

**THE ROLE OF AUTOMATION, BOTS AND AI IN INFLUENCING
KNOWLEDGE OF THE PAST**

Benedict Wills-Eve

This thesis is submitted for the degree of Doctor of Philosophy.

Lancaster University

Department of History

December, 2023

Word count: 73,948 words

Abstract

The Role of Automation, Bots and AI in Influencing Knowledge of the Past

Benedict Wills-Eve

This thesis explored the role of algorithms, automated computer programs ('bots') and artificial intelligence (AI) in the creation, interpretation and sharing of knowledge of the past online. It aimed to explore how these specific types of computational approaches influence the diversity of historical and cultural narratives that are researched and search through the Web, especially by institutions like museums and art galleries. The roles of the wide range of different stakeholders involved in these processes, from researchers to museum curators to software developers to interested users, were explored to better understand the human interactions often at the heart of automated approaches.

The historical context of automation, bots and algorithms on the Web is explored with regard to their roles in influencing the production, interpretation and sharing of knowledge about the past. Three case studies follow to explore specific types of automated approaches in more detail: the sharing of art and artefacts from online gallery and museum collections via automated social media accounts ('social bots'); the interpretation of historical figures and events by 'chatbots' through the Amazon Alexa platform; and the creation and sharing of historical knowledge and interpretation of online museum collections through the AI tool ChatGPT.

I argue that these processes and their influence on knowledge of the past can only be properly understood through a theoretical approach that focusses on the complex human and computational interactions inherent in the production and sharing of knowledge online. Automated approaches all offer potential benefits to cultural heritage institutions looking to engage wider audiences with interpretations of the past, but existing biases in the data available and algorithmic methods involved poses significant risks to the historical accuracy and trustworthiness of such generated material. Greater collaboration and

regulation are needed to prevent singular, uncritical interpretations dominating narratives of the past.

Table of Contents

List of Tables	5
List of Figures	6
Acknowledgements.....	10
Declaration.....	11
1. Introduction	12
1.1 Rationale for Case Studies	17
1.2 Why are Algorithms, Bots and AI Important for the Future of Knowledge about the Past?	23
2. A History of AI, Algorithms and Bots: The State of the Art in Context	26
2.1 The Nature of Bots and the Web.....	27
2.2 The History of Bots and the Web	31
2.2.1 Crawler Bots and the Early Years of the Web	31
2.2.2 Social Bots and the Web 2.0 Era	38
2.2.3 Chatbots: Fifty Years of History	57
2.3 Web 3.0: From Social Bots to Social Machines.....	64
2.4 Algorithms.....	72
2.4.1 Recommendation Algorithms.....	76
2.4.2 Algorithms and Computational Creativity	80
2.4.3 Foundation Models, Large Language Models, and Generative AI.....	81
3. Case Study: Bots Sharing GLAM Collections on Twitter	90
3.1 Survey of Bots Sharing History and Cultural Heritage on Twitter	93
3.2 Individual Twitter Bots and Their Approach to Sharing Content	101
3.3 The 'Off The Easel' Twitter Bot Community and the Sharing of GLAM Collections	110

3.4 Discussion: Influence is Key to Engaging a Twitter Audience with Online GLAM Collections	130
4. Case study: Talking to Bots about History via Amazon Alexa	140
4.1 Survey of Chatbots Relating to History and Cultural Heritage	142
4.2 Amazon Alexa Skills Survey	146
4.3 Discussion: Early Stages for Conversational Engagement with History and Cultural Heritage.....	179
5. Case study: Is the Novel AI Tool ChatGPT the Future of the Past?	187
5.1 Exploring ChatGPT and its Plugins in Context	189
5.2 Researching ChatGPT: Novel Technologies Require New Approaches	195
5.3 Exploring Knowledge about the Past Using ChatGPT with Plugins	201
5.4 Discussion: Exploring the Past with ChatGPT Involves Humans and Machines for Meaningful Interpretations	229
6. The Future of the Past: from Agents to ‘Agencies’	240
6.1 Information Inequalities and Historical and Cultural Biases	242
6.2 Multimodal Generative Approaches and their Implications	250
6.3 Digitisation and Digitalisation for Historical Research and GLAMs	257
6.4 Historical Information, Misinformation and Disinformation	263
6.5 The Future of Historical Social Machines: What Would a Foundation Model for History Look Like?	269
7. Conclusion	281
Appendices	287
Glossary	291
Bibliography	302

List of Tables

Table 1: Table taken from Tsvetkova et al. categorising online bots. This broad schema includes the types of bot which will be discussed below, namely crawler bots, social bots and chatbots.....	31
Table 2: Table taken from Just and Latzer outlining a classification of algorithmic selection applications.....	76
Table 3: Table showing Twitter engagement metrics for top ten individual bot accounts with the most followers.....	108
Table 4: Table showing Twitter engagement metrics for top ten Off The Easel bot accounts with the most followers.....	121
Table 5: Table showing network statistics for top ten most influential Off The Easel Twitter bot accounts in terms of highlighting content across the botnet. Note these bots focus on retweets for influence, shown by their low percentages of suggestion tweets.....	126
Table 6: Table showing network statistics for top ten recommenders within the Off The Easel Twitter bots network in terms of pointing followers towards content shared by other bot accounts. Note these bots focus on tweets suggesting other art bot content.....	127
Table 7: Table showing results of the chatbot survey across social messenger platforms; relevance determined as above in the methodology.....	146
Table 8: Search results returned for bots on Amazon Alexa; relevance determined as above in the methodology.....	157
Table 9: Example quiz Skill set of questions and answers for different historical periods and places.....	165
Table 10: The recommendation network statistics for top ten Skills ranked by stress centrality.....	177
Table 11: The recommendation network statistics for top ten Skills ranked by number of user ratings (popularity measure). N.B. the four Skills with zero in each column play no part in the network, showing that popular Skills often don't recommend each other.....	178
Table 12: Relevant plugins identified from ChatGPT plugin store.	204

List of Figures

Figure 1: Overview of the already crowded online social bot landscape. Source: venturebeat.com.	51
Figure 2: Overview of current chatbot landscape. Source: Carylyne Chan.	61
Figure 3: Diagram showing a graph made up of triples, the circles being subjects and objects and the arrows linking them together the predicates. Source: W3C RDF 1.1 Primer 2.	67
Figure 4: Examples of Twitter bot accounts (from top left to bottom right): @artistmonet (Off The Easel); @met_medievalart (Off The Easel); @DeathMedieval, (individual bot account); @MechCuratorBot (The British Library).	92
Figure 5: Overview of the Twitter bot survey data collection and analysis process	96
Figure 6: Bar chart showing categorisation of Twitter bots by sharing approach and content source.	102
Figure 7: Bar chart showing locations of institutions, collections or textual content shared by Twitter bots.	103
Figure 8: Pie chart showing regions covered by the content shared by bots (Worldwide means that a bot account shares content from multiple regions). 104	104
Figure 9: Pie chart showing the historical time period covered by the content shared by bots (Mixed means that a bot account shares content from multiple time periods).	105
Figure 10: Corpus linguistics analysis for all individual bot accounts. Table showing top ten terms in order of relative frequency. Measures of dispersion across the corpus are shown for each term. The keyword 'century' is highlighted and its collocates displayed.	107
Figure 11: Pie chart showing regional spread of artist Twitter bot accounts in the 'Off The Easel' network.	112
Figure 12: Pie chart showing regional spread of GLAM Twitter bot accounts in the 'Off The Easel' network.	113
Figure 13: Bar chart showing proportion of content representing different regions for GLAM bots in the 'Off The Easel' network; worldwide represents bots sharing content covering multiple regions.	113

Figure 14: Pie chart showing proportion of content representing different time periods for GLAM bots in the ‘Off The Easel’ network; mixed represents bots sharing content covering multiple periods.	114
Figure 15: Pie chart showing proportion of content representing different categories for GLAM bots in the ‘Off The Easel’ network; mixed represents bots sharing both art and artefacts in roughly equal measure.	114
Figure 16: Corpus linguistics analysis for the ‘Off The Easel’ art bot accounts. Table showing top ten terms in order of relative frequency. Measures of dispersion across the corpus are shown for each term. ‘Woman’ is highlighted and its collocates displayed.	116
Figure 17: Corpus linguistics analysis for Off The Easel GLAM bot accounts. Table showing top ten terms in order of relative frequency. Measures of dispersion across the corpus are shown for each term. ‘century’ is highlighted and its collocates displayed.	118
Figure 18: Bar chart showing the distribution of highly engaged followers by profession or area of interest.	122
Figure 19: Visualisation showing how the eighteen Off The Easel Metropolitan Museum of Art bot accounts form a connected and coordinated network via the hashtags that are included in some of their tweets.	129
Figure 20: Screenshots of interactions with Alexa on a smartphone. The top left request was carried out by Alexa itself and not an additional Alexa Skill. Top right to bottom left show three screenshots of interactions with different Alexa Skills.	148
Figure 21: Flow chart detailing the interview schema for interactions with Alexa Skills.	153
Figure 22: Visualisation showing part of the recommendation network for the history Skills. The direction of the arrows indicates which Skills recommend others (outdegree) and which are recommended by others (indegree).	156
Figure 23: Bar chart showing the number of Alexa Skills in each manually-assigned content category.	158
Figure 24: Top: word cloud showing most frequent terms used in Skills’ descriptions. Bottom: collocate network showing terms’ most frequent associations	160

Figure 25: Bar chart showing distribution of content sources across Skills for the search term 'history'	161
Figure 26: Bar chart showing the distribution of Skills' content per region and/or culture.....	169
Figure 27: Graph showing in-degree distribution of Skills in recommendation network	176
Figure 28: Graph showing out-degree distribution of Skills in recommendation network	176
Figure 29: Flowcharts showing the stages of Human Reinforcement Feedback Learning (HRL) involved in training GPT-4 (attribution: OpenAI; licence: CC BY 4.0 DEED).....	192
Figure 30: Flow chart showing the different layers of interaction in the PaaS/SaaS framework that encompasses OpenAI, ChatGPT and its plugins.	194
Figure 31: Graph of the datasets, organisations and software influenced by OpenAI, ChatGPT and GPT-4.	195
Figure 32: Screenshot of a ChatGPT-generated response giving information about the Battle of Hastings. This comes purely from the model's training data and does not involve any web browsing or plugins.	207
Figure 33: Screenshot of a ChatGPT conversation showing sources for the Battle of Hastings summary generated by ChatGPT purely from training data with no input from plugins.	208
Figure 34: Screenshot of an example ChatGPT conversation asking about the 1st Earls of Lonsdale, comparing responses from the model's training data and the Wikipedia plugin.....	212
Figure 35: Screenshot of a summary generated showing ChatGPT's inherent uncertainty about specific and potentially out of date information.....	213
Figure 36: Screenshot of a ChatGPT summary answer to the same question as in Figure 35 showing the content, links, attribution and clarity provided by the Wikipedia plugin.....	214
Figure 37: Screenshot of ChatGPT example history quiz question generations with the prompt.	216

Figure 38: Screenshot of the ChatGPT response showing the code for the Filtir fact-checking plugin reveals that it uses Wikipedia as its source of information for checking the accuracy of statements.	217
Figure 39: Screenshot of the first four generated questions and answers for a 'Samurai history' quiz created by combining the inputs of multiple ChatGPT plugins.	219
Figure 40: Screenshot of ChatGPT showing an example search result returned from the NASA Media Explorer plugin.	222
Figure 41: Screenshot of the ChatGPT follow-on response from the image in Fig 9 showing the use of the Web Requests plugin to find out more information about the ISS mock-up lab pictured in the image.	223
Figure 42: Screenshot of a ChatGPT response to the prompt asking for a SPARQL query.	226
Figure 43: Example screenshot of a conversation showing the trial and error process involved in attempting to get working SPARQL queries for some natural language searches. Note how ChatGPT explains its steps in the response.	228
Figure 44: An example screenshot of ChatGPT providing a confusing and contradictory response in the face of a user challenging its previously correct response.	233

Acknowledgements

I cannot thank my supervisors enough for all their support, encouragement and expertise throughout the process of writing this thesis (which has taken an eternity!). Paty and Jaiver, your endless patience and encouragement in getting this work across the line is much appreciated and I am extremely grateful for all your help over the past few years.

My family, especially wife and parents, have also managed to put up with the process of writing this thesis and have been a constant and much-needed source of support throughout. Big thanks in particular to Sam for making sure I stuck at it and finally got it all written!

And finally, I have to mention the person who has overseen the writing of the latter part of this thesis, despite also doing his best to delete large chunks of it, my son Toby. I can only hope that he gets to live in a world where academic discussion and debating of knowledge will not be controlled by robots in thrall of a powerful few.

Declaration

I declare that that the thesis is my own work and has not been submitted in substantially the same form for the award of a higher degree elsewhere.

1. Introduction

The Web has significantly changed how we access information, which has had implications for the creation, consumption and transmission of historical, cultural heritage and archaeological knowledge. This has been further impacted by the growing role of automation, algorithms and AI (artificial intelligence) in these processes, especially in how knowledge is searched for or, more commonly nowadays, curated for and presented to users across the Web. How that knowledge might come to be constructed, and reconstructed, is also changing as the development of more sophisticated AI tools poses more questions than answers when considering the inherently creative act of trying to understand the past. Hence, this thesis was born from the question: what does all this mean?

For GLAM institutions (Galleries, Libraries, Archives and Museums) this has meant that online collections have become accessible via websites and social media platforms (especially following the COVID-19 pandemic), where algorithms and 'bots' ((semi-)automated social media accounts) have been used to share collections across audiences on social media and make their interactions with these collections more engaging. Users can then interact with in novel ways, creatively reinterpreting objects via the latest AI tools before sharing this content with others through the algorithmically mediated spaces of online social platforms, perhaps blissfully unaware of the myriad historical, cultural and social biases amplified in the process.

For researchers and educators, the latest developments in AI models offer both opportunities and challenges in interacting with online data and information about the past, improving on basic 'chatbots' (automated programs that converse with a user via text or speech) which have already been used to try and engage people with history. Such engagement can aim to decolonise historical narratives and share the resulting histories with the same communities once denied the opportunity to do the same, only to find that the social structures embedded within these technologies of the Global North prevent still prevent such dreams from being realised.

As these automated and algorithmic technologies, both older and newer, continue to develop and grow in popularity and value, it is vital for professionals in research, educational and GLAM institutions to better understand how automation, algorithms and AI influence and mediate online interactions with the past so that their potential benefits and challenges are critically discussed.

In doing so, this thesis seeks to highlight that these technologies, which can so easily be perceived as individual, highly complex and often mysterious entities (especially in popular depictions of AI), are actually the products and producers of tangled networks involving many different human users, from developers to museum visitors, and the computational hardware, software and methods they employ. All taken together, these 'sociotechnical networks' form a central argument of this thesis, positing that the theoretical agencies of these complex interactions are the key to understanding what it means to create, discover, share and reimagine knowledge about the past on the Web. In the face of technological change it is easy to focus on what tools can and cannot do, but this thesis reminds us that it is the humans who use these tools and the things they create that are always at the heart of such discussions.

As such, this work is firmly positioned within the interdisciplinary praxis of Digital Humanities research, which is well suited to exploring the complex interactions inherent in sociotechnical systems¹. Previous and ongoing work in this area has highlighted different aspects of such systems, from the physical and digital infrastructures which underpin and shape interactions², to the effects of existing data structures, algorithms and machine learning models on researchers and users when trying to make sense of complex, messy and ontologically diverse information about the past³.

¹ James O'Sullivan, *The Bloomsbury Handbook to the Digital Humanities* (Bloomsbury Publishing, 2022).

² Javier Pereda, 'Where Do I Stand? Deconstructing Digital Collections [Research] Infrastructures: A Perspective from Towards a National Collection' (presented at the Open and Engaged 2023: Community over Commercialisation, British Library, 2023), doi:10.23636/9fh3-ze12; Thomas S. Mullaney, 'QWERTY in China: Chinese Computing and the Radical Alphabet', *Technology and Culture*, 59.4 (2018), pp. S34–65.

³ Patricia Murrieta-Flores, Mariana Favila-Vázquez, and Aban Flores-Morán, 'Indigenous Deep Mapping: A Conceptual and Representational Analysis of Space in Mesoamerica and New Spain', in *Making Deep Maps* (Routledge, 2021); Gustavo Candela and others, 'An Ontological Approach for Unlocking the Colonial Archive', *Journal on Computing and Cultural Heritage*, 16.4 (2023), pp. 1–18, doi:10.1145/3594727; Piraye Hacıgüzeller, James Stuart Taylor, and Sara

This work also stems from a growing call within scholarship exploring the influences of the sociotechnical systems inherent in the Web on information about the past towards decolonisation of these systems themselves. The role of Digital Humanities research in drawing on approaches and insights from various disciplines has highlighted the inequalities and subjectivities of technologies that are often abstracted in the influences of society, history and culture, and that then come to be perpetuated by the sociotechnical systems of today⁴. The complexities that emerge from this become clear when systems force ambiguity upon the roles of researchers and their methods, whilst simultaneously attempting to strip such nuance from the commercialised end product of information⁵. Within this academic context, this thesis aims to extend such discussions of power, representation and perspective to include the entangled and ambiguous roles of automated and AI-enabled bots in these sociotechnical networks.

This thesis therefore set out to understand for the first time the roles of bots and algorithms in collecting, curating, creating and disseminating historical information via the Web, an area as yet unexplored despite the growing commercial and governmental interest in online automation and its social consequences. This impact of automation, bots and AI on public engagement and interactions with knowledge about the past, especially via GLAM institutions, was explored through addressing the following questions:

- Do AIs and algorithms exert influence over public knowledge of history and heritage primarily explicitly (dissemination of historical narratives), implicitly (algorithmic recommendation), or both?

Perry, 'On the Emerging Supremacy of Structured Digital Data in Archaeology: A Preliminary Assessment of Information, Knowledge and Wisdom Left Behind', *Open Archaeology*, 7.1 (2021), pp. 1709–30, doi:10.1515/opar-2020-0220.

⁴ Roopika Risam, 'Decolonizing the Digital Humanities in Theory and Practice', in *The Routledge Companion to Media Studies and Digital Humanities* (Routledge, 2018); Catherine D'Ignazio and Lauren F. Klein, *Data Feminism, Strong Ideas* (Penguin Random House, 2023) <<https://www.penguinrandomhouse.com/books/654051/data-feminism-by-catherine-dignazio-and-lauren-f-klein/9780262358538/>>.

⁵ Shawn Graham, *An Enchantment of Digital Archaeology: Raising the Dead with Agent-Based Models, Archaeogaming and Artificial Intelligence* (Berghahn Books, 2020); Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (MIT Press, 2000).

- What role(s) do developers, users and researchers play in influencing the content produced / disseminated by AIs and algorithms?
- How do the underlying internet and information geographies of the Web and social media shape the activities of AI and algorithms with regard to history and heritage?
- How might novel generative algorithms influence public and academic perceptions and interpretations of historical content and data?
- In this constantly evolving, rapidly expanding area, what might the future hold for researchers, developers, heritage professionals and public users?

These questions were addressed through three case studies. The first case study, detailed in Chapter 3, explored the role of bot accounts on social media, with a focus on Twitter (now 'X'), sharing and suggesting artworks and objects from online GLAM collections to users. The second case study, detailed in Chapter 4, looked at how chatbots are used to engage and educate users about history and cultural heritage, first surveying the use of chatbots more generally before focussing in on the conversational assistant Amazon Alexa and the wide range of Alexa apps called 'Skills' specifically related to history, such as historical quizzes or fun facts. The final case study, detailed in Chapter 5, explored the novel AI tool ChatGPT, a text-based conversational platform that can also have apps called 'Plugins' added to it, with the plugins most relevant to engaging with historical knowledge and GLAM collections forming the focus of analysis.

The rationale behind the choices of these case studies is that they have covered the majority of the current use cases for the combination of AI and automation across the time period of this thesis, which began back in 2018. It is important to note that the rapid pace of change over the ensuing five years has seen shifts in the technologies available, with the first two case studies reflecting the main use cases before Covid-19, and the ChatGPT case study highlighting the recent rise of AI tools that now form the most obvious avenues for further investigation and expansion. However, across all case studies

fundamental themes of infrastructure, power dynamics, social and historical biases and complex interactions between different people involved, from developers, to online museum visitors, emerged that are as relevant today as they were thirty years ago in the early years of the Web. Hence, the central arguments are of this thesis around automated and AI approaches having the potential to increase access to knowledge about the past, but simultaneously risk amplifying its inherent and associated biases, shows that the historical context for this work (detailed in Chapter 2), especially the very recent context, is vital for fully understanding these fundamental issues.

The study of automation, bots and AI in the context of online information about the past requires novel, adaptable methodologies, especially given the pace of technological change in this area. Although the specific methodologies for each case study are outlined in more detail in the following sections, it is important to note here that the different methodologies employed throughout this thesis, and the experimental approaches taken to developing and testing them, are in themselves important contributions to the field. Rooted in the interdisciplinary and exploratory nature of Digital Humanities research, exemplified in Graham's recognition of a 'digital enchantment' stemming from research insights gained from exploring the possibilities of digital tools and systems, viewing the process of methodology development as a crucial part of the research findings, the approaches described in this work required similarly playful and creative processes in both their development and implementation⁶.

The results of this novel methodological work, from attempting to create qualitative interview schema for chatbots to the slippery trial and error of nudging large language models into generating synthetic archaeological datasets, highlight the fact that any research grounded in the unfathomable sociotechnical complexities of the Web, and especially generative AI models, will require approaches that are necessarily and unambiguously subjective, in

⁶ Andrew Prescott, 'Mixed Methods and the Digital Humanities', in *Mixed Methods and the Digital Humanities* (Bielefeld University Press, 2023), pp. 27–42, doi:10.1515/9783839469132-004; Graham, 'An Enchantment of Digital Archaeology'.

contrast to the quantitative evaluations usually favoured by computer science methods⁷.

1.1 Rationale for Case Studies

On social media platforms like X (formerly Twitter), it has been shown that bots, intelligent or otherwise, vastly outnumber their human counterparts⁸. Some of these ‘social bots’ are harmless, some useful, whilst others secretly collect user information and spread misinformation⁹. Such activity has been cited as influencing the 2016 US Presidential Elections and the Brexit referendum in the UK, although unravelling the respective roles of bots and humans in these processes is always challenging¹⁰. It is impossible to separate history from politics, as clearly demonstrated in the Brexit debate on social media platforms where arguments around Britain and Europe’s history and heritage were invoked to entrench ideological positions¹¹. Hence, it is apparent that social media, and thus social bots (and their human creators), have significant potential to influence the ways in which historical and heritage-related content are disseminated to and consumed by a growing online audience.

The literature around bots disseminating information on social media largely focusses upon such examples of political and economic influence, overlooking the related and equally important areas of history and heritage. There are many benign bots on social media that share historical facts and interpretations that

⁷ Laura Weidinger and others, ‘Sociotechnical Safety Evaluation of Generative AI Systems’ (arXiv, 2023), doi:10.48550/arXiv.2310.11986.

⁸ Igal Zeifman, ‘Bot Traffic Report 2016’, *Incapsula Blog*, 2017
<<https://www.incapsula.com/blog/bot-traffic-report-2016.html>> [accessed 24 August 2018].

⁹ Samantha Bradshaw and Philip N Howard, ‘Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation’, 2017, p. 37.

¹⁰ Philip N. Howard and Bence Kollanyi, ‘Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum’, *arXiv:1606.06356 [Physics]*, 2016
<<http://arxiv.org/abs/1606.06356>> [accessed 13 February 2019]; Philip N. Howard, Samuel Woolley, and Ryan Calo, ‘Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration’, *Journal of Information Technology & Politics*, 15.2 (2018), pp. 81–93, doi:10.1080/19331681.2018.1448735.

¹¹ Chiara Bonacchi, Mark Altaweel, and Marta Krzyzanska, ‘The Heritage of Brexit: Roles of the Past in the Construction of Political Identities through Social Media’, *Journal of Social Archaeology*, 2018, p. 1469605318759713.

thousands of humans interact with on a daily basis ¹²; this activity needed to be explored. GLAM institutions share pieces of their collections, officially and unofficially, through such bots yet little was known about the scale, scope or results of this activity. The first case study set out to explore this area.

Carried out in 2019, three years before Elon Musk's takeover and rebranding of the Twitter platform to X, this first case study analysed automated Twitter 'bot' accounts that shared primarily artworks from publicly available online gallery and museum collections. This work compared the activity of different Twitter bots sharing content individually with that of a coordinated group of Twitter bots that shared content and engaged with each other as part of a community then known as 'Off The Easel' (now called 'Bot Frenz'). This case study sought to address the following points:

- How many social bots exist, including individuals and members of the 'Off The Easel' botnet
- What sort of content they disseminate across the fields of cultural heritage (including art) and history
- The sources of such content, for example Open Access collection of GLAMs
- The roles of algorithms, developers and GLAMs in curating content and determining how it is shared
- How many people follow these bots on social media and whether they follow to be entertained, educated or both
- How the automated and algorithmic activity of bots influences the ways in which users consume the content shared
- How all of this activity differs between individual bots and those that are part of the 'Off The Easel' botnet

¹² L. Kelly Fitzpatrick, 'Anatomy of a Museum Twitter Bot', *Berkman Klein Center Collection*, 2017 <<https://medium.com/berkman-klein-center/anatomy-of-a-museum-twitter-bot-2311d81de243>> [accessed 16 November 2018].

The main findings were that the automated sharing of collections reflected wider information inequalities, with the majority coming from GLAM institutions in Western Europe and the USA, and the choices of museum curators and bot developers in how collections were digitalised and shared. It was striking that the GLAM institutions themselves were not the ones actually running the automated accounts doing the sharing, the majority were run through a coordinated bot community overseen by an individual developer. Although the content shared by these bots came directly from the GLAM collections online, the fact that it was then shared on a social media feed in a necessarily fragmented way led to a different experience than viewing collections on the GLAM website itself. It was also clear that for many objects within such digitalised collections, little was known or had been added to their records, leading to absent or minimalistic descriptions with little or no interpretation.

The second case study moved from bots on social media to consider chatbots - bots and AIs that are to some extent conversational in nature – on a range of platforms. This chapter surveyed chatbots that appeared on social messaging platforms, such as Facebook Messenger and WhatsApp, before focussing on Amazon Alexa, a ‘virtual conversational assistant’ that is a chatbot which can be added to through apps called Alexa Skills. Chatbots have become popular in the corporate world and many large businesses now have chatbots to help customers carry out tasks online. The GLAM sector, especially museums, have been experimenting with using chatbots for the last decade.

This case study was carried out in 2019, when the interest in GLAM chatbots was still in its early stages and was growing, but when technological limitations meant that most chatbots were not truly conversational, providing basic information about museum opening times rather than interactive dialogues about collections or exhibitions¹³. By focussing on chatbots and the growing amount of historical ‘Alexa Skills’ (think conversational smartphone apps) available on Amazon Alexa, this chapter examined what content is shared to

¹³ Angeliki Tzouganatou, ‘Can Heritage Bots Thrive? Toward Future Engagement in Cultural Heritage’, *Advances in Archaeological Practice*, 6.4 (2018), pp. 377–83, doi:10.1017/aap.2018.32. ‘Can Heritage Bots Thrive? Toward Future Engagement in Cultural Heritage’

the public, where it comes from and how it is interpreted. It examined this from the perspectives of researchers, GLAM and educational professionals, app developers and users.

The chapter aimed to survey the landscape on chatbots related to history and heritage as of 2019, and in doing so sought to answer the following questions:

- How many history and heritage chatbots exist and where can they be found?
- What topics are covered by their content and how accurate and representative is it?
- Who develops these chatbots and from which sources does their content originate?
- How do users engage with chatbots and what roles do the platforms that host them play in shaping this engagement?
- How might chatbots be used to increase public engagement with historical information, from entertainment to education?

The initial chatbot survey found that while there were plenty of corporate chatbots available on apps like Facebook Messenger, there were very few relating to history or GLAMs and those that did exist were information / customer-service type bots that gave the opening hours for particular museums. It also highlighted an academic interest in GLAM chatbots, but very few actual bots being produced, and those that were developed remained in proof-of-concept stage and were not maintained for long.

The far more in depth survey of Alexa Skills, effectively chatbot add-ons for Amazon Alexa, related to history showed a different picture. While there were almost no Skills specifically related to GLAMs on the Alexa platform, there were well over a hundred active ones related to history, covering general history facts or quizzes to information about certain time periods, geographical areas or historical figures. The main findings from this work were that the historical knowledge shared by these Skills came primarily from the USA, as that was where the majority of these Skills individual developers were based. Although

there was an educational element to many of these Skills, with the most popular being produced by organisations like The HISTORY Channel in the USA, the fundamental limitations of the technology meant that Alexa Skill developers often had to automate extraction of information from Wikipedia or simply manually update the chatbots themselves, for example adding in new quiz questions at regular intervals. Couple this with the annoying and awkward voice interactions that were the baseline for Alexa to begin with, and overall its capability for the automated sharing historical knowledge, especially educationally, could definitely be improved

However, these findings do reinforce the wider points around automation of knowledge sharing happening in complex sociotechnical networks that are subject to all manner of information biases and limitations. As is highlighted time and again throughout the thesis, a key part of exploring these practices is to look for the humans in the loop at every stage of the process – from the individual developer who decides to make a What Happened on This Day in History chatbot, to the Wikipedia editor who maintains the Events that Happened On This Day page on Wikipedia from which the information is automatically extracted, to the user who wakes up every morning to hear Alexa talking about another random history fact to start that person's day.

It is these complex, often mundane interactions which are as central to the sharing of historical knowledge as the algorithms within Amazon's recommender system, or the technology that converts text-to-speech when Alexa conveys that information. This theme is also apparent in the final case study which examines the recent advances in AI technology that is touted to bring about a whole new range of chatbots, including an updated and much enhanced Amazon Alexa.

The third case study explores the more recent developments in the growing field of generative AI, in doing so highlighting the changes that occurred in the three-year gap in the timing of this case study and the first two. Carried out in 2023, this case study focussed on exploring the capabilities of the ChatGPT tool which has been the subject of much recent hype.

Rapid developments have been (and continue to be) made in the AI models that can generate text, images, video, audio and other types of data. These ‘large language models’, or ‘foundation models’, are now able produce text, images and audio, with tools like OpenAI’s ChatGPT achieving generations of sufficient sophistication that they have been described as the forerunners to AGI – ‘Artificial General Intelligence’ – as seen in Hollywood depictions of AI.

Of concern for this thesis is the extent to which these impressive abilities can change the way online audiences, GLAM professionals, researchers and developers interact with and influence knowledge about the past. The development, operation, use and potential misuse of such models also introduces a new networks of ‘social machines’, the entanglements humans, computer programs and the Web, that cover a far broader spectrum of automated machine and human interaction than seen in the previous case studies. ChatGPT and similar models have the potential to alter the Web in ways similar to that seen in the past by the likes of Google, Wikipedia or Facebook, hence this chapter focussed as much on the sociotechnical ecosystems in which ChatGPT operates as the platform itself.

To explore further how ChatGPT influenced that production, reinterpretation and sharing of knowledge about the past in research and GLAM institutions, this chapter aimed to:

- Assess the current capabilities of AI models in understanding and generating content about history and cultural heritage
- Explore potential applications of AI models in research, education and GLAM settings
- Examine the wider sociotechnical ecosystems of AI models, especially ChatGPT, and how these influence potential information biases
- Explore the ethical implications of using such models, with a focus on information biases, and the potential future opportunities and challenges

The main findings of this case study were that the complexity and inscrutability of the latest generative AI tools like ChatGPT effectively amplified the existing opportunities for sharing but also many of the concerns around biases and the potential for generated errors, the sources of which were harder to pinpoint in

the complex ecosystems that have already sprung up around ChatGPT and which GLAM institutions in particular would have to approach with caution.

ChatGPT, with its general natural language capabilities, offers the possibility of far more engaging interactions with historical knowledge and GLAM collections than either Twitter bots or Alexa Skills could provide, however given its ability to generate new information on demand, working out how these capabilities can be best implemented within the existing infrastructures of GLAMs, research and educational institutions will require much human input; such issues around copyright are particularly fraught. This highlights a future in which understanding the interactions between the humans and bots caught up in these 'historical social machines' will be the key to working together to share knowledge about the past in engaging ways that do not perpetuate current biases.

1.2 Why are Algorithms, Bots and AI Important for the Future of Knowledge about the Past?

An increasingly vast and global public accesses, interprets and disseminates historical information via the World Wide Web (or Web for short). Google searches, Wikipedia articles and the plethora of posts and pictures present on social media platforms all form a first port-of-call for most users looking for information online¹⁴. The Web is a vastly complicated sociotechnical system, that is one in which the technologies which support its existence and the people who use it act alongside one another to shape its reality and meaning, which is full of data, information and knowledge.

When someone searches for historical information on Wikipedia, they are processing a combination of text and images that has already been created, interpreted and reinterpreted, both as data and information, by human editors and computers. The accumulation of such information by a user to increase their understanding of the past is what produces knowledge of a given historical

¹⁴ Mostafa Mesgari and others, "The Sum of All Human Knowledge": A Systematic Review of Scholarly Research on the Content of Wikipedia.', *Journal of the Association for Information Science and Technology*, 66.2 (2015), pp. 219–45, doi:10.1002/asi.23172.

'fact', however they have chosen to interpret it¹⁵. Appreciating these subtleties in how information is created, processed and shared across the Web is important when it comes to considering what happens when algorithms and bots get involved.

This thesis argues that examining this automated and algorithmic activity in more detail highlights much of the unseen infrastructure, human work and implicit biases that pervade the current ways in which knowledge is shared, and that this in itself relies on such activity in ways that are often overlooked. As such, using AI and automation to create, reinterpret and share knowledge about the past online is much more about the very human processes of international regulation, collaboration between professional groups, involving all stakeholders and groups whose histories and worldviews have been suppressed and fundamentally recognising that these technologies might stem primarily from the USA, with data primarily sourced from the Global North, but they do not operate in a vacuum.

Each of the case studies in the thesis highlights the fundamental roles and influences of commercial entities in the sociotechnical systems surrounding bots, AI agents and the various groups of users with which they are involved. These fundamental influences often illustrate power imbalances, especially where online social platforms are concerned, that lead to commercial organisations, via their leading figures and development teams, directly and indirectly shaping the types of content and interactions that are possible and legitimised. From the algorithmic banning of art bots on Twitter and Facebook, to legal grey areas about responsibility for the accuracy and sourcing of text generated by ChatGPT, the actions of organisations apparently striving to democratise information for all users can often seem to be 'democratising' it only for some¹⁶.

¹⁵ Anett Hoppe and others, 'Wisdom - the Blurry Top of Human Cognition in the DIKW-Model?', in *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-2011)* (presented at the 7th conference of the European Society for Fuzzy Logic and Technology, Atlantis Press, 2011), doi:10.2991/eusflat.2011.91.

¹⁶ Katherine Cook, 'EmboDIYing Disruption: Queer, Feminist and Inclusive Digital Archaeologies', *European Journal of Archaeology*, 22.3 (2019), pp. 398–414 (p. 398), doi:10.1017/eaa.2019.23.

It is proposed that a theoretical framework shifting the focus from algorithms, bots and AI models as ‘agents’ within systems, to understanding the complex networks of interactions within such systems, their ‘agencies’, is crucial if progress is to be made in ensuring balanced, nuanced and equitable sharing of knowledge about the past online. Extending the idea of sociotechnical systems, this approach aims to highlight that the coming wave of generative AI-enabled technology has the ability to change knowledge about the past as much as it does to amplify it, mainly thanks to often unseen and interlinking actions of numerous automated bots and humans that underpin the current creation and sharing of information on the Web.

As this thesis goes on to demonstrate, continuing to overlook these complex and subtle networks of power poses serious risks for the future of our knowledge about the past. The first step in exploring and understanding these interactions is to start with a historical review of these technologies, their place on the Web and their interactions with humans.

2. A History of AI, Algorithms and Bots: The State of the Art in Context

The brief overview of bots and algorithms presented in the preceding section emphasised the ambiguity of terminology around ‘artificial intelligence’, ‘bots’ and ‘algorithms’. To the non-expert, the three are often conflated with little change in meaning, whilst technical definitions still struggle to capture and classify the sheer variety of approaches and phraseology to be found within each field: is it a bot, a script, a program, an algorithm or all of the above? ¹ Classifying the countless bots that inhabit the Web has become a task akin to that of an ecologist trying to determine the constitution of a rainforest ecosystem ². Each ‘species’, as Leonard put it ³, may be determined by function, operation, design or merely perceived wisdom, but these definitions often contain considerable fluidity. Latzko-Toth describes the idea of a ‘bot’ as an “imaginary of autonomy” ⁴, placing the definition squarely in the mind of the user; if they think it’s a bot then a bot it shall be, regardless of technicalities. Deciphering this complex world of bots and algorithms, their histories and the research activity around them will be the focus of this summary.

Changing terminologies and definitions often point to diverging histories with far from predictable evolutionary trajectories or starting points, as demonstrated by Naughton⁵ for the history of the Internet and by Schmidhuber⁶ for AI as a whole. In trying to relate a history of bots the same complex patterns, and tangled messes, soon emerge from countless historical threads. Amongst these are the technical histories of AI, the Internet and the Web, and social histories of

¹ Stan Franklin and Art Graesser, ‘Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents’, in *Intelligent Agents III Agent Theories, Architectures, and Languages*, ed. by Jörg P. Müller, Michael J. Wooldridge, and Nicholas R. Jennings, Lecture Notes in Computer Science (Springer Berlin Heidelberg, 1997), pp. 21–35.

² Robert Gorwa and Douglas Guilbeault, ‘Unpacking the Social Media Bot: A Typology to Guide Research and Policy’, *arXiv:1801.06863 [Cs]*, 2018 <<http://arxiv.org/abs/1801.06863>> [accessed 15 August 2018].

³ Andrew Leonard, *Bots: The Origin of New Species* (Penguin Books Limited, 1998).

⁴ Guillaume Latzko-Toth, ‘The Socialization of Early Internet Bots: IRC and the Ecology of Human-Robot Interactions Online’, *Socialbots and Their Friends*, 2016, p. 2, doi:10.4324/9781315637228-10.

⁵ ‘The Evolution of the Internet: From Military Experiment to General Purpose Technology’, *Journal of Cyber Policy*, 1.1 (2016), pp. 5–28, doi:10.1080/23738871.2016.1157619.

⁶ ‘2006: Celebrating 75 Years of AI - History and Outlook: The Next 25 Years’, *arXiv:0708.4311 [Cs]*, 2007 <<http://arxiv.org/abs/0708.4311>> [accessed 11 October 2018].

flourished and failed technologies and economic histories of automation, positive and negative. These broader ‘sociotechnical’ narratives are beyond the scope of this summary, but a few of the patterns and tangles which link them to the situation today cannot be overlooked, and the overarching thread of sociotechnical entanglement weaves its way into every aspect of this thesis.

With this in mind, it seems logical to start with the variety of bots in the present, necessarily delving back into their distinctive, yet related, pasts to establish some context. Next will come a typological examination of online bots that includes the ways that they currently, and may in future, affect historical and public perceptions of the past. This will be followed by a summary of the current and historical explorations of the algorithms that often work alongside or within such bots on the Web.

Although before we begin, a quick note on terminology is needed. Today, ‘the Internet’ and ‘the Web’ are often used synonymously and interchangeably in common speech, but technically they are not the same thing. The Internet is the physical infrastructure which connects computers (or more accurately, networks of computers) across the globe via hundreds of thousands of kilometres of undersea, underground and overground cables⁷. The Web, or more properly the World Wide Web, is the unimaginably vast network of linked documents and resources that sit on computers physically connected via the cables of the Internet. It is actually the ‘Web’ that most people are referring to when they talk about ‘the Internet’.

2.1 The Nature of Bots and the Web

The origins of the term ‘bot’ as a colloquial shortening of ‘robot’ infers an entity that carries out tasks with (some) autonomy, or at least seems to do so. The infamous Mechanical Turk²⁴, an 18th century chess-playing robot that was little

⁷ Business Insider Prachi Bhardwaj, ‘Fiber optic wires, servers, and more than 550,000 miles of underwater cables: Here’s what the internet actually looks like’, *Business Insider Deutschland*, 2018 <<https://www.businessinsider.de/how-internet-works-infrastructure-photos-2018-5>> [accessed 19 August 2019].

²⁴ Tom Standage, *The Mechanical Turk: The True Story of the Chess-Playing Machine That Fooled the World* (Penguin, 2003).

more than a man concealed beneath a chessboard pulling the strings of a puppet player above, has fittingly leant its name to a platform run by Amazon which pays humans to perform research tasks that computers still struggle to accurately and efficiently complete²⁵. The identification of whether a user is a person or a bot, especially on social networking platforms like Twitter, has ramifications for cybersecurity experts and policy makers trying to curb political manipulation and the dissemination of ‘fake news’²⁶. This tends to focus on bots impersonating humans, but the deception can work both ways with some humans starting to impersonate bots in order to gain followers²⁷, or in commercial sectors where some ‘AI’ solutions are actually just teams of human workers hidden inside the metaphorical ‘black box’ of algorithmic mystery and secrecy²⁸. These 21st century ‘Mechanical Turks’ demonstrate the first, and possibly hardest, choice faced by the bot taxonomist: bot or not? (Ironically, there’s a bot that tries to determine this automatically, with limited success)²⁹. Such deliberations were at the forefront of Alan Turing’s mind, back in 1951, in his famous paper ‘The Imitation Game’ proposing the ‘Turing Test’, in which a human communicates with another user via a text interface before deciding whether or not that user is human or an AI³⁰. As we shall see later, some AIs (and bots) are developed with the sole purpose of trying to pass the Turing Test.

The uncertain nature of bots’ online identity, and the implicit role of humans in their creation and often ongoing activity, has led to the term ‘cyborg’ being used

²⁵ Gabriele Paolacci and Jesse Chandler, ‘Inside the Turk: Understanding Mechanical Turk as a Participant Pool’, *Current Directions in Psychological Science*, 23.3 (2014), pp. 184–88, doi:10.1177/0963721414531598.

²⁶ David M. J. Lazer and others, ‘The Science of Fake News’, *Science*, 359.6380 (2018), pp. 1094–96, doi:10.1126/science.aa02998.

²⁷ Taina Bucher, ‘About a Bot: Hoax, Fake, Performance Art’, *M/C Journal*, 17.3 (2014) <<http://journal.media-culture.org.au/index.php/mcjourn/article/view/814>> [accessed 29 October 2018].

²⁸ Olivia Solon, ‘The Rise of “Pseudo-AI”: How Tech Firms Quietly Use Humans to Do Bots’ Work’, *The Guardian*, 6 July 2018, section Technology <<https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies>> [accessed 3 February 2019].

²⁹ Clayton Allen Davis and others, ‘BotOrNot: A System to Evaluate Social Bots’, in *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW ’16 Companion (International World Wide Web Conferences Steering Committee, 2016), pp. 273–74, doi:10.1145/2872518.2889302.

³⁰ Luciano Floridi, Mariarosaria Taddeo, and Matteo Turilli, ‘Turing’s Imitation Game: Still an Impossible Challenge for All Machines and Some Judges—An Evaluation of the 2008 Loebner Contest’, *Minds and Machines*, 19.1 (2009), pp. 145–50, doi:10.1007/s11023-008-9130-6.

for bots that are simultaneously automatic and under a degree of human control³¹. Although often associated with human users, ‘sockpuppets’ and ‘trolls’ can also be automated (bots)³², or semi-automated (cybots), and these terms have begun to be used in this way when considering social media in the political arena³³. Sockpuppets are accounts set up as fake or misleading identities to deceive users, manipulate conversations and increase or decrease the influence of one side of a given debate. Trolls, in this new “politically motivated”³⁴ context, can be automated bots or human users who create content for bots to then amplify across social media platforms³⁵. ‘Trolls’ usually refer to human users (and historically this was the only usage of the term), either on forums, chatrooms or social network platforms who bait, abuse and cyberbully other users³⁶, but are now increasingly thought of in terms of automated action. The situation becomes more confusing when considering terminology used in other countries: in Poland, for example, ‘bots’ and ‘trolls’ are often used interchangeably to mean automated users involved in spreading misinformation³⁷. Despite a recent ‘bot purge’ carried out by Twitter to remove fake or malicious automated accounts, which ended up locking out several legitimate human users, all of these problems remain³⁸.

Of course, for both algorithms and AIs it can be argued that the very act of development by a human, or even simply the use of data created from human activity, makes any bot inherently ‘cyborgian’, a term coined by Donna Haraway

³¹ Zi Chu and others, ‘Who Is Tweeting on Twitter: Human, Bot, or Cyborg?’, in *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC ’10 (ACM, 2010), pp. 21–30, doi:10.1145/1920261.1920265.

³² Marco Bastos and Dan Mercea, ‘The Public Accountability of Social Platforms: Lessons from a Study on Bots and Trolls in the Brexit Campaign’, *Phil. Trans. R. Soc. A*, 376.2128 (2018), p. 20180003, doi:10.1098/rsta.2018.0003.

³³ Gorwa and Guilbeault, ‘Unpacking the Social Media Bot’.

³⁴ *ibid.*

³⁵ Chengcheng Shao and others, ‘Anatomy of an Online Misinformation Network’, *PLoS ONE*, 13.4 (2018), pp. 1–23, doi:10.1371/journal.pone.0196087.

³⁶ Gabriella Coleman, ‘Phreaks, Hackers and Trolls: The Politics of Transgression and Spectacle’, in *The Social Media Reader*, ed. by Michael Mandiberg (New York University Press, 2012).

³⁷ Robert Gorwa, *Computational Propaganda in Poland: False Amplifiers and the Digital Public Sphere — Oxford Internet Institute* (2017) <<https://www.oi.ox.ac.uk/blog/computational-propaganda-in-poland-false-amplifiers-and-the-digital-public-sphere/>> [accessed 13 February 2019].

³⁸ BBC, ‘Twitter Bot Purge Prompts Backlash’, 21 February 2018, section Technology <<https://www.bbc.com/news/technology-43144717>> [accessed 15 October 2018].

to explore the relationships between the body and its extension by interaction with the surrounding ‘technologies’. In her theoretical examples, such a technology could be a fire for cooking or a cat for pest control (with the cat being as much a ‘cyborg’ as the human), or a phone or Web for communication³⁹. These ideas naturally blend with the likes of Bruno Latour’s Actor-Network Theory (ANT)⁴⁰, at the heart of which is the idea that any actions between entities, human and non-human, form a network of connections that are as important as the actors (those doing the acting) and actants (those being acted upon) themselves. The simultaneous simplicity and complexity of this approach has made it popular for scholars studying the Web, a literal network, especially as it is a sociotechnical system in which the roles of the technology itself and the multifaceted society that uses it are entangled and dependent upon each other⁴¹. The inherent ambiguities and complexities that arise and multiply when bots, humans and varying degrees of ‘cyborg’ are considered together makes these theories a useful starting point for discussion, as seen in recent papers exploring the role of criminal botnets as cyborgian hybrid actor-networks⁴², the sociotechnicity of Wikipedia and its cyborgian ‘immune system’⁴³, and the inherent sociotechnical nature of social bots influencing politics⁴⁴. These approaches hold great promise for making such highly complex areas easier to understand, especially when they are applied to concrete examples such as the wide range of stakeholders involved in the digitalisation and dissemination of GLAM collections⁴⁵, and so a theoretical framework will be developed for this thesis which ties together these ideas (see Chapter 6).

³⁹ Donna Haraway, ‘A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20th Century’, in *The International Handbook of Virtual Learning Environments*, ed. by Joel Weiss and others (Springer Netherlands, 2006), pp. 117–58, doi:10.1007/978-1-4020-3803-7_4.

⁴⁰ Bruno Latour and Centre de Sociologie de l’Innovation Bruno LaTour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (OUP Oxford, 2005).

⁴¹ see Wiebe E. Bijker, Thomas Parke Hughes, and Trevor Pinch, *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (MIT Press, 2012).

⁴² Wytske van der Wagen and Wolter Pieters, ‘From Cybercrime to Cyborg Crime: Botnets as Hybrid Criminal Actor-Networks’, *The British Journal of Criminology*, 55.3 (2015), pp. 578–95, doi:10.1093/bjc/azv009.

⁴³ J. Riedl and A. Halfaker, ‘Bots and Cyborgs: Wikipedia’s Immune System’, *Computer*, 2012, 79–82.

⁴⁴ Dhiraj Murthy, ‘Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital’, 2016, p. 20.

⁴⁵ Katja Kaiser and others, ‘Promises of Mass Digitisation and the Colonial Realities of Natural History Collections’, *Journal of Natural Science Collections*, 11 (2023), pp. 13–25.

There are various types of bot and classification of bots is a subjective, contested area with differing typologies often having been developed with the purpose of delineating ‘good’ bots from ‘bad’ (see, for example, the work of Gorwa and Guilbeault⁴⁶, Maus⁴⁷, and Tsvetkova et al⁴⁸). Table 1, from Tsvetkova et al⁴⁹, broadly categorises such bots and acts as a starting point for the historical and current exploration of online bots, which will be extended to cover their specific roles in the areas of history and cultural heritage.

Table 1. Categorization of Internet bots according to the intended effect of their operations and the kind of activities they perform, including some familiar examples for each type.

	Benevolent	Malevolent
Collect information	<ul style="list-style-type: none"> • Web crawlers • Bots used by researchers 	<ul style="list-style-type: none"> • Spam bots that collect e-mail addresses • Facebook bots that collect private information
Execute actions	<ul style="list-style-type: none"> • Anti-vandalism bots on Wikipedia • Censoring and moderating bots on chats and forums 	<ul style="list-style-type: none"> • Auction-site bots • High-frequency trading algorithms • Gaming bots • DDoS attack bots • Viruses and worms • Clickfraud bots that increase views of online ads and YouTube videos
Generate content	<ul style="list-style-type: none"> • Editing bots on Wikipedia • Twitter bots that create alerts or provide content aggregation 	<ul style="list-style-type: none"> • Spam bots that disseminate ads • Bot farms that write positive reviews and boost ratings on Apple App Store, YouTube, etc.
Emulate humans	<ul style="list-style-type: none"> • Customer service bots • @DeepDrumpf and poet-writing bots on Twitter • AI bots, e.g. IBM's Watson 	<ul style="list-style-type: none"> • Social bots involved in astroturfing on Twitter • Social bots on the cheater dating site Ashley Madison

Table 1: Table taken from Tsvetkova et al. categorising online bots. This broad schema includes the types of bot which will be discussed below, namely crawler bots, social bots and chatbots.⁵⁰

2.2 The History of Bots and the Web

2.2.1 Crawler Bots and the Early Years of the Web

Although Google is now the Web’s dominant search engine, it was not the first (as is often assumed) and was launched some seven years after the first

⁴⁶ Gorwa and Guilbeault, ‘Unpacking the Social Media Bot’.

⁴⁷ Gregory Maus, ‘A Typology of Socialbots (Abbrev.)’, in *Proceedings of the 2017 ACM on Web Science Conference - WebSci ’17* (presented at the the 2017 ACM, ACM Press, 2017), pp. 399–400, doi:10.1145/3091478.3098860.

⁴⁸ Milena Tsvetkova and others, ‘Even Good Bots Fight: The Case of Wikipedia’, *PLOS ONE*, 12.2 (2017), p. e0171774, doi:10.1371/journal.pone.0171774.

⁴⁹ Tsvetkova and others, ‘Even good bots fight’.

⁵⁰ ibid.

programs to try their hand at searching the Web. Between 1990 and 1993 three programs took up the mantle of being web search pioneers: ‘Archie’, ‘Veronica’ and ‘Jughead’⁵¹. All three were relatively simple, collecting filenames or titles from the few hundred webpages that existed at the time and storing them in a database that could be searched. The limitation here was that most of the text on a page was not stored, so a user looking for a page about ‘baseball’ would see only pages where ‘baseball’ featured in titles, potentially missing many relevant pages.

As the Web grew, storage space on servers increased and finding information was becoming ever more important, for which a broader approach was needed. A series of developments throughout 1993 culminated in the release of the first proper search engines; chief amongst these developments was the creation of the Web’s first bot – the World Wide Web Wanderer⁵². This bot was also the first of what would become known as ‘Web crawlers’, bots that ‘crawl’ from page to page recording their contents in the process. The Wanderer’s primary aim was to measure the size of the Web and this required it to visit every single webpage, which it then recorded and added to an index (a list of words found on each page) called ‘Wandex’, short for Wanderer Index⁵³. By mid-1994, the first recognisably ‘modern’ search engine, called ‘WebCrawler’, had been created. Not only was WebCrawler able to store *all* the text written on a webpage, rather than just the titles and headings, it was also the first to run its own crawler bot, indexing system and search form which was available online for people to use. With data storage capabilities improving and the number of websites and Web users dramatically increasing, WebCrawler was just the beginning of the search engine explosion.

All of this search activity on a growing Web saw an increasingly large army of crawler bots wandering from page to page, diligently recording everything they found and, in the process, sometimes causing trouble or finding themselves in areas where they were not welcome. Website administrators became

⁵¹ Tom Seymour, Dean Frantsvog, and Satheesh Kumar, ‘History Of Search Engines’, *International Journal of Management & Information Systems (IJMIS)*, 15.4 (2011), pp. 47–58, doi:10.19030/ijmis.v15i4.5799.

⁵² ibid.

⁵³ ibid.

increasingly irritated by the constant visits of bots that use up precious bandwidth, accidentally spam the website with unwanted requests to view pages or try and record temporary information which was of no use for searching text content⁵⁴. To counter this, a ‘Standard for Robot Exclusion’ (more commonly referred to as ‘robots.txt’) was agreed in 1994 that allowed administrators to put a file called robots.txt on the web server that would tell robots which parts of the website, if any, they were allowed to visit⁵⁵. Some sites initiated blanket bans, others were more specific and allowed some crawling to occur, recognising the benefits of being indexed by a search engine. However, these instructions were very much a gentleman’s agreement between website administrator and robot creator: social etiquette and politeness persuaded the robots to play by the rules, but there was nothing to stop the bots’ operators from programming their crawlers to simply ignore the new agreements and behave unscrupulously. The leading search engines of the time, including WebCrawler, made their robots adhere to the new rules, but inexperienced or inconsiderate bot operators (especially those spreading malicious code) paid no attention whatsoever⁵⁶. Twenty-five years on and that situation has not changed; there is still no official regulation of crawler bots with interesting consequences, as we shall see later.

Ranking search results in some order so that the user gets the seemingly most important, or relevant, webpages appearing at the top of the list had always been an important consideration for search engines. Having found plenty of pages that contain the search phrase, the simplest solutions then relied upon frequency statistics to count how many times the search phrase appeared on a page, whilst increasingly sophisticated algorithms also looked for other words closely related to the search term (e.g. ‘sheep’ and ‘shepherd’, ‘London’ and ‘Big Ben’) and measured the overall relevance based on such associations. The

⁵⁴ ‘The Web Robots Pages’ <<http://www.robotstxt.org/orig.html>> [accessed 11 April 2019].

⁵⁵ ibid.

⁵⁶ ibid.

more often the search word was counted, or the more relevant a series of words were across the entire page, the higher up the ranking it featured.⁵⁷

Google's ranking algorithm was starkly different because it determined precedence primarily on how many links there were to other pages on a given webpage (and how many links these pages had to others etc.), a measurement that had nothing to do with the search term itself. The algorithm, still running today in an altered form, is called PageRank and its aim of 'bringing order to the Web'⁵⁸, by measuring the relevance of search results through the quality of pages and links to other pages, has proven highly popular. Based upon research into citations between academic papers, this measurement appealed to the way that people subjectively surfed the web from page to page in everyday situations and it was this realisation that relevance was not tied to just one page that made Google's algorithm so much more successful than those based purely on search term statistics⁵⁹.

By 1998, when a new search engine called Google officially went live on the Web, the market was already overcrowded. Two years earlier the leading Web browser, Netscape, had offered an exclusive deal for a single search engine to be featured on its homepage; scores of competitors vied for this coveted spot and in the end a deal was done where five would be featured in rotation throughout the year⁶⁰. This shared prize cost each search engine company \$5M, a figure that paled into insignificance when compared to the eye-watering sums invested in such tech companies during the dot-com bubble of the mid-to-late 90's. Of all the competitors, and one of the five chosen, only one such company survived the subsequent dot-com collapse and continues to be a popular search engine today: Yahoo!. Although many of the mid-90's search companies perished when the dot-com bubble burst at the turn of the 21st century, the newcomer – Google – soon began to flourish. The root of Google's success was being able to perform more useful searches, something not

⁵⁷ Sergey Brin and Lawrence Page, 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', *Computer Networks and ISDN Systems*, 30.1–7 (1998), pp. 107–17, doi:10.1016/S0169-7552(98)00110-X.

⁵⁸ Brin and Page, 'The anatomy of a large-scale hypertextual Web search engine'. 'The Anatomy of a Large-Scale Hypertextual Web Search Engine'.

⁵⁹ *ibid.*

⁶⁰ Seymour, Frantsvog, and Kumar, 'History Of Search Engines'.

enabled by having more bots or larger indexes (although this helps), but by implementing an algorithm that took a different approach to ordering search results.

Back in March 1997, before Google was available as a demo on the Stanford University website⁶¹, the inaugural ‘Museums and the Web’ conference took place in Los Angeles, California. By this time the growth of the Web and the improving ability to find information on it had already seen a number of GLAM institutions, especially museums in the USA and UK, create their own websites attracting an increasing number of online visitors⁶². Although the browser technologies of the time limited how collections could be displayed, and existing database technology determined how they could be searched, a more central sociotechnical quandary arose from these early efforts: what exactly was the role of a museum website – what did it mean to be a ‘virtual museum’⁶³?

A particularly insightful paper presented at this conference by Kevin Donovan outlines the opportunities and challenges around the ‘virtual museum’ hype. As he put it: ‘The question isn’t What?, it’s So What?’⁶⁴. He pointed out that whilst Web search was attracting virtual visitors to museums’ website, the information available there was often little different to that found in the collection database, if the poor user could manage to find it in the first place given the clunky database search tools, and there was a lack of engaging, narrative content about objects or collections on these websites – ‘museums add so little value to the data they provide’⁶⁵. These discussions remain pertinent today and indeed can be seen running throughout the various examples of automated and AI-mediated interactions with online GLAM collections explored in the rest of the

⁶¹ First Versions, ‘Google’, *First Versions*, 2023 <<https://www.firstversions.com/2015/02/google-search-engine.html>> [accessed 3 December 2023].

⁶² Giuliano Gaia and others, ‘Museum Websites of the First Wave: The Rise of the Virtual Museum’ (presented at the Proceedings of EVA London 2020, BCS Learning & Development, 2020), pp. 24–31, doi:10.14236/ewic/EVA2020.4.

⁶³ Gaia and others, ‘Museum Websites of the First Wave’.

⁶⁴ Kevin Donovan, ‘The Best of Intentions: Public Access, the Web & the Evolution of Museum Automation’, 1997

<https://www.museumsandtheweb.com/mw97/speak/donovan.html?_gl=1*c718qy*_ga*MTEwNDMzMjIwOC4xNjgzODCxNjgz*_ga_4799M3BP4Z*MTcwMTM3OTM0My4xLjEuMTcwMTM3OTcwNi4wLjAuMA..&_ga=2.85551641.1065217541.1701379344-1104330208.1683871693> [accessed 30 November 2023].

⁶⁵ ibid.

thesis. They also became all the more pressing for those GLAM professionals back in the late 90's once Google's improved search engine began directing even more visitors to museum websites.

Searching the Web was the first, and arguably still the most influential, case which demonstrates the power of combining bots and algorithms to influence the social and economic connectivity enabled by the Internet. The role of crawlers, the automated bots traversing the Web and recording its contents, usually within the socially acceptable limits defined by the Standard for Robot Exclusion, has since diversified from storing information on the Web so that people can search it to collecting and archiving the Web itself as a historical record and resource. This is one of the main goals of the Internet Archive, an organisation that may come to surpass Google in its ability to influence what we know and believe about past, present and future.

Founded in 1995 by Brewster Kahle, the Internet Archive started out with the long-term goal of enabling "universal access to all knowledge, within our lifetime"⁶⁶, a task which has been taken up (at least in the popular imagination) by Wikipedia and the more recent WikiData. In the early years of the Web, webpages were often short-lived and broken links were a common problem, just as they often are today, therefore the Internet Archive set about recording everything on as many webpages as possible, much like a search engine did, initially using a crawler bot called Alexa (now Alexa Internet, a web traffic analytics company owned by Amazon, although no relation to their personal assistant)⁶⁷.

This reliance on crawling meant that the Internet Archive also decided to play by the rules of the 'robots.txt' Standard for Robot Exclusion which became widely accepted from 1997 onwards, meaning that some websites would now go unarchived in the interests of politeness and digital harmony. Twenty years later, in 2017, when far more websites were blocking all robot visitors outright, the Archive decided to break its gentleman's agreement with the Web and give

⁶⁶ M Kimpton and J Ubois, 'Year by Year: From an Archive of the Internet to an Archive on the Internet', in *Web Archiving*, by J Masanes (Springer-Verlag Berlin Heidelberg, 2006), p. 201 <<http://dx.doi.org/10.1007/978-3-540-46332-0>> [accessed 23 April 2019].

⁶⁷ *ibid.*

its crawlers free reign in the name of preserving as much Web history as possible⁶⁸. Web crawlers are just as relevant as ever given their role in collecting vast quantities of data from across the Web that is used to ‘train’ AI models like those underpinning ChatGPT, and the robots.txt protocol remains one of the few ways in which organisations can resist such automated ‘data scraping’ (read more on this at the end of this chapter).

Come 2001, when the dot-com bubble, the economic crisis caused by massive speculation in online companies that could not achieve the dreams peddled in the new technological age of the Web⁶⁹, had well and truly burst, the Internet Archive decided that it needed an easily accessible platform through which to offer all the content collected over the previous five years, a huge amount (more than 43TB) that included webpages, forums, books, images, tv shows, movies and music tracks, for public browsing and consumption. Hence, in October 2001 the ‘Wayback Machine’ was launched, just a few months after Wikipedia came into being, allowing any user to enter a website URL and see any (or all) archived versions of it⁷⁰. The following year, a physical mirror location (a copy of all the servers storing the Archive’s vast amount of information) was built in the Bibliotheca Alexandrina in Alexandria, Egypt - the modern-day successor to the fabled Great Library of Alexandria from antiquity.

Since its initial success, the Internet Archive has grown to a staggering 30+ Petabytes (30,000,000 GB) containing billions of webpages and millions of books, films, videos, images, audio recordings and software programs⁷¹. The Wayback Machine has become an important tool for historical and social research⁷², but more significantly the combination of the Archive’s crawler bots, linked to bots on other platforms like Wikipedia, has enabled the Archive to

⁶⁸ ‘Robots.Txt Meant for Search Engines Don’t Work Well for Web Archives | Internet Archive Blogs’ <<https://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/>> [accessed 23 April 2019].

⁶⁹ Debra Howcroft, ‘After the Goldrush: Deconstructing the Myths of the Dot.Com Market’, *Journal of Information Technology*, 16.4 (2001), pp. 195–204, doi:10.1080/02683960110100418.

⁷⁰ Kimpton and Ubois, ‘Year by Year: From an Archive of the Internet to an Archive on the Internet’.

⁷¹ ‘Internet Archive: About IA’ <<https://archive.org/about/>> [accessed 23 April 2019].

⁷² Anat Ben-David and Hugo Huurdeman, ‘Web Archive Search as Research: Methodological and Theoretical Implications’, *Alexandria*, 25.1–2 (2014), pp. 93–111, doi:10.7227/ALX.0022.

restore more than nine-million broken links on Wikipedia, mostly citations, so that many online sources of knowledge can be read and checked again⁷³. This is significant for public interaction with much academic information, including that relating to the past. Similarly, WikiData works closely with the Internet Archive to try and mend broken links to historical and cultural items in online museum collections, from paintings to manuscripts, which are often accidentally condemned to digital purgatory when a website is redeveloped⁷⁴.

Therefore, as much as bots are paramount in finding historical information amidst the tangled Web, their role in saving that tied up in strands that are accidentally cut is even more significant for researchers and the public alike. However, these bots act behind the scenes of the Web, collecting and curating, whereas the rise of social media has offered up many new opportunities for new types of bots to be involved in disseminating information in a very public manner, sometimes almost 'face to face'.

2.2.2 Social Bots and the Web 2.0 Era

The rise of human impersonation of bots over the past few years brings our first taxonomic category of bots to the fore: the 'social bots'. There is an important distinction between 'socialbot' and 'social bot' in academic and policy literature⁷⁵: the former is a term used primarily in cybersecurity to describe a program which infiltrates a network of human users for malicious purposes⁷⁶; the latter describes the bots of interest to this section, autonomous users of social media platforms that interact with humans, more strictly defined as

⁷³ 'More than 9 Million Broken Links on Wikipedia Are Now Rescued | Internet Archive Blogs' <<https://blog.archive.org/2018/10/01/more-than-9-million-broken-links-on-wikipedia-are-now-rescued/>> [accessed 23 April 2019].

⁷⁴ 'Wikidata:WikiProject Sum of All Paintings - Wikidata' <https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings> [accessed 16 April 2019].

⁷⁵ Gorwa and Guilbeault, 'Unpacking the Social Media Bot'.

⁷⁶ Yazan Boshmaf and others, 'The Socialbot Network: When Bots Socialize for Fame and Money', in *Proceedings of the 27th Annual Computer Security Applications Conference on - ACSAC '11* (presented at the the 27th Annual Computer Security Applications Conference, ACM Press, 2011), p. 93, doi:10.1145/2076732.2076746.

“automated social agents”⁷⁷ that often try to mimic humans. It is these social bots, recently described as “a new organism in the human social ecosystem”⁷⁸, that most members of the public, and the media, would refer to when talking about bots. If social bots are seen as a new kind of automated organism, then social media has come to represent their densely populated habitat. Understanding of an organism requires an appreciation of its habitat today and how that habitat has changed over time. This is particularly important given the increasing importance of social media as a platform for the dissemination of history and cultural heritage, both by humans and bots⁷⁹.

2.2.2.1 Historical Background of Web 2.0 and Social Bots

‘Social media’ describes both a concept and a set of technologies, defined by Kaplan and Haenlein as a “group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content.”⁸⁰ This definition highlights two phenomena of historical importance for social media and social bots: ‘Web 2.0’ and ‘user-generated content’. This requires a brief diversion into the history of the developing World Wide Web, a linked collection of information available over the Internet, and an exploration of social media.

As a piece of infrastructure, the Internet has been described as an unparalleled digital tool for innovation and producing the unexpected⁸¹. Invented by Tim Berners-Lee in 1989, the World Wide Web was certainly unexpected. The ‘Web’, as it is commonly known, consists of all the documents made available

⁷⁷ Norah Abokhodair, Daisy Yoo, and David W. McDonald, ‘Dissecting a Social Botnet: Growth, Content and Influence in Twitter’, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW ’15 (ACM, 2015), pp. 839–51, doi:10.1145/2675133.2675208.

⁷⁸ Maus, ‘A Typology of Socialbots (Abbrev.)’.

⁷⁹ Kirsten Drotner and Kim Christian Schrøder, *Museum Communication and Social Media: The Connected Museum* (Routledge, 2014).

⁸⁰ Andreas M. Kaplan and Michael Haenlein, ‘Users of the World, Unite! The Challenges and Opportunities of Social Media’, *Business Horizons*, 53.1 (2010), pp. 59–68, doi:10.1016/j.bushor.2009.09.003.

⁸¹ Richard S. Whitt and Stephen Schultze, *The New ‘Emergence Economics’ of Innovation and Growth, and What It Means for Communications Policy* (Social Science Research Network, 1 September 2008) <<https://papers.ssrn.com/abstract=1311904>> [accessed 28 February 2019].

(or ‘served’) by computers (‘servers’) connected to the underlying infrastructure of the Internet. Web browsers, programs run on a computer connected to the net, make all the linked documents of the World Wide Web accessible to a user. Up until the start of the 21st century, the Web was primarily a space in which users could only read the content of webpages but had no opportunity to alter or add to it. Given the Web’s origin, as a tool to link together all academic research documents on CERN’s internal computer network so that they were easily accessible, this is understandable: this was content to be viewed, not rewritten. However, as the Web branched out to a wider public who wanted to share and peruse information for all kind of purposes, the lack of interaction available became a source of irritation. With hindsight, this became known as the ‘Web 1.0’ era, when webpages were largely ‘static’ entities meant for viewing only.

In 2004, fifteen years after the Web had been invented, a new phrase began to permeate into the consciousness of those in the Internet industry: ‘Web 2.0’. That year, O’Reilly Media, known for being one the leading predictors of the future of the tech industry, hosted a conference entitled ‘Web 2.0’ (the first of many which continue to this day) to explore a new phenomenon described by some as ‘the Social Web’.⁸²

The preceding years had seen the Internet industry boom during the ‘dot-com bubble’ which burst in 2001, bankrupting several high-profile ‘dot.com’ companies and leading to the inevitable conclusion that the Web could not herald the technologic and economic revolution that many had hoped, and expected, to occur.⁸³ During the boom of the mid-to-late 1990’s, the commercial interest in everything ‘Web-based’ had seen large-scale investments in improving the physical infrastructure of the Internet in the developed world, giving many households access to the Internet via broadband rather than dial-up connections. This increase in bandwidth mean that far more information could be transmitted to and from servers and users’ PCs without taking a prohibitive amount of time or overloading the server.

⁸² Tim O'Reilly, 'What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software', 65, 2007, p. 22.

⁸³ Howcroft, 'After the goldrush'.

Simultaneously, innovations in web browser programming, such as the invention of the Javascript programming language and Ajax, a set of techniques that allowed web browsers, via Javascript, to update and change a webpage ‘dynamically’ in real-time, made it possible to make webpages that users could fully interact with and change: ‘read/write’ pages rather than ‘read-only’.⁸⁴ During the 90’s these possibilities had been tentatively explored by individuals like Ward Cunningham, who designed a type of website he named a ‘Wiki’ (the Hawaiian for ‘quick’) which allows multiple users to collaboratively edit and modify the site’s content and structure from a web browser.⁸⁵ With such technical advances having been made, by 2004 a series of interactive websites had emerged that relied upon users creating the sites’ content, either collaboratively or individually. The prime example, launched in 2001 and based upon Cunningham’s software, was Wikipedia: a free online encyclopaedia that anyone could edit, add to and read. Wikipedia was the epitome of ‘Web 2.0’, a combination of technology, ideology and social interaction of users that was not previously possible on any significant scale in the Web 1.0 era.

Wikipedia underwent significant growth in its early years, thanks in no short measure to the work of bots that created thousands of articles from databases of information, for instance directories of American towns⁸⁶. Over time, bots gradually became an integral part of the Wikipedia community, as much a social construct as a technological one (although this does not make them ‘social’ bots, as they do not seek to mimic human editors, merely assist them). The number of human editors has increased as the community has grown and even today bots represent a tiny proportion of all editors on Wikipedia, despite being responsible for the vast majority of edits⁸⁷. Recent studies have investigated the

⁸⁴ O'Reilly, 'What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software'.

⁸⁵ Bo Leuf and Ward Cunningham, *The Wiki Way: Quick Collaboration on the Web* (Addison-Wesley Professional, 2001) <<http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/020171499X>> [accessed 11 February 2019].

⁸⁶ Andrew Lih, 'Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Resource', *5th International Symposium on Online Journalism*, 2004, pp. 1–31.

⁸⁷ Tsvetkova and others, 'Even good bots fight'.

ways in which humans and bots edit⁸⁸, how they interact as a community⁸⁹, and how the Wikipedia ecosystem interacts with bots and users on other social channels⁹⁰. However, more work is needed to fully understand the role of bots in helping to shape the interpretations of historical fact that Wikipedia presents to millions of users, along with the factual historical data accessed by humans and bots through WikiData⁹¹. This is especially important given the fact that all the current AI models like ChatGPT have been trained on Wikipedia's content, and that many automated bots retrieving and sharing knowledge about the past get this content from Wikipedia. Yet, the perception of Wikipedia as an objective encyclopaedia, a repository of facts unfettered by politics and subjective interpretations, is a dangerous fallacy; the realities are far more complex and wrapped up in the sociotechnical interactions between human and bot editors⁹².

Wikipedia, however, is not what is usually thought of as a 'social media' platform. Yes, it allows users to collaborate to produce and edit content, but not to interact with each other in the process, or produce such content for the sole purpose of interaction with other users. One of the by-products of so much collaborative, community-driven editing was the concept of 'folksonomies'⁹³, the collections of various keywords and tags used by different editors on Wikipedia which represented a diverse, inherently social type of knowledge base quite distinct from the rigid, authoritative ones, such as library categorisation

⁸⁸ Thomas Steiner, 'Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata', in *Proceedings of The International Symposium on Open Collaboration, OpenSym '14* (ACM, 2014), p. 25:1-25:7, doi:10.1145/2641580.2641613; Maxime Clément and Matthieu J. Guitton, 'Interacting with Bots Online: Users' Reactions to Actions of Automated Programs in Wikipedia', *Computers in Human Behavior*, 50 (2015), pp. 66–75, doi:10.1016/j.chb.2015.03.078.

⁸⁹ Tsvetkova and others, 'Even good bots fight'.

⁹⁰ H. Ford, E. Dubois, and C. Puschmann, 'Keeping Ottawa Honest—One Tweet at a Time? Politicians, Journalists, Wikipedians and Their Twitter Bots', ed. by P. Howard and S. Woolley, *International Journal of Communication*, 10 (2016), pp. 4891–4914.

⁹¹ Denny Vrandečić and Markus Krötzsch, 'Wikidata: A Free Collaborative Knowledgebase', *Commun. ACM*, 57.10 (2014), pp. 78–85, doi:10.1145/2629489; Alessandro Piscopo, 'Wikidata: A New Paradigm of Human-Bot Collaboration?', *arXiv:1810.00931 [Cs]*, 2018 <<http://arxiv.org/abs/1810.00931>> [accessed 11 October 2018].

⁹² Heather Ford, *Writing the Revolution: Wikipedia and the Survival of Facts in the Digital Age* (The MIT Press, 2022); Robin Gieck and others, 'Cultural Differences in the Understanding of History on Wikipedia', in *Designing Networks for Innovation and Improvisation*, ed. by Matthäus P. Zylka and others, Springer Proceedings in Complexity (Springer International Publishing, 2016), pp. 3–12.

⁹³ Hend S. Al-Khalifa and Hugh C. Davis, 'Towards Better Understanding of Folksonomic Patterns', in *Proceedings of the Eighteenth Conference on Hypertext and Hypermedia*, HT '07 (Association for Computing Machinery, 2007), pp. 163–66, doi:10.1145/1286240.1286288.

schemas, which had traditionally been used. Such social technology did not just enable collaboration to pool vast amounts of knowledge for the cultural sector as well as for individuals, it meant that the knowledge itself came to be represented differently. In this vein, the interactive possibilities that enabled the creation of Wikipedia also saw the rise of inherently social sites with the purpose of enabling and promoting interaction between users, whether this was sharing doses of their daily lives or proudly displaying the contents of their camera. ‘Blogger’ (1999) and ‘MySpace’ (2003) were the giants of this new era, but new players were emerging all the time, most notably ‘Flickr’ and ‘Facebook’, both formed in 2004,⁹⁴ and YouTube for video in 2005.

It is worth noting that such social interaction was not a new Internet phenomenon; far from it. In 1979, before the World Wide Web made the Internet accessible for most people, the ‘Usenet’ was created. A worldwide discussion system distributed across connected computers (via the early Internet), it became a popular means of reading and posting text to ‘newsgroups’ dedicated to any number of conceivable topics. Considered the predecessor of Web forums, and ultimately social media communities (for instance Reddit which still uses a newsgroup-like categorical naming for threads), Usenet still operates today but with far fewer users, most having turned to social media. Taking inspiration from users who kept ‘online diaries’ on Usenet, as the Web grew in the 1990’s so did the communities of diarists that could now attract thousands of public users. In 1998, ‘Open Diary’ became the primary site for hosting a person’s diary, or ‘web log’, which morphed into the term ‘we blog’⁹⁵, emphasising the social nature and community spirit central to the activity. By 1999 ‘blogging’ had become a past-time popular enough that competitor sites like Blogger were born.

Social networking sites, the most successful of which was MySpace, sought to offer the same community spirit, initially based around users’ appreciation of

⁹⁴ Jose van Dijck, *The Culture of Connectivity: A Critical History of Social Media* (Oxford University Press, 2013), pp. 6–7 <<http://site.ebrary.com/id/10686674>> [accessed 11 February 2019].

⁹⁵ Paul Bausch and others, *We Blog: Publishing Online with Weblogs* (Wiley, 2002) <<http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0764549626>> [accessed 11 February 2019].

music bands, whilst also offering the chance to connect with other users through customisable profiles. MySpace even attracted a few GLAM institutions, such as Brooklyn Museum which used the social networking space to build an online community and share video content from exhibitions⁹⁶. Facebook offered a similar service, albeit with less customisation, resulting in profiles that were much easier to read and navigate, and presented the user with a 'feed' of their friends' posts and activities. Feeds were at the heart of blogging, allowing users to keep track of their favourite blogs in a single aggregator that would notify them when a blog had been updated. Feeds would come to form the basis of the 'microblogging' platform Twitter (known as 'X' since 2022 following a takeover by Elon Musk)⁹⁷. Feeds were based upon RSS (Really Simple Syndication), an approach also pioneered in the late 90's which allowed blogging, and later microblogging, to become so popular and widespread.⁹⁸

All of this activity, be it collaborative, social or both, relied upon users, both to produce and share the content for their sites and, in turn, to interact with others and their content. This led to the phrase 'User-Generated Content' (UGC) becoming the buzz-word of the media industry and eventually worldwide fame when *Time* magazine named their 2006 Person of the Year as 'You' in response to the massive growth of YouTube formed only a year earlier.⁹⁹ Although this was initially seen as a revolutionary move towards freedom and 'produsers' (creators, users and distributors)¹⁰⁰, and away from traditional media of which users were merely consumers, van Dijck and others rightly criticise the naivety of this position, arguing that consumers had always been active participants in the media ecosystem and that individuals may be able to create

⁹⁶ N Caruth and S Bernstein, 'Building an On-Line Community at the Brooklyn Museum: A Timeline', 2007 <<https://www.archimuse.com/mw2007/papers/caruth/caruth.html>> [accessed 3 December 2023].

⁹⁷ 'Acquisition of Twitter by Elon Musk', *Wikipedia*, 2023 <https://en.wikipedia.org/w/index.php?title=Acquisition_of_Twitter_by_Elon_Musk&oldid=1186289585> [accessed 3 December 2023].

⁹⁸ Judith Wusterman, 'RSS: The Latest Feed', *Library Hi Tech*, 22.4 (2004), pp. 404–13, doi:10.1108/07378830410570511.

⁹⁹ José van Dijck, 'Users like You? Theorizing Agency in User-Generated Content', *Media, Culture & Society*, 31.1 (2009), pp. 41–58, doi:10.1177/0163443708098245.

¹⁰⁰ Axel Bruns, 'The Future Is User-Led: The Path towards Widespread Produsage', in *Fibreculture Journal* (presented at the PerthDACP: Digital Arts & Culture, Fibreculture Publications, 2008) <http://journal.fibreculture.org/issue11/issue11_bruns.html> [accessed 11 February 2019].

and share their own content as part of a community, but to say that everyone benefits from such a community is presumptuous at best.¹⁰¹ Expressed in 2009, when social media was an embedded phenomenon with enough traction to attract the attention of businesses keen to make commercial use of these new networked communities of millions of potential customers¹⁰², such views seem to predict the rise of the current social media buzz-word: influence.

In his seminal work of 2006, Yochai Benkler put forward the economic idea that the social media, with its basis of user-generated content, was effectively a “networked public sphere”¹⁰³ that would enable a new area of information and creative production outside of any market influence. During the early years of Web 2.0, from 2000-2005, this had indeed been the case, with users creating and sharing information, engaging with social media and becoming members of online social networks without any thought of commercial gain or profit. That, after all, was not the point of such online spaces. However, as more users flooded in and the networks grew unfathomably large, two important changes occurred. Firstly, the small organisations or companies running social networking sites, like Facebook, realised that in order to maintain their service, they needed to invest large sums of money to deal with the increasing load on their servers and staff. This meant developing a business model based around advertising, and beginning to think commercially, rather than just ideologically.¹⁰⁴ The real value would turn out to lie in users’ data, which would become a more controversial and far more lucrative resource.

At the same time, such huge numbers of users had not escaped the attention of big business, which was finally waking up to the fact that this ‘networked public sphere’ was the perfect place to influence people, and their networks of friends, towards their products and brands.¹⁰⁵ Hence ‘social media marketing’ was born; but this need not only be a pursuit of major corporate entities, individuals could

¹⁰¹ van Dijck, ‘Users like you?’

¹⁰² Kaplan and Haenlein, ‘Users of the world, unite! The challenges and opportunities of Social Media’.

¹⁰³ Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press, 2006), p. 10.

¹⁰⁴ van Dijck, ‘The culture of connectivity’, pp. 15–16.

¹⁰⁵ Andrew Lipsman and others, ‘The Power of “Like”: How Brands Reach (and Influence) Fans Through Social-Media Marketing’, *Journal of Advertising Research*, 52.1 (2012), pp. 40–52, doi:10.2501/JAR-52-1-040-052.

make the most of their ability to create, share and distribute content in a one-person marketing machine. The democratising, liberating ideologies that had inspired the creation of platforms like Facebook and Twitter now made it possible, in theory, for any user to wield significant power and influence over the network.¹⁰⁶ Of course, such 'influencers' remain a tiny minority, but the opportunity to reach their heights is still open to anyone.

The importance of the 'social' within social networks and their platforms has become only too apparent with the rise of Web 2.0. The last decade has seen a shift from 'produsers' creating and sharing content to 'produsers' creating and sharing influence. Influence lies in the network itself, in the connections between users and their networks, the maintenance of which now relies upon content. This has come to form the underlying economics of these platforms as places where user-generated content has become central to the advertising business model, the value of which still relies on the content itself (and the algorithms that make it visible, see more later on) but is increasingly determined by the role of a community around that content. This shift has become more pronounced recently, being touted as a move from an 'attention economy' of users having content to a 'creator economy' of building communities to engage with that content, something of particular relevance to GLAM institutions that are now found across social media platforms¹⁰⁷.

The tendency for similar people to form social groups, a phenomenon known as homophily, is seemingly magnified by social network sites¹⁰⁸, resulting in echo chambers that discourage information diversity¹⁰⁹. This means that although the

¹⁰⁶ Karen Freberg and others, 'Who Are the Social Media Influencers? A Study of Public Perceptions of Personality', *Public Relations Review*, 37.1 (2011), pp. 90–92, doi:10.1016/j.pubrev.2010.11.001.

¹⁰⁷ Clara Lindh Bergendorff, 'From The Attention Economy To The Creator Economy: A Paradigm Shift', *Forbes*, 2021 <<https://www.forbes.com/sites/claralindhbergendorff/2021/03/12/from-the-attention-economy-to-the-creator-economy-a-paradigm-shift/>> [accessed 3 December 2023]; Alexandra Weilenmann, Thomas Hillman, and Beata Jungselius, 'Instagram at the Museum: Communicating the Museum Experience Through Social Photo Sharing', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13 (ACM, 2013), pp. 1843–52, doi:10.1145/2470654.2466243.

¹⁰⁸ Mike Thelwall, 'Homophily in MySpace', *Journal of the American Society for Information Science and Technology*, 60.2 (2009), pp. 219–31, doi:10.1002/asi.20978.

¹⁰⁹ Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson, 'Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data', *Journal of Communication*, 64.2 (2014), pp. 317–32, doi:10.1111/jcom.12084.

content produced still has to serve a given community, the emphasis is more on quantity and regularity than quality so long as it feeds the social and cultural expectations of the group. An influencer need not even produce much of their content when so much is available to share and reshare, especially on platforms like Twitter where quoting and ‘retweeting’ are a key part of influence production. The same is true of image-based microblogging sites like Flickr and Instagram. In this ecosystem a new type of organism has emerged, one designed to fit perfectly into an evolving and ever-expanding network of content and influence: the automated user, or social bot. Whilst GLAMs may have begun to embrace social media platforms as ways of engaging audiences and building communities, these were nearly all human-curated and communicated endeavours and of the few social bots that do exist in these contexts only a handful are officially linked to and managed by GLAM institutions themselves.

To say that social bots are an entirely new phenomenon of the present social media age would be to overlook their own historical background. Indeed, the first wave of Internet bots included the first social bots, which could be found on Internet Relay Chat (IRC) as early as 1989. IRC is an Internet-based, geographically distributed chat service that is its own Internet protocol separate from HTTP, the protocol of the World Wide Web.¹¹⁰ Computers connected using IRC form an IRC network that lets users chat about topics organised into ‘channels’, similar to the way in which Usenet users organised content into newsgroups. Within the context of IRC, a bot was a program capable of signing on to the network itself and interacting with other users (or bots); scripts were also often used on IRC networks, the main difference between bot and a script being that a bot could log itself on, whereas a user would have to start a script, after which it could run automatically.¹¹¹

One of the first, most popular and longest-lasting IRC bots is called ‘Eggdrop¹¹²’ which functions as a ‘channel guard’, a way of preventing users from hijacking a certain channel. Channels were controlled by certain users, who determined the rules for the conversation around the given topic, but this control lasted only as

¹¹⁰ Latzko-Toth, ‘The Socialization of Early Internet Bots’.

¹¹¹ ibid.

¹¹² ‘Eggheads’ <<https://www.eggheads.org/>> [accessed 13 February 2019].

long as a user was logged-in.¹¹³ Other users could then try and takeover said channel, claiming it for themselves and potentially changing the rules, or the topic of conversation, to the annoyance of the channel's current devotees. Eggdrop, still popular on IRC today, is designed to automatically manage channels to prevent such takeovers, kick out users flouting the rules and generally make the channel controller's life as easy as possible. Eggdrop is social in that it could interact with users, albeit in a highly scripted way, and the fact that it is designed to try and reduce anti-social behaviour has doubtless helped its apparent sociality. Just as the outward appearance of automation can make an automated agent a bot, so a social demeanour can make a bot 'social'.

Latzko-Toth approaches the issues of sociality, like many others in the fields of social technology who have turned their gaze towards bots, through the lens of Actor Network Theory (ANT) which describes a network of 'actants', human and non-human, whose interactions form their sociality, rather than being an indication of its prior existence within the actors themselves.¹¹⁴ These links also constitute the agencies present in the network, rather than ascribing them to actants or presuming their omnipotence. In the context of IRC, Latzko-Toth argues that bots' sociality, as non-human actants, comes from their interactions with human and non-human (bot) users.¹¹⁵

Sociologists and communication theorists have long argued about whether social bots have agency, or rather certain kinds of agency (e.g. moral¹¹⁶, authorial¹¹⁷), have no agency whatsoever and are merely engaged in the agency of a human user¹¹⁸, or whether the social networks which they inhabit may give them agency as perceived by human users¹¹⁹. The last of these positions comes

¹¹³ Latzko-Toth, 'The Socialization of Early Internet Bots'.

¹¹⁴ Latour, 'Reassembling the Social'.

¹¹⁵ Latzko-Toth, 'The Socialization of Early Internet Bots'.

¹¹⁶ Luciano Floridi, *The Cambridge Handbook of Information and Computer Ethics* (Cambridge University Press, 2010), pp. 208–9.

¹¹⁷ Krista A Kennedy, 'Textual Curators and Writing Machines: Authorial Agency in Encyclopedias, Print to Digital' (University of Minnesota, 2009).

¹¹⁸ c.f. Douglas Guilbeault, 'Automation, Algorithms, and Politics| Growing Bot Security: An Ecological View of Bot Agency', *International Journal of Communication*, 10.0 (2016), 19; c.f. Gina Neff and Peter Nagy, 'Automation, Algorithms, and Politics| Talking to Bots: Symbiotic Agency and the Case of Tay', *International Journal of Communication*, 10.0 (2016), 17.

¹¹⁹ Oliver Leistert, 'Social Bots as Algorithmic Pirates and Messengers of Techno-Environmental Agency', in *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*, ed. by Jonathan Roberge and Robert Seyfert (Taylor & Francis, 2016).

closest to that methodologically describe by ANT, often employed to argue that societal and cultural separations of human and technology are unhelpful illusions¹²⁰, but which itself has been criticised for treating humans (perhaps users and programmers) and non-humans (software bots or even physical computers) on an equal footing in terms of agency¹²¹. Across the literature, despite the details debated, the consensus on what represents sociality when considering bots boils down to adopting a pragmatic approach. In essence, however ‘social’ comes to be defined and understood, these bots acting within social networks and on social platforms will also, in some way, be social. To borrow from Gorwa and Guilbeault’s framework¹²², we need to ask not what makes bots social, but how and why they are social and what their sociality achieves.

Stieglitz¹²³ attempts to answer these questions through further classification of social media bots. As noted from the discussions above, bot sociality and agency are concepts tangled up in the literal and theoretical ‘networks’ which they inhabit, but ‘intentionality’, what a bot is trying to do, seems to have a more concrete and direct association with its creator(s). The varying definitions of a bot all stipulate that bots have ‘goals’¹²⁴, necessarily determined by programmers (at least for now). Stieglitz, therefore, attempts an initial classification by bot/creator intent, focussing on bots which hold social media accounts rather than bots, or botnets, which often use social media as a covert method of communication. Although most commonly discussed in terms of malicious activity, a botnet is simply a network of bots; in rare examples of benign activity, this might better be described as a ‘community’ of linked bots. This framework serves as a useful starting point to examine the various types of

¹²⁰ J. Macgregor Wise, ‘Intelligent Agency’, *Cultural Studies*, 12.3 (1998), pp. 410–28, doi:10.1080/095023898335483.

¹²¹ Frederic Vandenberghe, ‘Reconstructing Humans: A Humanist Critique of Actant-Network Theory’, *Theory, Culture & Society*, 19.5–6 (2002), pp. 51–67, doi:10.1177/026327602761899147.

¹²² Gorwa and Guilbeault, ‘Unpacking the Social Media Bot’.

¹²³ Stefan Stieglitz and others, ‘Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts’, *arXiv:1710.04044 [Cs]*, 2017 <<http://arxiv.org/abs/1710.04044>> [accessed 23 January 2019].

¹²⁴ Michael Wooldridge and Nicholas R. Jennings, ‘Intelligent Agents: Theory and Practice’, *The Knowledge Engineering Review*, 10.2 (1995), pp. 115–52, doi:10.1017/S0269888900008122.

social bots currently active on the Web and their role in this complex sociotechnical system.

2.2.2.2 The Current Roles of Social Bots

Amongst the ‘benign’ (i.e. those with ‘good’ or useful intentions) bots designed to operate on social networks, the vast majority provide information services to users by automatically disseminating content from other sources. These include: news bots, traffic bots, weather bots, sport bots, earthquake bots, job recruitment bots etc. covering a huge array of topics, from the mainstream to the obscure¹²⁵ (for an overview of this rapidly expanding field, see Figure 1). Falling within the same benign remit are a collection of bots that “merely are”¹²⁶ as Stieglitz puts it, with entertainment, education or metaphysical explorations of reality through increasingly nonsensical utterances amongst their collective ‘goals’. Some of the weirder of these harmless bots turn out to be useful for research into computational linguistics and human-computer interaction¹²⁷, the more humorous managing to provide some light relief in the process¹²⁸. One bot, for instance, automatically generates amusing exhibition titles for a fictitious virtual museum¹²⁹.

¹²⁵ Stieglitz and others, ‘Do Social Bots Dream of Electric Sheep?’

¹²⁶ *ibid.*

¹²⁷ Tony Veale, ‘Exploring the Placebo Effect in Computational Creativity’, *Proceedings of the Sixth International Conference on Computational Creativity*, 2015.

¹²⁸ Alex Wilkie, Mike Michael, and Matthew Plummer-Fernandez, ‘Speculative Method and Twitter: Bots, Energy and Three Conceptual Characters’, *The Sociological Review*, 63.1 (2015), pp. 79–101, doi:10.1111/1467-954X.12168.

¹²⁹ ‘MuseumGenerator @MuseumGenerator’ <<https://twitter.com/MuseumGenerator>>.

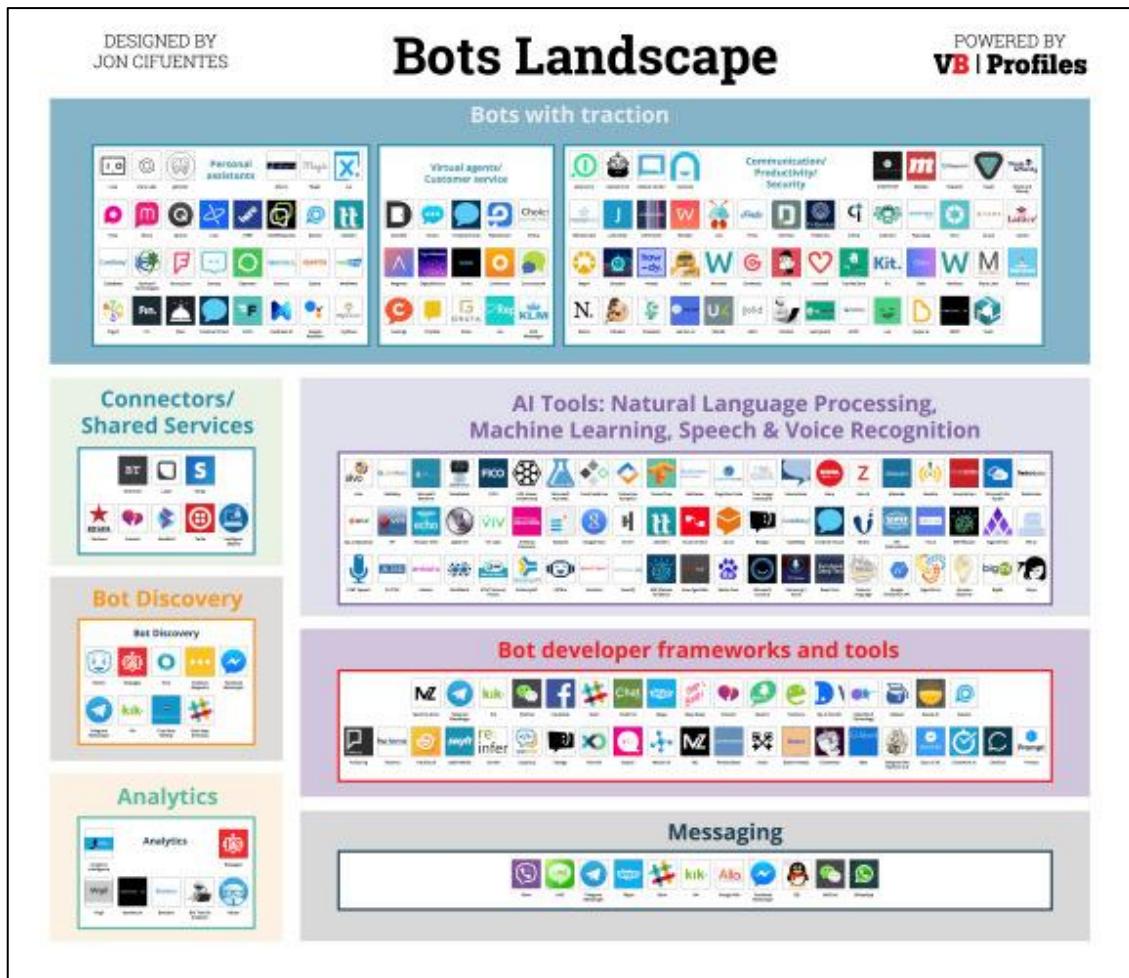


Figure 1: Overview of the already crowded online social bot landscape. Source: venturebeat.com.¹³⁰

Bots seeking to educate or entertain (or both), spanning the ‘good’ and ‘harmless’ categories, combine entertaining content with something of an educational approach (whether by design or not), and are perhaps those most relevant for examining public consumption of history and heritage on social media. An extremely popular example is the collection of ‘art bots’ managed by one user that run across Twitter and the photo-blogging site Tumblr, disseminating artworks from publicly available archive collections on museum websites. This collection currently covers over three-hundred artists, from

¹³⁰ ‘Introducing the Bots Landscape: 170+ Companies, \$4 Billion in Funding, Thousands of Bots | VentureBeat’ <https://venturebeat.com/2016/08/11/introducing-the-bots-landscape-170-companies-4-billion-in-funding-thousands-of-bots/?utm_content=buffereddf6&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer> [accessed 26 May 2020].

Monet¹³¹ to Cezanne¹³², and each bot has hundreds of followers who often thank their creator for providing them with this life-affirming service. Some heritage institutions and archives also have their archives disseminated through ‘unofficial’ bots, disseminating images of everything from items in the DPLA (Digital Public Library of America)¹³³ to Australian newspaper articles from the 1800’s¹³⁴.

The category of malicious (or ‘bad’) social media bots is just as large and diverse. There are overtly criminal bots whose creators seek to spread spam, steal users’ identities, copy content or links for advertising fraud or spread malware¹³⁵, some of which do not even bother to try and mimic human users¹³⁶. Others combine to produce and spread the current phenomenon of ‘fake news’, as a new form of computational propaganda and misinformation¹³⁷. This includes bots that carry out activities like ‘astroturfing’, giving the impression of having fake political user bases back a given opinion¹³⁸, ‘smoke-screening’, flooding a debate with hashtags to distract users from the subject under discussion or to misrepresent its context¹³⁹, and more blatant forms of misdirection, such as posting fake articles with unrelated hashtags or adding mentions for public figures or companies to sway opinion and influence, hence their classification as ‘influence bots’¹⁴⁰.

If social media itself is viewed as a network of influences, as described earlier, especially as bot activities can take place across platforms (or in the case of botnets, ‘through’ social media more generally), then it could be argued that any

¹³¹ ‘Claude Monet @artistmonet’ <<https://twitter.com/artistmonet?lang=en>> [accessed 14 February 2019].

¹³² ‘Paul Cezanne @cezanneart’ <<https://twitter.com/cezanneart?lang=en>> [accessed 14 February 2019].

¹³³ ‘DPLA Bot @DPLABot’ <<https://twitter.com/DPLABot>> [accessed 14 February 2019].

¹³⁴ ‘Trove Goldfields News Bot @TroveGoldfields’ <<https://twitter.com/trovegoldfields>> [accessed 14 February 2019].

¹³⁵ Abokhodair, Yoo, and McDonald, ‘Dissecting a Social Botnet’.

¹³⁶ Stieglitz and others, ‘Do Social Bots Dream of Electric Sheep?’

¹³⁷ Edson C. Tandoc Jr, Zheng Wei Lim, and Richard Ling, ‘Defining “Fake News”’, *Digital Journalism*, 6.2 (2018), pp. 137–53, doi:10.1080/21670811.2017.1360143.

¹³⁸ Jacob Ratkiewicz and others, ‘Truthy: Mapping the Spread of Astroturf in Microblog Streams’, in *Proceedings of the 20th International Conference Companion on World Wide Web - WWW ’11* (presented at the the 20th international conference companion, ACM Press, 2011), p. 249, doi:10.1145/1963192.1963301.

¹³⁹ Abokhodair, Yoo, and McDonald, ‘Dissecting a Social Botnet’.

¹⁴⁰ V.S. Subrahmanian and others, ‘The DARPA Twitter Bot Challenge’, *Computer*, 49.6 (2016), pp. 38–46, doi:10.1109/MC.2016.183.

social media bot is, in some form, using and abusing its and others' influence. Other bots are often deployed to 'fight back' against those which spread fake news by churning out 'facts' of their own¹⁴¹, or by linking to more reputable sources. Savage et al. have even developed a platform, called 'Botivist'¹⁴², that uses Twitter bots to organise volunteers for social activism both online and offline. This leads into the current major area of concern surrounding social media bots: their engagement with politics¹⁴³.

The actions of bots in the political realm have led to the formation of another subcategory of social bots, the 'political bots'¹⁴⁴. These bots have been at the forefront of much media and academic attention since the US Presidential election of 2016¹⁴⁵, with claims of Russian manipulation abounding, and the Brexit referendum of the same year¹⁴⁶, both of which saw coordinated mass campaigns of political 'influencing' and manipulation across social media¹⁴⁷. Often, small numbers of bots, whether coordinated or not, can automatically generate a huge amount of content¹⁴⁸. Social media had been used in political contexts before to dramatic effect, most notably the Arab Spring uprisings of 2011, but bots did not feature in this online activity, which was orchestrated by human users¹⁴⁹. The key point about 'political bots' harks back to the pragmatic approach of identifying the 'social' of social media, social networks and social bots as inherently involving human users. In their typology, Gorwa and

¹⁴¹ Emilio Ferrara and others, 'The Rise of Social Bots', *Communications of the ACM*, 59.7 (2016), pp. 96–104, doi:10.1145/2818717.

¹⁴² Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer, 'Botivist: Calling Volunteers to Action Using Online Bots', in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16 (ACM, 2016), pp. 813–22, doi:10.1145/2818048.2819985.

¹⁴³ Samuel Woolley, 'The Political Economy of Bots: Theory and Method in the Study of Social Automation', in *The Political Economy of Robots*, International Political Economy Series (Palgrave Macmillan, Cham, 2018), pp. 127–55, doi:10.1007/978-3-319-51466-6_7.

¹⁴⁴ Gorwa and Guilbeault, 'Unpacking the Social Media Bot'.

¹⁴⁵ c.f. Howard, Woolley, and Calo, 'Algorithms, bots, and political communication in the US 2016 election'.

¹⁴⁶ c.f. Philip N. Howard and Bence Kollanyi, *Bots, #StrongerIn, and #Brexit: Computational Propaganda During the UK-EU Referendum* (Social Science Research Network, 20 June 2016) <<https://papers.ssrn.com/abstract=2798311>> [accessed 13 August 2018].

¹⁴⁷ Bradshaw and Howard, 'Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation'.

¹⁴⁸ Howard and Kollanyi, 'Bots, #StrongerIn, and #Brexit'.

¹⁴⁹ Philip N. Howard and others, *Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?* (Social Science Research Network, 2011) <<https://papers.ssrn.com/abstract=2595096>> [accessed 13 February 2019].

Guilbeault highlight the role of ‘cyborg accounts’¹⁵⁰, nicely shortened to ‘cybots’ by Lamo and Calo¹⁵¹, as bots under human control or, as discussed at the start of this chapter, human / bot users of (deliberately) ambiguous identity. Whilst cybots can and do exist in almost every subcategory of social ‘bot’, they may be particularly prevalent in the political zone where various semi-automated human/bot ‘creatures’, or networks of such, can exert significant influence¹⁵².

The political motivations behind much of the automated and semi-automated computational propaganda activity suggests that the posts and articles created and disseminated may include content discussing history and heritage, particularly in nationalist contexts. Whilst there is not much evidence for this in the literature discussing social media and bots in the US Elections of 2016, focussed as it is more upon methodological and technical operations of digital political campaigning¹⁵³, there is evidence for public perceptions of their national pasts playing a part in the Brexit debates and the nationalist ‘Internet Hindus’ community in India.

Bonacchi et al examined the role of social media in shaping public and political discourses towards Brexit with relation to the past, namely Britain’s heritage, history and national identity¹⁵⁴. This study showed the close connections between politics and history within online debates and displays numerous examples of history being portrayed from biased perspectives, heritage being misappropriated or historical ‘truths’ being fabricated. Whilst this is nothing new in the field of computational propaganda, the coordinated actions around these campaigns on social media sites, particularly Facebook and Twitter, show the power of users to influence perceptions of the past. Although most of the users in Bonacchi’s study were humans, given that it focused upon discussion in Facebook groups, it is highly likely that some of the other users disseminating

¹⁵⁰ Gorwa and Guilbeault, ‘Unpacking the Social Media Bot’.

¹⁵¹ Madeline Lamo and Ryan Calo, *Regulating Bot Speech* (Social Science Research Network, 16 July 2018) <<https://papers.ssrn.com/abstract=3214572>> [accessed 13 February 2019].

¹⁵² Andree Thielges and others, ‘Effects of Social Bots in the Iran-Debate on Twitter’, *arXiv:1805.10105 [Cs]*, 2018 <<http://arxiv.org/abs/1805.10105>> [accessed 15 August 2018].

¹⁵³ c.f. Howard, Woolley, and Calo, ‘Algorithms, bots, and political communication in the US 2016 election’.

¹⁵⁴ Bonacchi, Altaweel, and Krzyzanska, ‘The heritage of Brexit’.

and linking to these discussions from platforms like Twitter were either fully- or semi-automated bots.

In a similar vein, the ‘Internet Hindus’ community, proponents of Indian nationalism and Hinduism mainly based within India¹⁵⁵, where social media usage has grown exponentially in recent years, often uses far-right, anti-Islamic rhetoric to invent “quasi-historical narratives”¹⁵⁶ of their homeland. They are strongly associated with the Hindu nationalist political party, the Bharatiya Janata Party (BJP), and Indian Prime Minister Narendra Modi, a BJP member. Like other politically motivated, nationalist groups who have formed communities on social media, the majority of their activities are carried and coordinated by human users, some living outside of India¹⁵⁷. However, a small proportion are doubtless automated and semi-automated user accounts engaged in the dissemination of their views and disruption of their opponents’, both of which involve creating biased, single-minded perceptions of national and religious history¹⁵⁸. Other groups use similar online tactics as part of spreading their nationalist agendas, but to what extent these are automated is unclear in many cases¹⁵⁹.

It is clear from both of the examples outlined above that the vast campaigns of politically motivated misinformation, and sometimes the countercampaigns of opposing communities seeking to state ‘facts’, have an effect upon the portrayal of national pasts and communities’ historical identities. However, such content is still only a fraction of the total produced on social media by political humans and political bots. The rise of automation in this area is only likely to increase this proportion and the significant amount of research now being undertaken with regard to computational propaganda and ‘fake news’ may start to provide a clearer picture about what this means for public understanding, or misunderstanding, of multiple pasts. Naturally, the role of the user cannot be

¹⁵⁵ Siram Mohan, ‘Locating the “Internet Hindu”: Political Speech and Performance in Indian Cyberspace’, *Television & New Media*, 16.4 (2015), pp. 339–45, doi:10.1177/1527476415575491.

¹⁵⁶ Mohan, ‘Locating the “Internet Hindu”’. ‘Locating the “Internet Hindu”: Political Speech and Performance in Indian Cyberspace’

¹⁵⁷ Ibid.

¹⁵⁸ Juli L. Gittinger, *Hinduism and Hindu Nationalism Online* (Routledge, 2018).

¹⁵⁹ Christian Fuchs, *Nationalism on the Internet: Critical Theory and Ideology in the Age of Social Media and Fake News* (Routledge, 2019).

overlooked here; messages on social media can easily be interpreted in a way that fits a user's pre-existing set of beliefs, or they may simply choose to ignore or block any sources that share such content. Within the framework of ANT, these users are not passive actants and their interpretation of materials disseminated can be as influential as the content itself. This makes the mediation of historical and cultural heritage information online a complex, two-way process that requires a nuanced knowledge of the potential audiences, both human and non-human, that may be present. For example, it has been shown that users searching specifically for cultural heritage information do not follow typical Google search patterns, instead browsing in a more open-minded manner¹⁶⁰; but the potential audience for content disseminated by social bots is likely to be concentrated on social media where different patterns of information consumption exist¹⁶¹.

The combination of 'humanness' and 'botness' present in the various instances of misinformation networks discussed can be extended to individual bots themselves, where a social bot can appear to be human whilst acting automatically, even though it may receive sporadic, if any, human input. Such an arrangement is common within 'troll farms' that create and coordinate these enormous networks, many of which dissipate once the election or event of interest has passed¹⁶². This gives the impression that these networks, or at least the majority of their automated aspects, can be social 'on demand', a situation which reflects as much upon the organisation of the social media ecosystem as the human and bot actors involved. Many social bots, especially the more benign, will often persist on social media networks for years, interacting in a more recognisably social manner and building up networks of followers in the

¹⁶⁰ Paul Clough and others, 'Europeana: What Users Search for and Why', in *Research and Advanced Technology for Digital Libraries*, ed. by Jaap Kamps and others, Lecture Notes in Computer Science (Springer International Publishing, 2017), pp. 207–19.

¹⁶¹ Sharad Goel, Jake M. Hofman, and M. Irmak Sirer, 'Who Does What on the Web: A Large-Scale Study of Browsing Behavior', in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012 <<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4660>> [accessed 26 May 2020].

¹⁶² Marco T. Bastos and Dan Mercea, 'The Brexit Botnet and User-Generated Hyperpartisan News', *Social Science Computer Review*, 37.1 (2019), pp. 38–54, doi:10.1177/0894439317734157.

process. However, sociality ‘on demand’ is the primary function of another increasingly popular type of bot: the ‘chatbot’.

2.2.3 Chatbots: Fifty Years of History

A ‘chatbot’ is a program designed to “approximate human speech and interact with humans directly through some sort of interface¹⁶³”. Chatbots are common across social messaging platforms, such as Facebook Messenger, and cover many types of AI including Virtual Conversational Assistants (VCAs) like Amazon Alexa. Museums are already experimenting with chatbots as a way of engaging with visitors online and making experiences more interactive once at the museum¹⁶⁴. Alexa hosts a vast array of ‘Skills’, small app-like chatbots, dedicated to history, including numerous quizzes and daily facts for different historical periods and topics (see Chapter 3). This section will explore the history of chatbots before returning to the present day and summarising their current and potential uses in the engagement of online users with information about the past.

Chatbots may appear to be just another type of social bot, and whilst those deployed on social media platforms and messaging apps are often referred to as ‘social’ chat bots, a chatbot need not be inherently social¹⁶⁵. Many chatbots, especially those used in commercial settings, work in isolation and interact only with customers through prescribed live chat boxes on company websites or, like Microsoft Office’s notorious ‘Clippy’ (an incessantly irritating talking paperclip), inside applications¹⁶⁶. However, one of the main properties expected of chatbots, and this has been the case since the very first to be developed in the 1960’s (see below), is embodiment: to “provide the function of presence... Even

¹⁶³ Gorwa and Guilbeault, ‘Unpacking the Social Media Bot’.

¹⁶⁴ Giuliano Gaia, Stefania Boiano, and Ann Borda, ‘Engaging Museum Visitors with AI: The Case of Chatbots’, in *Museums and Digital Culture: New Perspectives and Research*, ed. by Tula Giannini and Jonathan P. Bowen, Springer Series on Cultural Computing (Springer International Publishing, 2019), pp. 309–29, doi:10.1007/978-3-319-97457-6_15.

¹⁶⁵ Gorwa and Guilbeault, ‘Unpacking the Social Media Bot’. ‘Unpacking the Social Media Bot’.

¹⁶⁶ Robert Dale, ‘The Return of the Chatbots’, *Natural Language Engineering*, 22.5 (2016), pp. 811–17, doi:10.1017/S1351324916000243.

the earliest chatbots were given names”¹⁶⁷. Issues of embodiment and human interaction with, and acceptance of, artificial intelligences were just as keenly felt in the 1950’s and 60’s when the industry hype around AI had fed into the public consciousness, often associated with the question of whether artificial minds could be used to describe and explain natural ones¹⁶⁸. It is against this backdrop that the first attempts were made to design a computer program that could converse with a human using natural language.

2.2.3.1 Historical Background of Chatbots

ELIZA, now considered to be the first chatbot, was designed in 1966 and managed to converse so successfully with human users that many believed it a real person¹⁶⁹, thus making it the first computer program potentially capable of passing the Turing Test. ELIZA was trained to imitate the style of a Rogerian psychotherapist, asking the human to explain their feelings in response to statements and asking such questions in response to being asked a question itself¹⁷⁰. This solid conversational foundation was a key part of ELIZA’s success in convincing humans that it was real rather than any great algorithmic sophistication; in the field of algorithms (discussed later) data quality is all important to the functioning of any algorithm, just as it is for any research study¹⁷¹.

Having shown that passing the Turing Test was a goal within reach, ELIZA spawned a new era of interest in ‘chatterbots’ (shortened to ‘chatbots’ in later years) and others followed over the subsequent decades¹⁷². In 1972, Kenneth

¹⁶⁷ Jack Cahn, ‘CHATBOT: Architecture, Design, & Development’ (University of Pennsylvania, 2017).

¹⁶⁸ Schmidhuber, ‘2006: Celebrating 75 years of AI - History and Outlook: the Next 25 Years’.

¹⁶⁹ Joseph Weizenbaum, ‘ELIZA—a Computer Program for the Study of Natural Language Communication Between Man and Machine’, *Commun. ACM*, 9.1 (1966), pp. 36–45, doi:10.1145/365153.365168.

¹⁷⁰ Weizenbaum, ‘ELIZA—a Computer Program for the Study of Natural Language Communication Between Man and Machine’., ‘ELIZA—a Computer Program for the Study of Natural Language Communication Between Man and Machine’

¹⁷¹ David J. Hand, ‘Principles of Data Mining’, *Drug Safety*, 30.7 (2007), pp. 621–22, doi:10.2165/00002018-200730070-00010.

¹⁷² Cahn, ‘CHATBOT: Architecture, Design, & Development’.

Colby created PARRY at Stanford University¹⁷³. Also designed with the input of psychologists, PARRY was programmed to imitate the behaviour of a paranoid schizophrenic by producing hostile and defensive outbursts¹⁷⁴. Some clinicians struggled to differentiate between PARRY and real patients, and it is interesting to note that clinical application and psychological research were as much at the heart of these early chatbot experiments as were computational questions around natural language processing¹⁷⁵. Shortly after PARRY's creation, ELIZA and PARRY had a conversation¹⁷⁶ which highlighted both their limitations and moments of surprising lucidity within the context of clinical psychology.

After the dawn of the World Wide Web, when bots were seen as needing to be inherently online entities, chatbots started to make the most of the potential reach offered by the growing Web¹⁷⁷. Along with this, chatbot development had received enough interest to spark the creation of an annual prize, the Loebner Prize¹⁷⁸, offered for the chatbot that came closest to passing the Turing Test across a panel of judges. With this in mind, Richard Wallace created A.L.I.C.E in 1995 and she has since gone onto win the Loebner Prize three times¹⁷⁹. The placement of A.L.I.C.E on the Web, where any user can still engage in a conversation with her, heralded a new era for chatbots as the potential availability of conversations increased enormously, and with it the commercial desire to automate such conversations¹⁸⁰. Microsoft has had contrasting fortunes when developing and deploying two chatbots, Xiaolce and Tay, on social media platforms¹⁸¹. Xiaolce, a social chatbot designed to interact like a teenage girl and used across Chinese social media since 2014, has been

¹⁷³ Kenneth Mark Colby, 'Modeling a Paranoid Mind', *Behavioral and Brain Sciences*, 4.4 (1981), pp. 515–34, doi:10.1017/S0140525X00000030.

¹⁷⁴ J Epstein and W. D Klinkenberg, 'From Eliza to Internet: A Brief History of Computerized Assessment', *Computers in Human Behavior*, 17.3 (2001), pp. 295–314, doi:10.1016/S0747-5632(01)00004-8.

¹⁷⁵ Colby, 'Modeling a paranoid mind'.

¹⁷⁶ Megan Garber, 'When PARRY Met ELIZA: A Ridiculous Chatbot Conversation From 1972', *The Atlantic*, 2014 <<https://www.theatlantic.com/technology/archive/2014/06/when-parry-met-eliza-a-ridiculous-chatbot-conversation-from-1972/372428/>> [accessed 14 February 2019].

¹⁷⁷ Cahn, 'CHATBOT: Architecture, Design, & Development'.

¹⁷⁸ 'Loebner Prize', *AISB - The Society for the Study of Artificial Intelligence and Simulation of Behaviour* <<http://www.aisb.org.uk/events/loebner-prize>> [accessed 14 February 2019].

¹⁷⁹ Cahn, 'CHATBOT: Architecture, Design, & Development'.

¹⁸⁰ Dale, 'The return of the chatbots'.

¹⁸¹ Heung-yeung Shum, Xiao-dong He, and Di Li, 'From Eliza to Xiaolce: Challenges and Opportunities with Social Chatbots', *Frontiers of Information Technology & Electronic Engineering*, 19.1 (2018), pp. 10–26, doi:10.1631/FITEE.1700826.

remarkably successful, establishing a significant human following through its largely believable conversations¹⁸². However, Tay, the equivalent teenage chatbot for the Western social media market (primarily the USA), was shut down only days after its launch in 2016 after it began spouting fascist and anti-Semitic hate speech on Twitter¹⁸³. This stark contrast reinforced the lesson demonstrated by ELIZA fifty years earlier: a good conversational AI needs quality data, and many unscrupulous users of Western social media soon worked out how to feed Tay ‘bad’ data from which to learn, resulting in the inevitable. Many commentators have suggested that the case of these two chatbots represents not technical differences, but social and cultural ones amplified by the respective social media ecosystems¹⁸⁴.

2.2.3.2 Chatbots in the Present

The explosion of online chatbots over recent years has led to enormous commercial interest that has seen chatbots pop up everywhere (see Figure 2: for context this was the situation in 2019 - a chatbot map for 2023 would be significantly larger and more complex). Some industry experts even predict that Web browsing will soon become a thing of the past, instead users will simply interact with chatbots and allow them to do all the hard work¹⁸⁵. Commercial chatbots have gained increasing usage with companies now looking to implement automated agents in a variety of settings, especially in the customer service sector¹⁸⁶. Studies have shown that major motivations for deploying chatbots are to increase productivity¹⁸⁷ and improve customer experience in using a service; for instance, Bank of America’s ‘erica’ has made the job of

¹⁸² Shum, He, and Li, ‘From Eliza to Xiaolce’., ‘From Eliza to Xiaolce: challenges and opportunities with social chatbots’

¹⁸³ Neff and Nagy, ‘Automation, Algorithms, and Politics| Talking to Bots’.

¹⁸⁴ James Vincent, ‘Twitter Taught Microsoft’s Friendly AI Chatbot to Be a Racist Asshole in Less than a Day’, *The Verge*, 2016 <<https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>> [accessed 14 February 2019].

¹⁸⁵ Dale, ‘The return of the chatbots’.

¹⁸⁶ Petter Bae Brandtzaeg and Asbjørn Følstad, ‘Why People Use Chatbots’, in *Internet Science*, ed. by Ioannis Kompatsiaris and others, Lecture Notes in Computer Science (Springer International Publishing, 2017), pp. 377–92.

¹⁸⁷ Brandtzaeg and Følstad, ‘Why People Use Chatbots’. ‘Why People Use Chatbots’

searching for information in transaction histories far easier and smoother for more than a million customers¹⁸⁸.

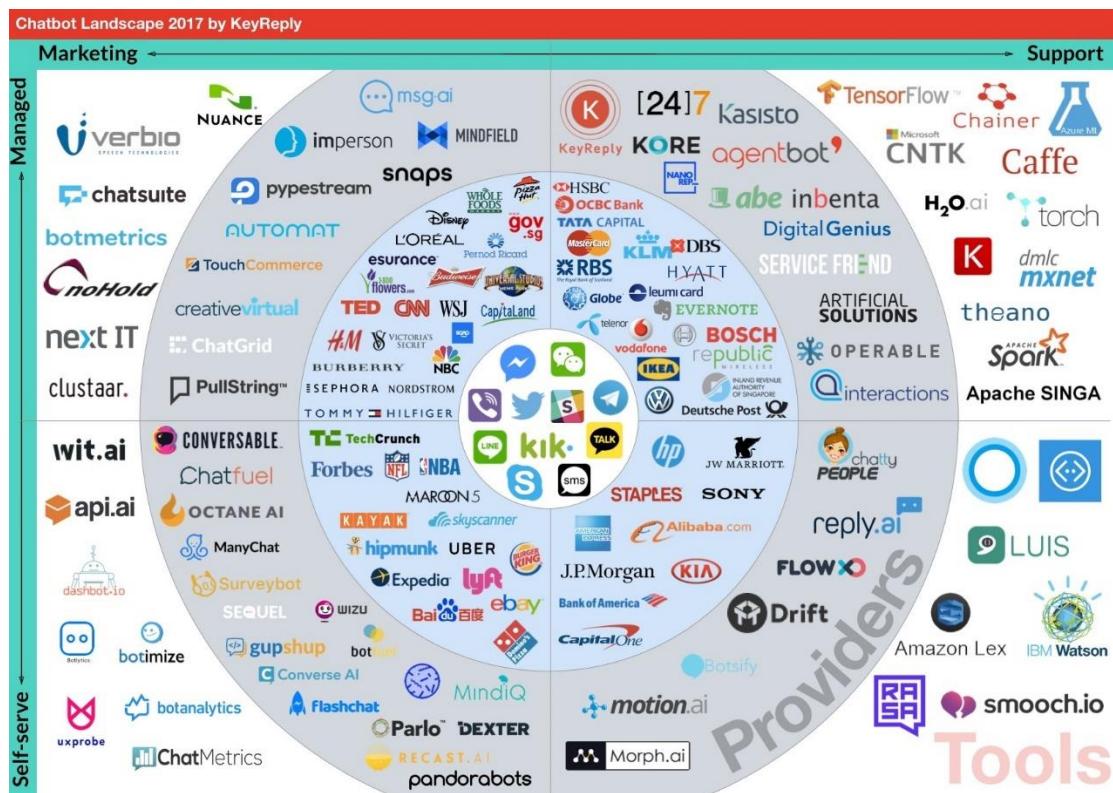


Figure 2: Overview of current chatbot landscape. Source: Carylyne Chan.¹⁸⁹

Chatbots, including many social chat bots, are now to be found on dedicated platforms which are designed with both bots and humans in mind. Several popular platforms now exist, including Facebook Messenger¹⁹⁰, WhatsApp¹⁹¹, Slack¹⁹², Kik¹⁹³, WeChat¹⁹⁴, Telegram¹⁹⁵ and Discord¹⁹⁶. These platforms, which

¹⁸⁸ Penny Crosman, 'Mad about Erica: Why a Million People Use Bank of America's Chatbot', *American Banker*, New York, N.Y. (14 June 2018) <<http://search.proquest.com/docview/2054327212/abstract/8855C1FD21E0437FPQ/1>> [accessed 14 February 2019].

¹⁸⁹ Carylyne Chan, 'The Chatbot Landscape, 2017 Edition', *Medium*, 2017 <<https://medium.com/keyreply/the-chatbot-landscape-2017-edition-ff2e3d2a0bdb>> [accessed 26 May 2020].

¹⁹⁰ 'Messenger', *Facebook* <<https://www.messenger.com/>> [accessed 14 February 2019].

¹⁹¹ 'WhatsApp', *WhatsApp.Com* <<https://www.whatsapp.com/>> [accessed 14 February 2019].

¹⁹² Slack, 'Where Work Happens', *Slack* <<https://slack.com/>> [accessed 14 February 2019].

¹⁹³ 'Kik' <<https://www.kik.com/>> [accessed 14 February 2019].

¹⁹⁴ 'WeChat for Web' <<https://web.wechat.com/>> [accessed 14 February 2019].

¹⁹⁵ 'Telegram – a New Era of Messaging', *Telegram* <<https://telegram.org/>> [accessed 14 February 2019].

¹⁹⁶ 'Discord - Free Voice and Text Chat', *Discord* <<https://discordapp.com>> [accessed 14 February 2019].

are all designed primarily to function as apps on smartphones and other Internet-connected mobile devices, also have web-based sites for use on desktop and laptop PCs. However, most users interact with their app (or mobile application) incarnations through a smartphone.

The stratospheric rise of apps and global smartphone usage¹⁹⁷ has seen a boom in social messaging applications like those listed above. Making use of the near-ubiquity of Internet access, at least in the developed world, such apps are creating a new type of coordination, ‘Microcoordination 2.0’¹⁹⁸, by allowing networks of individuals to communicate with each other through speech, text, image and video with consummate ease. Ling and Lai suggest that this has the effect of digitally propagating cultural phenotypes, such as memes, and changes the nature of both human group interaction and the cognitive efforts of social interaction¹⁹⁹. In this context, the significant role of chatbots on these platforms is hardly surprising. Platforms like Telegram and Slack have their own ‘bot directories’, in effect ‘bot stores’ that work in a similar fashion to the ‘app stores’ that act as repositories of mobile apps for smartphone users to download²⁰⁰. As Klopfenstein et al. point out²⁰¹, the variety of bots available for these platforms moves many of them away from being mere ‘chat’ bots; there are bots that organise diaries, automate payments, sort out travel arrangements or order pizza for everyone in a given chat group²⁰². These bots still interact with humans via natural language, hence why they retain the ‘chat’ moniker, but some are hardly ‘conversational’, instead displaying lists of options (‘button tree’) that effectively do all the talking²⁰³.

¹⁹⁷ Thanasis Petsas and others, ‘Rise of the Planet of the Apps: A Systematic Study of the Mobile App Ecosystem’, in *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC ’13 (ACM, 2013), pp. 277–90, doi:10.1145/2504730.2504749.

¹⁹⁸ Rich Ling and Chih-Hui Lai, ‘Microcoordination 2.0: Social Coordination in the Age of Smartphones and Messaging Apps’, *Journal of Communication*, 66.5 (2016), pp. 834–56, doi:10.1111/jcom.12251.

¹⁹⁹ Ling and Lai, ‘Microcoordination 2.0’., ‘Microcoordination 2.0: Social Coordination in the Age of Smartphones and Messaging Apps’

²⁰⁰ Lorenz Cuno Klopfenstein and others, ‘The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms’, in *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS ’17 (ACM, 2017), pp. 555–65, doi:10.1145/3064663.3064672.

²⁰¹ Klopfenstein and others, ‘The Rise of Bots’., ‘The Rise of Bots’.

²⁰² The Vinny G. Bot, ‘Order a Pizza, Through Slack, Today!’, *The Vinny G Bot*, 2016 <<https://medium.com/@VinnyGBot/order-a-pizza-through-slack-today-3f371764d417>> [accessed 14 February 2019].

²⁰³ Klopfenstein and others, ‘The Rise of Bots’.

What does this new, vast array of chatbots mean for users seeking to engage with history, either for entertainment or serious fact-finding, on such messaging platforms? Firstly, much as is the case on Twitter, a series of entertaining bots purport to offer conversations with historical figures, such as George Washington²⁰⁴, provide a snippet of historical information that happened 'on this day'²⁰⁵ or highlight lesser-known historical perspectives, such as 'Black History Facts'²⁰⁶ for Amazon Alexa. The likes of Alexa, Apple's Siri and Google's Assistant are classified as 'virtual private assistants' and interact with users primarily through speech and voice recognition²⁰⁷. Bots are becoming more popular for these platforms too, particularly Amazon's Alexa, but these assistants are also chatbots in their own right, attempting to answer questions, carry out tasks and find information. In this last case, such bots usually rely upon existing information platforms and / or search services to do the job, for instance asking Alexa about a historical figure will most likely result in the corresponding Wikipedia summary being recited to the user. This is an established method of bot action and communication, and when combined with the bots of Wikipedia and the Google search algorithm, a complex set of automated interactions is involved in the processing of such a request (why this might matter for those seeking historical information is discussed in the next sub-section). In the heritage sector, museums have also been experimenting with chatbots, but so far the vast majority have offered rigidly structured customer service and marketing functions, answering questions around museum opening times or current exhibitions, rather than conversing naturally or creatively to educate users about the historical content of collections²⁰⁸. There are a few notable exceptions (see Chapter 4), but the real potential for chatbots in the heritage sector is yet to be realised.

However, the advent of generative AI models and tools like ChatGPT (see end of this chapter for more detail), offers GLAMs new opportunities to create truly

²⁰⁴ Will Knight, 'A Plague of Dumb Chatbots', *Technology Review; Cambridge*, August 2016, 28.

²⁰⁵ 'History Buddy Chatbot on BotList' <<https://botlist.co/bots/history-buddy>> [accessed 14 February 2019].

²⁰⁶ 'Black History Facts Chatbot on BotList' <<https://botlist.co/bots/black-history-facts>> [accessed 14 February 2019].

²⁰⁷ Klopfenstein and others, 'The Rise of Bots'.

²⁰⁸ Tzouganatou, 'Can Heritage Bots Thrive?'

conversational chatbots that will allow users to engage with collections in far more natural ways. These new AI models are multimodal, which means they can handle text, image, audio, video and potentially 3d objects, giving a far greater range of options to curators, GLAM professionals and users in how they could be implemented. Indeed, research is currently ongoing to use the latest models as conversational tour guides, and this is an area where a significant amount could be gained from further study²⁰⁹. It also serves to show the potential ubiquity, diversity and social power of AI-driven chatbots, making them a phenomenon worthy of consideration for researchers interested in the consumption, dissemination and protection of the past, as well as GLAMs looking to novel ways of engaging physical and virtual visitors with their collections and communities.

2.3 Web 3.0: From Social Bots to Social Machines

By 2009, when the social networks enabled by Web 2.0 technologies were rapidly expanding, there were already signs that a new era was well under way: 'Web 3.0'. Ten years earlier, Sir Tim Berners-Lee first introduced the idea of a 'social machine' in a book entitled *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. In an oft-quoted passage, Berners-Lee outlines a broad vision for social machines:

Real life is and must be full of all kinds of social constraint – the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: processes in which people do the creative work and the machine does the administration.²¹⁰

The term 'social machine' still encompasses that idea of humans and computers working together as part of a machine. On the Web, this really means a variety of humans, from people who consume and/or create content (the produsers), to developers who create and maintain platforms like Wikipedia

²⁰⁹ Georgios Trichopoulos and others, 'Crafting a Museum Guide Using GPT4' (Preprints, 2023), doi:10.20944/preprints202306.1618.v1.

²¹⁰ Tim Berners-Lee and Mark Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor* (Harper San Francisco, 1999), p. 172.

or Twitter, and a number of different kinds of program run by computers: editing bots on Wikipedia, recommender algorithms on social media or an AI customer service chatbot on a company's website. The interactions between all the people, algorithms, bots and AI tools involved, at every level, together constitute the potentially vast and complex workings of the social machine. The Web can therefore be seen as a network of many such social machines, of varying size and composition, that may interact with each other to different extents.

Within the original formulation of this model by Berners-Lee, there is one key aspect of this idea that is often overlooked – the word “abstract”. With hindsight, as Berners-Lee himself notes²¹¹, it is easy to critique the overly optimistic (perhaps naïve) vision of ‘real life’ and ‘society’ being distinct from the Web, and that with misaligned goals machine-mediated algorithmic administration leads to ‘creative’ work becoming a mundane and potentially harmful chore central to precarious Web-based livelihoods²¹². However, ‘abstract’ belies a pervasive mode of thinking that even historical context cannot fully account for, now sometimes seen as a Silicon Valley stereotype, that the Web is an ‘experimental model’, ‘sandbox’ or ‘playground’ in which new toys can be enjoyed, broken, snatched from other children and then hurled back at them disdainfully.

Unfortunately, the toys which prove most popular end up as social machines so powerful that they pull in millions of users, each becoming a cog in a vast and complex system whose insatiable appetite requires huge amounts of resource to ensure its continued existence. ‘Real life’ and ‘society’ are suddenly drawn into focus, and everyone realises too late that they’d been there all along. Across social media platforms, and more recently AI tools like OpenAI’s ChatGPT, safety measures are often implemented piece-meal and their effectiveness cannot be guaranteed, with such safety systems by design including algorithmic approaches that can be subverted or penalise those not

²¹¹ John Harris, ‘Tim Berners-Lee: “We Need Social Networks Where Bad Things Happen Less”’, *The Guardian*, 15 March 2021, section Life and style <<https://www.theguardian.com/lifeandstyle/2021/mar/15/tim-berners-lee-we-need-social-networks-where-bad-things-happen-less>> [accessed 2 December 2023].

²¹² Laura Herman, ‘For Who Page? TikTok Creators’ Algorithmic Dependencies’, in *IASDR 2023: Life-Changing Design* (presented at the IASDR 2023: Life-Changing Design, Design Research Society, 2023), doi:10.21606/iasdr.2023.576.

acting maliciously²¹³. And this is on top of a wholesale abandonment by the sector of ethics as a worthwhile topic of discussion, let alone implementation²¹⁴. Wikipedia is an example showing that these depressing outcomes are not the inevitable products of the societal reality of social machines, but it is still firmly rooted in the reality of a sociotechnical community with debate and disagreement a healthy part of that ecosystem²¹⁵. This is perhaps closest to the vision that came to mind for Berners-Lee back in 1999, although Web 3.0 might yet supplant it.

Back to 2001 – to realise the dream of social machines, it was clear to Sir Tim Berners-Lee, Jim Hendler and Ora Lassila that there would need to be a new kind of Web, the ‘Semantic Web’, which would make information available online more understandable to computers, allowing them to work with it more intelligently so that users could ask more complex search questions rather than just typing in keywords²¹⁶. At the heart of this idea is the technology of Linked Data, a set of rules and data models that represent links between two pieces of information to create meaning, hence the ‘Semantic’ Web. Just as a grammatical sentence follows the rule ‘subject’ – ‘verb’ – ‘object’, so a Linked Data ‘triple’ consists of ‘subject’ – ‘predicate’ – ‘object’; it is the linking predicate that enables meaning. For example: ‘The Mona Lisa’ – ‘was created by’ – ‘Leonardo Da Vinci’ is one triple; ‘Le Joconde à Washington’ – ‘is about’ – ‘The Mona Lisa’ is another (see Figure 4). As more and more triples link together, so a vast ‘knowledge graph’ is constructed which can expand ad infinitum and multiply meaning exponentially. Such a graph allows a user to ask: ‘What are all the buildings owned by the President of the USA?’, and the computer could return an accurate list. This is the technology upon which WikiData is built, with

²¹³ Andrei Taraschuk on Twitter: “I Stopped Sharing Art on Facebook a While Ago but They Just Restricted My Account for Something @artist_dali Shared in 2017... That Artwork Tho, Lol [Https://T.Co/94Wixwstpv](https://T.Co/94Wixwstpv)” / Twitter’, *Twitter* <<https://twitter.com/andreitr/status/1254645595760783360>> [accessed 22 May 2020].

²¹⁴ Dataethics, ‘Big Tech Is Cutting Down on Ethics · Dataetisk Tænkehandletank’, 2023 <<https://dataethics.eu/big-tech-is-cutting-down-on-ethics/>> [accessed 2 December 2023].

²¹⁵ Anasuya Sengupta and Brooke A. Ackerly, ‘Wikipedia Edit-A-Thons: Sites of Struggle, Resistance, and Responsibility’, *PS: Political Science & Politics*, 55.2 (2022), pp. 434–38, doi:10.1017/S1049096521001220.

²¹⁶ Tim Berners-Lee, Jim Hendler, and Ora Lassila, ‘The Semantic Web’, *Scientific American*, 2001.

the added stipulation that its Linked Data must be ‘Open’, i.e. freely available for use by anyone, giving the term ‘Linked Open Data’ (or LOD).

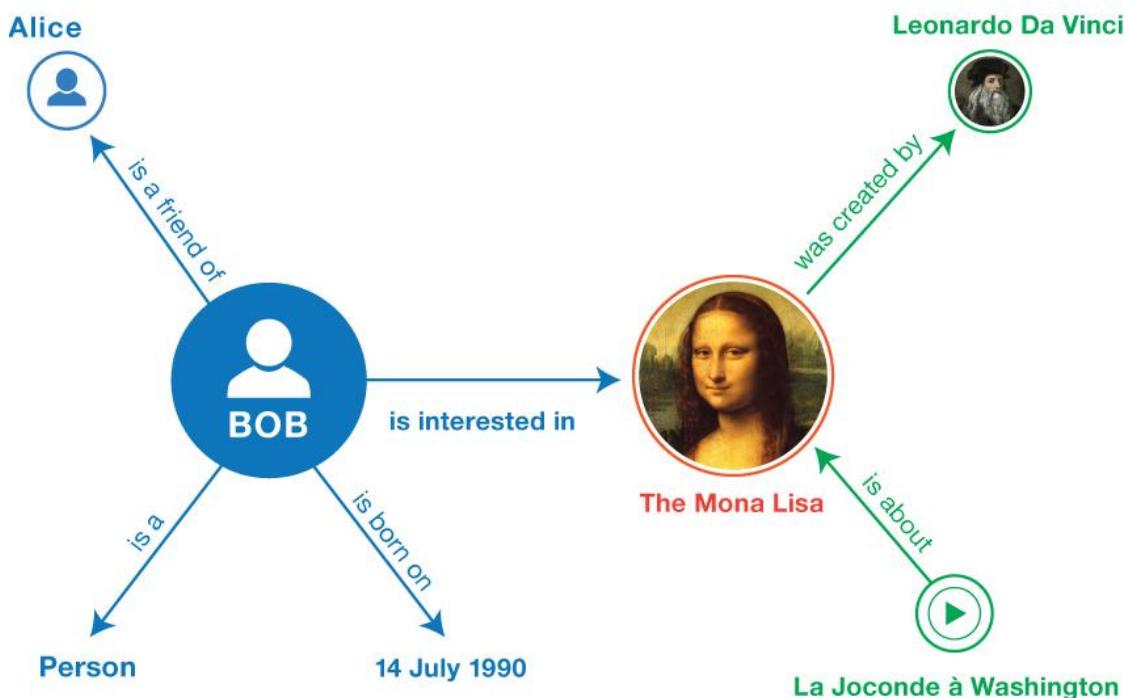


Figure 3: Diagram showing a graph made up of triples, the circles being subjects and objects and the arrows linking them together the predicates. Source: W3C RDF 1.1 Primer 2.²¹⁷

The cultural heritage sector has made great efforts to explore the potential of the Semantic Web through Linked Open Data over the past decade and its involvement in the areas of heritage, archaeology and history is only increasing²¹⁸. ‘GLAMs’ (Galleries, Libraries, Archives and Museums) is an acronym born from a collaborative effort across the sector to encourage open sharing of data, enable digitisation and online access to collections and archives and to improve the usability and usefulness of data and searches through semantic technologies like Linked Open Data. One of the main examples of this approach is Europeana, a ten-year-old, EU-funded project that

²¹⁷ ‘RDF 1.1 Primer’ <<https://www.w3.org/TR/rdf11-primer/>> [accessed 26 May 2020].

²¹⁸ Julia Marden and others, ‘Linked Open Data for Cultural Heritage: Evolution of an Information Technology’, in *Proceedings of the 31st ACM International Conference on Design of Communication*, SIGDOC ’13 (ACM, 2013), pp. 107–12, doi:10.1145/2507065.2507103; Allana Mayer, ‘Linked Open Data for Artistic and Cultural Resources’, *Art Documentation: Journal of the Art Libraries Society of North America*, 34.1 (2015), pp. 2–14, doi:10.1086/680561.

has grown to cover the aggregation of data from GLAMs and other cultural heritage institutions across Europe (not just the EU) to collect and create a shared heritage²¹⁹. This sort of project exemplifies Tim Berners Lee's vision of a semantic web of data, but although it has achieved much success in promoting the digitisation and digitalisation of cultural heritage whilst improving access to it, Europeana also exemplifies some of the problems with using linked, open and semantic approaches in the real world; these problems apply to many other projects in the same ilk, most notably WikiData, the semantic knowledge-graph aiming to underpin Wikipedia and bring Linked Open Data to the masses²²⁰.

One of the main stumbling blocks, both technically and pragmatically, is the different systems of logic, or 'ontologies', that different Linked Open Data platforms use. Returning to the idea of a triple being subject – predicate – object, an ontology decides the definitions of the predicates and the rules governing their use²²¹. For example, in the case recording a person as the creator of a painting, one ontology might state that Van Gogh 'is a person' and 'is creator of' 'Sunflowers', whereas another may describe this relationship as: 'Sunflowers' 'underwent production' which 'was achieved by' Van Gogh (if these example predicates sound stiltedly confusing, they are easier to read than the real things). Some ontologies are simpler to understand than others but lack the ability to give real depth to the information and relationships described, whereas others prioritise these aspects at the cost of being extremely complex and often confusing, such as the frequently used CIDOC-CRM²²².

There is also the point that these existing ontologies are predicated on and reinforce predominantly Western conceptions of knowledge and links between

²¹⁹ Bjarki Valtysson, 'Europeana', *Information, Communication & Society*, 15.2 (2012), pp. 151–70, doi:10.1080/1369118X.2011.586433.

²²⁰ European Commission, *REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the Evaluation of Europeana and the Way Forward* (European Commission, 2018) <<https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1536235661093&uri=COM:2018:612:FIN>> [accessed 23 August 2019].

²²¹ *Semantic Web Information Management: A Model-Based Perspective*, ed. by Roberto De Virgilio, Fausto Giunchiglia, and L. Tanca (Springer, 2010), pp. 25–37.

²²² This gives an overview of the various projects and ontologies used across cultural heritage as of 2018 Vladimir Alexiev, 'Museum Linked Open Data: Ontologies, Datasets, Projects', *Digital Presentation and Preservation of Cultural and Scientific Heritage*, VIII, 2018, pp. 19–50.

entities²²³. Using the example of Aboriginal maps as cases where knowledge of beings, places and relationships are all wrapped up in single entities, it is clear that Indigenous ontologies would require a more flexible, complex and interpretive structure that goes beyond the subject-predicate-object model.

Within this context, approaches focussing on decolonising existing ontologies and data structures have also highlighted the need to recognise and represent non-Western epistemologies if truly diverse perspectives on the past are to be made possible in digitally-mediated spaces²²⁴. Methodologies seeking to achieve this include attempts to bring nuanced information in iconographic representations in Mesoamerican maps into novel LOD-based ontologies, the inherent complexities of which are highlighted in these examples of representations reflecting the collision and combination of Western and non-Western epistemologies as part of the processes of colonisation²²⁵. Such work also extends to more practical aspects of cultural heritage management, where Western concepts of 'heritage' built in to existing software tools do not reflect the ontological perspectives of communities seeking to decolonise their heritage so that it can then be represented and understood from their perspective²²⁶.

Ontologies are only one part of the system; there are different ways to describe the triples, the most popular being the combination of the RDF schema and the OWL language. This means that RDF is the language used to define the concepts in the ontologies, such as of subjects and objects and what types of predicates (links) are possible, such as 'is_a_person', while OWL supplies the semantic structure which can give a specific context to the general type of link 'is_a_person', such as 'is_an_art_lover'. This combination gives the resulting triple through which meaning is created.

²²³ Ramesh Srinivasan, 'Re-Thinking the Cultural Codes of New Media: The Question Concerning Ontology', *New Media & Society*, 15.2 (2013), pp. 203–23, doi:10.1177/1461444812450686.

²²⁴ Ramesh Srinivasan, *Whose Global Village?: Rethinking How Technology Shapes Our World* (NYU Press, 2018).

²²⁵ Candela and others, 'An Ontological Approach for Unlocking the Colonial Archive'.

²²⁶ Agnes Sofia Mpingana Shiningayamwe, 'Decolonizing Heritage Management Systems: New Directions in Digital Heritage Management from Namibia', *Conservation and Management of Archaeological Sites*, 0.0 (2023), pp. 1–21, doi:10.1080/13505033.2023.2287894.

There are different vocabularies for various specialist areas, for instance the Getty vocabularies define the ways in which subjects and objects should be described when recording artworks to ensure consistency (i.e. that the author, art type, material, date of creation etc are always included in an object's information)²²⁷. All these various standards and systems of naming, logic and coding can be used by different projects in the cultural heritage space, causing utter bafflement to the researcher trying to navigate the different systems, and so 'interoperability' – the goal of making the different systems in some way compatible – is still high on the Linked Open Data agenda.

However, the fact that all the information and relationships present within Linked Data are by design machine readable, meaning that this complex information can be processed automatically, means that it still holds great potential particularly as an information source for bots that could be accessed programmatically once diverse datasets are linked together into a single knowledge-graph. While such complex data systems would seem to lend themselves to the potential power of automation and AI approaches, these are still developing and are not yet fully realised, despite much work in semantic search and recommendation especially for cultural heritage. The gradual rise in profile of Linked Open Data across cultural heritage, medicine, law and increasingly other areas is indicative of Web 3.0 as a whole.

A 2006 Deutsche Telekom report²²⁸ heralded Web 3.0 as the next great innovation achievable through the convergence of Web 2.0 and the Semantic Web, highlighting the potential power of linking together people and data through the Web. Indeed, the 'futuristic' examples put forward in the original 2001 vision for the Semantic Web talk of agents booking appointments and finding out information for users through speech recognition, something that many modern chatbots can now achieve (with varying degrees of success). But much of this is still reliant on the social links (and social data) enabled and

²²⁷ Patricia Harpring, 'Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA', *Art Documentation: Journal of the Art Libraries Society of North America*, 29.1 (2010), pp. 67–72, doi:10.1086/adx.29.1.27949541.

²²⁸ Wolfgang Wahlster and others, 'Web 3.0: Convergence of Web 2.0 and the Semantic Web', in *In Technology Radar, Feature Paper, 2nd Ed.*; Deutsche Telekom Laboratories, 2006, pp. 1–23 (p. 0).

created by Web 2.0, and today, although Web 3.0 is seen as a probable, rather than potential, future, its full realisation is still some way off. There is nothing more reliant upon semantics than social interaction, as demonstrated by the deeply interwoven strands of the social Web 2.0 and the Semantic Web from which Web 3.0 emerges, and which themselves make a mockery of the arbitrary, pseudo-logical classification of these Web 'eras'²²⁹. Perhaps with the benefit of hindsight, that much-loved historical equivalent of geological time, Webs 1.0, 2.0 and 3.0 will appear as distinct entities, but at the moment they are still messily intermingled and interdependent, especially the latter two.

This fact was not lost on Berners-Lee and Hendler in their 2009 article 'From the Semantic Web to social machines'²³⁰ in which they argue that the social and technological aspects of Web 2.0 make platforms like Wikipedia 'social machines' which can be made more powerful and pervasive primarily through greater collaboration and the linking together of online communities. This linking unsurprisingly requires the ubiquity, consistency and intelligence supplied by the Semantic Web of Linked Data; WikiData is a prime example of this vision, and possibly the first Web 3.0 social machine. The article's vision for social machines also highlights the important role of AI in realising their full potential, and as discussed in the following section the emergence of modern chatbots connected to various knowledge sources and able to communicate in natural language may appear to be the cogs in a new social engine, but they only partially tap into the Semantic Web, relying more broadly on the vast swathes of data introduced by Web 2.0 technologies.

Web 3.0 technologies, namely the Semantic Web fully realised through ubiquitous use of Linked Open Data, is far from being a universal reality, but it is expanding, especially within the GLAMs sector. However, with the recent rapid advances in generative AI capabilities, and the growing spread of tools like ChatGPT, this now poses a question for the Semantic Web movement: if AI can

²²⁹ Veronica Barassi and Emiliano Treré, 'Does Web 3.0 Come after Web 2.0? Deconstructing Theoretical Assumptions through Practice', *New Media & Society*, 14.8 (2012), pp. 1269–85 (p. 0), doi:10.1177/1461444812445878.

²³⁰ Jim Hendler and Tim Berners-Lee, 'From the Semantic Web to Social Machines: A Research Challenge for AI on the World Wide Web', *Artificial Intelligence*, Special Review Issue, 174.2 (2010), pp. 156–61, doi:10.1016/j.artint.2009.11.010.

now read the natural language of human text as well as any human, is the core aim of making information machine-readable actually that relevant anymore? In the short term, it is becoming clear that these capabilities of AI models are still at the stage where knowledge graphs are beneficial to helping them extract useful meaning from vast databases of information²³¹. Currently, the likes of Microsoft and Amazon are developing generative AI-based chatbots that include links to knowledge graphs to act as memory stores for conversations and wider sources of information, but these are not open knowledge graphs, they are proprietary parts of commercially-driven tools²³².

Hence, while the future for Linked Data approaches is optimistic, with the likes of ChatGPT offering the opportunity to unlock their true potential, that of Web 3.0 and all that Linked Open Data stands for is far more uncertain; perhaps the GLAM sector, as an early adopter of the collaborative nature of Linked Open Data, can show how such open approaches may also be enhanced by the latest AI tools.

2.4 Algorithms

The discussions around automation and bots above have emphasised the sudden increase in both usage and awareness of bots in the last couple of years; the same could be said of algorithms, but their rise has been steadier, more closely monitored and far more significant for society both online and offline²³³. When considering GLAMs specifically, but this also goes for many of the Web social platforms upon which GLAMs and users interact, algorithms are

²³¹ Juan Sequeda, Dean Allemang, and Bryon Jacob, 'A Benchmark to Understand the Role of Knowledge Graphs on Large Language Model's Accuracy for Question Answering on Enterprise SQL Databases' (arXiv, 2023), doi:10.48550/arXiv.2311.07509.

²³² Will Knight, 'Amazon Upgrades Alexa for the ChatGPT Era', *Wired*, 2023 <<https://www.wired.com/story/amazon-upgrades-alexa-for-the-chatgpt-era/>> [accessed 1 December 2023]; Maggie Harrison, 'Microsoft Patents AI-Powered Therapy App', *Futurism*, 2023 <<https://futurism.com/microsoft-patent-ai-therapy>> [accessed 18 November 2023].

²³³ See for example: David Beer, 'The Social Power of Algorithms', *Information, Communication & Society*, 20.1 (2017), pp. 1–13, doi:10.1080/1369118X.2016.1216147; Natascha Just and Michael Latzer, 'Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet', *Media, Culture & Society*, 39.2 (2017), pp. 238–58, doi:10.1177/0163443716643157; Nicholas Carah, 'Algorithmic Brands: A Decade of Brand Experiments with Mobile and Social Media', *New Media & Society*, 19.3 (2017), pp. 384–400, doi:10.1177/1461444815605463.

used for recommending collections or objects to users, with the aim of helping them engage with collection content in a more enjoyable and personalised way.

Defining an algorithm is technically far easier than defining a bot; an algorithm is simply a set of instructions that produces a result through a finite number of processes²³⁴. In theory, this can apply to any set of instructions, therefore a cake recipe could count as an algorithm, but when applied in a computational setting each step of an algorithm is usually a mathematical operation expressed in computer code. Compared to the other definitions so far, this seems very straightforward; however, this simplicity means that an algorithm can account for a huge number of automatic computer programs, or parts of programs, from complex statistics to the simplest arithmetic. In theory, any simple automatic script could be considered an algorithm, but in practice it is usually a set of more complex statistical algorithms, referred to as machine learning algorithms (see below), that are brought to mind when the term ‘algorithmic’ is used²³⁵.

Algorithms often constitute the inner workings of chatbots and social bots, the ‘bot’ simply providing the voice, face and personality for its algorithms (this aptly describes most commercial chatbots). Although, as already discussed, many bots are simple automated programs that would not be considered ‘algorithmic’, even if some managed to produce an illusion of ‘intelligence’. Similarly, despite public perceptions, many complex statistical algorithms would not be considered ‘intelligent’ either; often it is the combination of different algorithms to give the impression of ‘learning’ that confers the status of ‘artificial intelligence’ upon a computer program²³⁶. Therefore, this can be as much a social construct as a technological one.

Indeed, online algorithms are often portrayed as being everywhere and controlling everything in a faceless, insidious manner reminiscent of secretive World Orders in Hollywood blockbusters, whereas bots can be seen as cute,

²³⁴ Jean-Luc Chabert, Évelyne Barbin, and Chris Weeks, *A History of Algorithms: From the Pebble to the Microchip* (Springer, 1999), p. 2.

²³⁵ Just and Latzer, ‘Governance by algorithms’.

²³⁶ Simone Natale and Andrea Ballatore, ‘Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence’, *Convergence: The International Journal of Research into New Media Technologies*, 2017, p. 135485651771516, doi:10.1177/1354856517715164.

funny and even endearing²³⁷. All of this begs the questions: what is this intangible mass of algorithms doing and why?

One answer is for commercial and economic purposes, although research and non-commercial data collection do have a small part to play. A recent, extensive survey of online algorithms concluded that the vast majority were used “mostly for purely commercial goals”²³⁸, primarily advertising. Web-based advertising has been around almost as long as the Web itself, with advertisers keen to find ways of exploiting the new communication medium with its enormous potential reach²³⁹. Indeed, by the mid-1990’s, when Larry Page and Sergey Brin were working on a new kind of search algorithm that would become the core of the Google search engine (see below), they noted that the majority of search technology was “advertising oriented”.²⁴⁰

The following two decades have seen new forms of online advertising and marketing emerge, from the exploitation of new social platforms to the beginnings of personalised advertising which has helped to spark the current data-driven approach to targeted marketing and ad placement²⁴¹. Modern advertising is an algorithmic affair, with algorithms using data (often collected from users) to decide which ads should appear on different pages for different users, whilst automated auctions determine which companies get their ads displayed in each available space. This is tied in with which pages are deemed mostly likely to be attracting the largest, most relevant audience, a metric determined by different algorithms on different platforms (i.e. Google’s ranking algorithm works differently to Facebook’s, which is different to Twitter’s etc), and gaming this system to favour one company or another has become a lucrative

²³⁷ Natale and Ballatore, ‘Imagining the thinking machine’; Florian Daniel and others, ‘Toward Truly Personal Chatbots: On the Development of Custom Conversational Assistants’, in *Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services*, SE4COG ’18 (ACM, 2018), pp. 31–36, doi:10.1145/3195555.3195563.

²³⁸ Just and Latzer, ‘Governance by algorithms’, p. 251.

²³⁹ David S Evans, ‘The Online Advertising Industry: Economics, Evolution, and Privacy’, *Journal of Economic Perspectives*, 23.3 (2009), pp. 37–60, doi:10.1257/jep.23.3.37.

²⁴⁰ Brin and Page, ‘The anatomy of a large-scale hypertextual Web search engine’, p. 109.

²⁴¹ Evans, ‘The Online Advertising Industry’.

digital industry, as exemplified by Search Engine Optimization (SEO – see below)²⁴².

All of the above activity comes under the heading of 'c selection', described by Just and Latzer²⁴³ as: "the automated assignment of relevance to certain selected pieces of information". In an attempt to bring order to the complexity demonstrated above, Latzer²⁴⁴ produced a typology of online selection algorithms which describes nine potential categories (see Table 2). Each type influences users in subtly different ways, and while the potential for some of these algorithms to influence historical information and knowledge is seemingly low, for others it is undoubtedly significant, such as those involved in ranking search results from Google and Wikipedia, or those underlying creative content production apps.

²⁴² Dimitrios Giomelakis and Andreas A. Veglis, 'Search Engine Optimization', *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*, 2019, pp. 1789–1800, doi:10.4018/978-1-5225-7598-6.ch132.

²⁴³ Just and Latzer, 'Governance by algorithms'.

²⁴⁴ Michael Latzer et al, *The Economics of Algorithmic Selection on the Internet* (University of Zurich, 2014), doi:<http://dx.doi.org/10.5167/uzh-100400>.

Types	Examples
Search	General search engines (e.g. Google search, Bing, Baidu) Special search engines (e.g. Mocavo, Shutterstock, Social Mention) Meta search engines (e.g. Dogpile, Info.com) Semantic search engines (e.g. Yummly) Question and answer services (e.g. Ask.com)
Aggregation	News aggregators (e.g. Google News, nachrichten.de)
Observation/surveillance	Surveillance (e.g. Raytheon's RIOT) Employee monitoring (e.g. Spector, Sonar, Spytec) General monitoring software (e.g. Webwatcher)
Prognosis/forecast	Predictive policing (e.g. PredPol) Predicting developments: success, diffusion etc. (e.g. Google Flu Trends, scoreAhit)
Filtering	Spam filter (e.g. Norton) Child protection filter (e.g. Net Nanny)
Recommendation	Recommender systems (e.g. Spotify, Netflix)
Scoring	Reputation systems: music, film, and so on (e.g. eBay's reputation system) News scoring (e.g. reddit, Digg) Credit scoring (e.g. Kreditech) Social scoring (e.g. Klout)
Content production	Algorithmic journalism (e.g. Quill, Quakebot)
Allocation	Computational advertising (e.g. Google AdSense, Yahoo!, Bing Network) Algorithmic trading (e.g. Quantopian)

Source: Latzer et al. (2014).

Table 2: Table taken from Just and Latzer outlining a classification of algorithmic selection applications.²⁴⁵

2.4.1 Recommendation Algorithms

Many of us are familiar with the 'recommender systems' of the type listed in Table 2, which include not just Netflix and Spotify but also Amazon, Youtube, Facebook, eBay, Twitter (X), TikTok and pretty much every other ecommerce or social networking site. The recommendation algorithms behind these systems aim to suggest products, movies, videos or people that may be 'similar' to you

²⁴⁵ Just and Latzer, 'Governance by algorithms'.

or your previous buying, viewing, or socialising habits and do so with varying degrees of success and user irritability²⁴⁶. GLAMs also use recommender systems to suggest exploration pathways around their online collections to virtual visitors.

There are three main approaches to recommendation. Content-based recommender algorithms compute the similarity of, say, a pair of shoes already bought with others up for sale by comparing images of the shoes and their other metadata; that is the accompanying information such as their description, size, brand etc²⁴⁷. This is the 'you bought this so you might like this' approach. Content-based approaches of various kinds were the earliest types of recommendation systems employed on the web, but the rise of Web 2.0 offered new opportunities for suggestion.

Collaborative filtering recommender algorithms rely on the social nature of the Web and the huge visitor numbers to popular ecommerce sites like Amazon and eBay. These systems take the 'someone else, apparently like you, bought this so you might like to buy it too' approach. Here, the details of the product itself are not taken into consideration, merely the purchasing habits of other people who have previously bought or viewed the same, or very similar, things to you. This forms part of your 'user profile', a personal shopping record from which 'categories that may interest you' can be inferred as well as individual products²⁴⁸. Targeted advertising takes a similar approach, except that it often has access to all of your browsing habits through small files called 'cookies' which can track your activity on the Web.

Lastly, hybrid recommender systems seek to take both the content-based and social aspects into account when deciding upon similarity of fit between product and potential buyer or Facebook user and suggested friend-to-be²⁴⁹. This may

²⁴⁶ J. Bobadilla and others, 'Recommender Systems Survey', *Knowledge-Based Systems*, 46 (2013), pp. 109–32, doi:10.1016/j.knosys.2013.03.012.

²⁴⁷ Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro, 'Content-Based Recommender Systems: State of the Art and Trends', in *Recommender Systems Handbook*, ed. by Francesco Ricci and others (Springer US, 2011), pp. 73–105, doi:10.1007/978-0-387-85820-3_3.

²⁴⁸ Michael D. Ekstrand, John T. Riedl, and Joseph A. Konstan, 'Collaborative Filtering Recommender Systems', *Foundations and Trends® in Human–Computer Interaction*, 4.2 (2011), pp. 81–173, doi:10.1561/1100000009.

²⁴⁹ Robin Burke, 'Hybrid Recommender Systems: Survey and Experiments', *User Modeling and User-Adapted Interaction*, 12.4 (2002), pp. 331–70, doi:10.1023/A:1021240730564.

seem the logical final frontier for recommender algorithms, but as Web 2.0 has matured and users have grown accustomed to having their free-time recommended out of existence, the semantic possibilities of Web 3.0 are beginning to be leveraged as the purveyors of algorithmic suggestion strive for new levels of personalisation in the mass spoon-feeding of digital content. Semantic recommender systems have been a hot research topic in the past few years, especially in the world of GLAMs and cultural heritage, a sector already adopting the Linked Open Data approaches needed to enable semantic suggestion²⁵⁰.

Cultural heritage institutions, particularly museums, have often tried to implement recommender systems to make their vast online and physical collections more easily accessible to users who may have limited time to enjoy them. Although the greater depth of information and context made available by Linked Open Data lends itself to this task, more standard systems (as described above) have been used or are still in use today²⁵¹. For example, the monitoring of visitor movements and time spent looking at different pieces of art can be used to suggest similar pieces that the visitor might want to view²⁵², or by looking at the routes other visitors have taken around a heritage site to give a visitor the optimal sensory experience²⁵³.

Semantic recommendation systems may have similar goals, although some focus on virtual interaction online rather than physically visiting the museum²⁵⁴, and the computation of similarity may still be achieved by comparing objects' or artworks' attributes, but the detail and complexity of information and reasoning

²⁵⁰ Wang Y (Yiwen), 'Semantically-Enhanced Recommendations in Cultural Heritage', 2011, doi:10.6100/ir694408.

²⁵¹ Liliana Ardissono, Tsvi Kuflik, and Daniela Petrelli, 'Personalization in Cultural Heritage: The Road Travelled and the One Ahead', *User Modeling and User-Adapted Interaction*, 22.1–2 (2012), pp. 73–99, doi:10.1007/s11257-011-9104-x.

²⁵² Svebor Karaman and others, 'Personalized Multimedia Content Delivery on an Interactive Table by Passive Observation of Museum Visitors', *Multimedia Tools and Applications*, 75.7 (2016), pp. 3787–3811, doi:10.1007/s11042-014-2192-y.

²⁵³ Georgios Alexandridis and others, 'Personalized and Content Adaptive Cultural Heritage Path Recommendation: An Application to the Gournia and Çatalhöyük Archaeological Sites', *User Modeling and User-Adapted Interaction*, 29.1 (2019), pp. 201–38, doi:10.1007/s11257-019-09227-6.

²⁵⁴ Alessio De Angelis and others, 'A Social Cultural Recommender Based on Linked Open Data', in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP '17* (ACM, 2017), pp. 329–32, doi:10.1145/3099023.3099092.

enabled by the underlying Linked Open Data models should enhance the ability to personalise suggestions. Semantic systems take the standard approach of the hybrid content-social approach and use the greater detail available to make more relevant recommendations; it is easier to attribute a user's appreciation of an object or artwork to a specific style, creator or colour palette as these are encoded in the Linked Open Data model, meaning that subsequent suggestions have more informed reasoning behind them²⁵⁵.

Whether taking a traditional approach to recommendation, and searching for related content based on keywords, or using a semantic system to suggest content that may be linked in far more nuanced ways, both of these types of recommender systems still rely on data from users' initial interactions with collections to and build up profiles of their likely interests²⁵⁶. But even though such systems aim to make visiting or browsing experiences more interesting by suggesting relevant items, especially using a semantic approach, their role is still in the vein of recommender systems used in other commercial fields: promote interest by suggesting based on previous interest. This makes it extremely difficult for a user or visitor to stumble upon something they had no idea they were interested in because it is in a completely unrelated area. Some recommender systems within cultural heritage are developed with elements of serendipity built into the algorithm²⁵⁷ and it seems likely that this will continue.

However, just as with the use of Linked Open Data per se, interoperability and access to the 'Linked Open Data cloud' – all the Linked Open Data available to institutions, developers and, eventually, users – can still pose problems for automation and the Semantic Web. Recommendation systems have proven to be one of the most popular forms of algorithmic automation for the cultural heritage industry, perhaps unsurprisingly given its huge amount of multimedia

²⁵⁵ Tuukka Ruotsalo and others, 'SMARTMUSEUM: A Mobile Recommender System for the Web of Data', *Journal of Web Semantics*, 20 (2013), pp. 50–67, doi:10.1016/j.websem.2013.03.001.

²⁵⁶ L. Deladienne and Y. Naudet, 'A Graph-Based Semantic Recommender System for a Reflective and Personalised Museum Visit: Extended Abstract', in 2017 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), 2017, pp. 88–89, doi:10.1109/SMAP.2017.8022674.

²⁵⁷ Tilman Deusel and others, 'Finding without Searching - A Serendipity-Based Approach for Digital Cultural Heritage', *Digital Intelligence*, 2014, p. 8.

data, but they are also making their way into other areas like healthcare²⁵⁸. This is because, on the semantic web, recommendation is easier than simply searching for something, which can be a daunting process for the uninitiated user trying to ask precisely the right question to navigate a path through the seething mass of a knowledge-graph overflowing with information and meaning. If the future of data storage, access and interrogation is semantic, automation will be indispensable to researchers, users and curators alike.

2.4.2 Algorithms and Computational Creativity

Recent developments in the field of computational creativity, which explores the use of AI and algorithms in areas of the arts such as creative writing and digital image production, have produced sophisticated algorithms capable of generating entire articles from a short prompt. One example, called GPT-2, was only released in full by its developers after a long debate over its ability to aid the spread of fake news²⁵⁹. However, it is currently the focus of experimentation to see whether historical figures can be ‘digitally resurrected’ through generation of new texts based on old notebooks and diaries²⁶⁰. Bots based on this algorithm are beginning to appear on social media, but it is unclear yet whether any exist dedicated to creatively reinterpreting or sharing history. On the artworks side of things, algorithms that can copy the style of one artwork onto another (neural style transfer), or generate an image having learnt associations from a large set of examples (GANs – generative adversarial networks), have already been used to create new artworks in the style of famous painters, such as ‘The Next Rembrandt’²⁶¹. Google Arts and Culture experimented with using such algorithms to enable users to apply a style of any artwork in their vast

²⁵⁸ Carlos Luis Sanchez Bocanegra and others, ‘HealthRecSys: A Semantic Content-Based Recommender System to Complement Health Videos’, *BMC Medical Informatics and Decision Making*, 17.1 (2017), p. 63, doi:10.1186/s12911-017-0431-7.

²⁵⁹ Aaron Mak, ‘When Is Technology Too Dangerous to Release to the Public?’, *Slate Magazine*, 2019 <<https://slate.com/technology/2019/02/openai-gpt2-text-generating-algorithm-ai-dangerous.html>> [accessed 6 April 2019].

²⁶⁰ Shawn Graham, ‘The Resurrection of Flinders Petrie’, 2019 <<https://electricarchaeology.ca/2019/10/23/the-resurrection-of-flinders-petrie/>> [accessed 6 April 2020].

²⁶¹ Werner Schweibenz, ‘The Work of Art in the Age of Digital Reproduction’, *Museum International*, 70.1–2 (2018), pp. 8–21, doi:10.1111/muse.12189.

online collections to a given photo²⁶². This offers new avenues for creative reinterpretation and exploration of online collections and promises to engage the viewing public in new ways.

However, these algorithms were just the beginning. The rapid progress in this area over the last couple of years culminated in 2023 with the realisation of 'foundation models' enabling 'multimodal generative AI'; basically single models are now powerful enough to create text, images and audio (potentially soon video and 3d objects too) on demand, with the user simply having to provide a description of what they would like creating (the 'prompt') by typing or talking. This has already led to debates around copyright law and sparked wider concerns over the place and prospects of human creatives, and many of these points are crucial to the future approaches the GLAM sector might take²⁶³. Understanding foundation models and generative AI is the first step in this process.

2.4.3 Foundation Models, Large Language Models, and Generative AI

The advent of foundation models marks a new chapter in the story of artificial intelligence (AI) and AI-hype. 2023 has been the year of 'ChatGPT', an AI tool (some call it a bot) that has thrust AI back into the forefront of public consciousness and seen growing debate around its potential uses and safety concerns, which has proven remarkably proficient at generating text, images and audio as part in a truly conversational manner²⁶⁴. Most people see ChatGPT

²⁶² 'Google's Arts & Culture App Taps AI to Transform Your Photos into Works of Art', *VentureBeat*, 2020 <<https://venturebeat.com/2020/04/02/google-arts-culture-app-taps-ai-to-transform-your-photos-into-works-of-art/>> [accessed 6 April 2020].

²⁶³ Susan Hazan, 'The Dance of the Doppelgängers: AI and the Cultural Heritage Community' (presented at the Proceedings of EVA London 2023, BCS Learning & Development, 2023), pp. 77–84, doi:10.14236/ewic/EVA2023.13; Magdalena Pasikowska-Schnass and Lim Young-Shin, *Artificial Intelligence in the Context of Cultural Heritage and Museums: Complex Challenges and New Opportunities* (European Parliamentary Research Service, 2023) <[https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2023\)747120](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)747120)> [accessed 18 November 2023].

²⁶⁴ Benj Edwards, 'ChatGPT Is One Year Old. Here's How It Changed the Tech World.', *Ars Technica*, 30 November 2023 <<https://arstechnica.com/information-technology/2023/11/chatgpt-was-the-spark-that-lit-the-fire-under-generative-ai-one-year-ago-today/>> [accessed 3 December 2023].

as the ‘AI’, but the ‘foundation model’ underpinning it, GPT-4 (Generative Pre-Trained Transformer 4), is where the algorithmic work actually happens²⁶⁵.

The journey of GPT models began with the creation of a non-profit research lab called OpenAI in 2015. OpenAI’s own extremely short history, tied in with the development of these GPT models, is another example of the how the history of the Internet and the Web has shaped the AI organisations of today. Founded by a group of computer scientists and Silicon Valley investors, including Elon Musk, with investment to the tune of \$1 billion, OpenAI started out with a research focus on general artificial intelligence in line with other organisations working in the area, most notably Google’s DeepMind. By 2018, this had progressed to the point where the GPT-1 model was introduced, performed better on text understanding tasks like question answering compared to previous approaches, and crucially it could do this generally across different types of task without having to be specifically trained for each one.

GPT was a new kind of ‘Transformer’ model that marked a new approach for computational text generation, largely thanks to a technique called ‘attention’ which allowed for more sophisticated prediction of what the next word would be in a sequence²⁶⁶. In older models, more akin to the predictive text function found on smartphones, the next word was often predicted based on the previous one, or sometimes previous few. ‘Attention’ takes each word in a sentence and compares its position to every other word in the sentence, building up a far more complex picture of which words are likely to appear closer together in the context of different sentences. Do this for a huge number of words across millions of sentences, and a statistical model of written language begins to emerge (at least, the written language which the model has been trained to predict).

The scale of this ‘training data’, effectively the amount of text from which the model was able to learn these statistical patterns, was crucial to achieving improvements in results. Although each successive upgrade to the GPT model would come with tweaks to the Transformer approach, perhaps the biggest

²⁶⁵ OpenAI, ‘GPT-4 Technical Report’ (arXiv, 2023), doi:10.48550/arXiv.2303.08774.

²⁶⁶ Ashish Vaswani and others, ‘Attention Is All You Need’ (arXiv, 2017), doi:10.48550/arXiv.1706.03762.

difference between GPTs 1 to 3 was the sheer volume of and different types of text used to train them. The release of GPT-2, a much larger model capable of generating far more coherent text than its predecessor, demonstrated the potential of transformer-based models for general natural language processing tasks²⁶⁷. This was primarily achieved by massively increasing the amount of training text, from 7,000 unpublished books for GPT-1 to over 8 million web pages of text from the social platform Reddit for GPT-2.

This transition is significant for a number of reasons. First and foremost, the sheer amount of text used to train GPT-2 resulted in much more coherent and generalised text generation capabilities than its predecessor. Add to this the source of the text, the social platform Reddit which is known to contain not insignificant amounts of hateful, biased, inappropriate and potentially harmful content, and the end result is a text generation model that OpenAI were hesitant about releasing to the public for fear of potential misuse²⁶⁸. Controversially, OpenAI only released a partial version of the model initially, waiting nine months before releasing the full version once concerns around safety had been somewhat allayed. However, the sheer amount of training text had also made it clear that continuing to develop such models was going to require significant amounts of computing power and money: it is estimated that GPT-2 cost \$256 per hour to train, likely for a period of weeks based on data from similar models²⁶⁹. During 2019, OpenAI moved from being a non-profit organisation to a capped-profit one in the hope of attracting more investment, which led to a highly lucrative and controversial \$1bn investment deal with Microsoft which

²⁶⁷ Alec Radford and others, 'Language Models Are Unsupervised Multitask Learners', 2019.

²⁶⁸ Alex Hern, 'New AI Fake Text Generator May Be Too Dangerous to Release, Say Creators', *The Guardian*, 14 February 2019, section Technology
<<https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction>> [accessed 7 December 2023]; Xavier Ferrer and others, 'Discovering and Categorising Language Biases in Reddit', *arXiv:2008.02754 [Cs]*, 2020
<<http://arxiv.org/abs/2008.02754>> [accessed 15 December 2020].

²⁶⁹ SYNCED, 'The Staggering Cost of Training SOTA AI Models', 27 June 2019
<<https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/>> [accessed 7 December 2023].

saw the two companies become partners and gave OpenAI access to Microsoft's enormous cloud computing resources through its Azure system²⁷⁰.

Less than a year later, the next iteration of GPT models was introduced to the world. GPT-3, a model effectively 100 times larger than GPT-2 and which had read trillions of words, began to demonstrate the capability for producing text across a generalised range of cases, from newspaper articles to computer code, that could convincingly have been written by a human. GPT-3 could also perform various tasks with few examples, without needing lots of extra training ('fine-tuning') as had been the case for GPT-2, highlighting the power of large-scale language models, both for better and worse²⁷¹. Although GPT-3 performed significantly better on a range of language tasks than its predecessors, this increased capability also made clearer than ever the inherent information and societal biases present in the model and its training data²⁷². To make this issue of bias even thornier, OpenAI was no longer 'open': in a new-found spirit of commercialism, the source code of the GPT-3 model was not released to the public, instead the model was exclusively accessible to Microsoft, meaning that its internal workings could not be picked apart by independent researchers seeking to understand such biases²⁷³.

GPT-3 also introduced a new problem for those seeking to understand it – an unimaginably vast amount of text constituted its training data. Just as the development of GPT-2 had required OpenAI to create its own dataset of text from the Reddit social platform, called 'WebText', so the development of GPT-3

²⁷⁰ OpenAI, 'Microsoft Invests in and Partners with OpenAI to Support Us Building Beneficial AGI', 2019 <<https://openai.com/blog/microsoft-invests-in-and-partners-with-openai>> [accessed 7 December 2023].

²⁷¹ Tom B. Brown and others, 'Language Models Are Few-Shot Learners', *arXiv:2005.14165 [Cs]*, 2020 <<http://arxiv.org/abs/2005.14165>> [accessed 1 April 2021].

²⁷² Luciano Floridi and Massimo Chiriatti, 'GPT-3: Its Nature, Scope, Limits, and Consequences', *Minds and Machines*, 30.4 (2020), pp. 681–94, doi:10.1007/s11023-020-09548-1; Li Lucy and David Bamman, 'Gender and Representation Bias in GPT-3 Generated Stories', in *Proceedings of the Third Workshop on Narrative Understanding* (presented at the NAACL-NUSE 2021, Association for Computational Linguistics, 2021), pp. 48–55, doi:10.18653/v1/2021.nuse-1.5; Conrad Borchers and others, *Looking for a Handsome Carpenter! Debiasing GPT-3 Job Advertisements* (arXiv, 23 May 2022), doi:10.48550/arXiv.2205.11374.

²⁷³ Karen Hao, 'OpenAI Is Giving Microsoft Exclusive Access to Its GPT-3 Language Model', *MIT Technology Review*, 2020 <<https://www.technologyreview.com/2020/09/23/1008729/openai-is-giving-microsoft-exclusive-access-to-its-gpt-3-language-model/>> [accessed 7 December 2023].

required an enormous amount of text data scraped from the Web. To achieve this, OpenAI used existing sources such as CommonCrawl, a dataset assembled by a non-profit organisation of the same name through the automated scraping of webpages using crawler bots, and deployed their own scraping crawler bots to extract as much text information as possible, before adding to this through digitised datasets of millions of books across all subjects and the entirety of Wikipedia. This mass data extraction from the Web was key to the model's capabilities, but has also become a source of contention as authors claim that such use of text from their books infringes copyright. Cases are ongoing at the time of writing, but OpenAI followed the lead of other Big Tech companies when announcing their latest update to ChatGPT by stating that any developer or paid customer who uses their AI model will have any copyright claim against them covered by OpenAI²⁷⁴. This also highlights copyright as one of the key areas of concern for GLAMs when thinking about digitising and digitalising collections, and using automated and AI approaches to sharing them online, especially as this is still a legal grey area²⁷⁵.

These huge training datasets and their importance for AI models also underline the role of a middle layer of non-profit 'research' organisations like CommonCrawl, not household names, whose data mining of the Web has laid the foundations for a slew of AI models generating text and images which can have important but subtle consequences when it comes to those models producing representations of the past. For example, Stable Diffusion is a popular image generation tool that was trained on the LAION-5b dataset of billions of images scraped from the web by the non-profit organisation

²⁷⁴ Blake Montgomery, 'OpenAI Offers to Pay for ChatGPT Customers' Copyright Lawsuits | Artificial Intelligence (AI) | The Guardian', *The Guardian*, 6 November 2023 <https://amp-the guardian.com.cdn.ampproject.org/v/s/amp.theguardian.com/technology/2023/nov/06/openai-chatgpt-customers-copyright-lawsuits?amp_gsa=1&_js_v=a9&usqp=mq331AQIUAKwASCAAgM%3D#amp_tf=From%20%251%24s&aoh=16993418795294&referrer=https%3A%2F%2Fwww.google.com&share=https%3A%2F%2Fwww.theguardian.com%2Ftechnology%2F2023%2Fnov%2F06%2Fopenai-chatgpt-customers-copyright-lawsuits> [accessed 18 November 2023].

²⁷⁵ Enrico Bonadio and Luke McDonagh, *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity* (Social Science Research Network, 2 June 2020) <<https://papers.ssrn.com/abstract=3617197>> [accessed 19 August 2020].

LAION²⁷⁶. These images were linked to text descriptions through ‘alt text’, a tag on a webpage which tells it what text to display when an image can’t be loaded, and this text is also used by screen readers for those who are visually impaired. This means that any image from a collection on a GLAM website will have likely been drawn up into the massive dataset, again raising issues around copyright for those online collections which are not Open Access, and that its description will have been determined by the alt text, sometimes itself auto-generated by an algorithm. Therefore, the AI model might have learned representations of images of cultural heritage artefacts, for example common objects like Roman coins, from places like eBay just as much as from museum websites, given the prevalence of modern replicas on eBay.

OpenAI first released ChatGPT in late 2022, followed by GPT-4 in 2023 and then a multimodal version of ChatGPT that encompasses GPT-4 and the image AI model DALLE-3 and the audio AI model Whisper, which are both based on GPT-4. Wrapping these together in one conversational platform, which is the current version of ChatGPT, marks the latest milestone in this journey²⁷⁷. A vital part of these upgrade models’ more sophisticated generation abilities was the introduction of Human Reinforcement Feedback Learning (HRFL), which basically involves getting humans to assess generated text from AI models to help train them to produce better and less biased answers. One of the main problems with GPT-3 was that the volume of text used to train it from the Web contained all the existing biases of Web content, which were then often amplified in hateful and harmful generations²⁷⁸. Therefore, for future models OpenAI paid people to do the jobs its AI models could not – filtering out harmful content from training data and then helping to retrain the models themselves to produce less biased outputs (although some research now suggests that the content generated by GPT-4 on average politically leans to the left)²⁷⁹. If nothing

²⁷⁶ Romain Beaumont, ‘LAION-5B: A NEW ERA OF OPEN LARGE-SCALE MULTI-MODAL DATASETS’, *LAION*, 2022 <<https://laion.ai/blog/laion-5b>> [accessed 8 December 2023].

²⁷⁷ OpenAI, ‘ChatGPT Can Now See, Hear, and Speak’, 2023 <<https://openai.com/blog/chatgpt-can-now-see-hear-and-speak>> [accessed 29 September 2023].

²⁷⁸ Borchers and others, ‘Looking for a Handsome Carpenter! Debiasing GPT-3 Job Advertisements’; Abubakar Abid, Maheen Farooqi, and James Zou, ‘Persistent Anti-Muslim Bias in Large Language Models’, *arXiv:2101.05783 [Cs]*, 2021 <<http://arxiv.org/abs/2101.05783>> [accessed 1 April 2021].

²⁷⁹ Paul F Christiano and others, ‘Deep Reinforcement Learning from Human Preferences’, in *Advances in Neural Information Processing Systems* (presented at the Advances in Neural

else, this highlights the often hidden roles humans play in the entangled sociotechnical ecosystems in which such AI models sit.

As explored in Chapters 5 and 6, these models hold great promise for the GLAM sector as novel ways to engage a vast audience with their collections in much more interesting ways than simply by sharing existing images or text descriptions in a fragmented way. However, these generative models still make mistakes, and their ability to misrepresent the past is just as powerful as their ability to represent it²⁸⁰. Nonetheless, these models are the blueprint for generative AI that competitors are already attempting to outdo, with Google and Amazon due to release their own versions shortly. It is perhaps the inevitable, commercialised AI chatbot development race, with lessening considerations on AI safety and ethics, that poses the biggest headaches for research, educational and GLAM institutions seeking to balance opportunities and challenges of these new approaches in the coming months and years²⁸¹.

While the development and implementation of such models can highlight specific challenges faced cultural heritage institutions in tackling data and content bias in the information available about the past, there are more fundamental biases present in the sociotechnical systems surrounding these models, as has been illustrated by work focussing on enabling GLAMs to decolonise collections and the information they contain²⁸². This illustrates the dual nature of technological systems in perpetuating colonial digital infrastructures and policies, such as copyright, whilst also offering solutions to help decolonise collections, such as facilitating participatory design practices within communities²⁸³. The potential of technology to aid in decolonisation

Information Processing Systems, Curran Associates, Inc., 2017), xxx, 1–9 <https://proceedings.neurips.cc/paper_files/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html> [accessed 20 July 2023]; Solon, ‘The rise of “pseudo-AI”’; Fabio Motoki, Valdemar Pinho Neto, and Victor Rodrigues, ‘More Human than Human: Measuring ChatGPT Political Bias’, *Public Choice*, 2023, doi:10.1007/s11127-023-01097-2.

²⁸⁰ Zihao Li, ‘The Dark Side of ChatGPT: Legal and Ethical Challenges from Stochastic Parrots and Hallucination’ (arXiv, 2023), doi:10.48550/arXiv.2304.14347.

²⁸¹ Nicola Jones, ‘What the OpenAI Drama Means for AI Progress — and Safety’, *Nature*, 623.7989 (2023), pp. 898–99, doi:10.1038/d41586-023-03700-4.

²⁸² Andrea Wallace, ‘Decolonization and Indigenization’, *Open GLAM*, 2021 <<https://openglam.pubpub.org/pub/decolonization/release/1>> [accessed 27 April 2024].

²⁸³ Siiri Paananen, Mari Suoheimo, and Jonna Häkkilä, ‘Decolonizing Design with Technology in Cultural Heritage Contexts - Systematic Literature Review’, in [] *With Design: Reinventing Design Modes*, ed. by Gerhard Bruyns and Huixin Wei (Springer Nature, 2022), pp. 1839–55,

efforts can also be seen in attempts to create a heritage management system to work with the ontologies and epistemologies of indigenous communities in Namibia²⁸⁴. These are just some examples of how wider, systemic biases can begin to be addressed with the help of technology, but it remains to be seen whether a decolonial large language model is currently possible or would even be helpful for GLAMs seeking to decolonise collections²⁸⁵.

As the history of the Web shows us, change does not happen in isolation and the network of complex sociotechnical ecosystems involved is already vast and complex. OpenAI, with its strong ties to Microsoft, may be the leader in the foundation model / generative AI race for now, but Amazon, Google and lesser-known names like Anthropic have also released their own foundation models, while organisations like Stable Diffusion and Midjourney have done the same for models focussing on generating images. On the non-profit side, BigScience, an organisation of volunteer researchers partly-funded by the French Government and supported by tech startup HuggingFace, have created an open-source, free-to-use Large Language Model called BLOOM that it hopes can democratise AI research and development²⁸⁶.

This also highlights the reality that foundation models are exclusive and expensive things to develop and run, as seen in OpenAI's controversial move from non-profit research lab to Microsoft-backed capped-profit organisation, and the vast majority of those in existence or under development (outside of China) are in the USA and either directly linked to, or with backing from, the Big Tech companies that own the required computing power and can attract the necessary investment: namely Microsoft, Amazon and Google. As generative AI forms the latest chapter in the history of automation, algorithms and bots on the Web, it highlights that even the earliest attempts to democratise access to knowledge about the past have always been influenced by underlying, long-

doi:10.1007/978-981-19-4472-7_119; Hacıgüzeller, Taylor, and Perry, 'On the Emerging Supremacy of Structured Digital Data in Archaeology'.

²⁸⁴ Shiningayamwe, 'Decolonizing Heritage Management Systems'.

²⁸⁵ Nan Yan and Cheng Xu, 'Decolonizing African NLP: A Survey on Power Dynamics and Data Colonialism in Tech Development', 2024 <<https://openreview.net/forum?id=D698BEfwgv>> [accessed 29 April 2024].

²⁸⁶ Melissa Heikkilä, 'Inside a Radical New Project to Democratize AI', *MIT Technology Review*, 2022 <<https://www.technologyreview.com/2022/07/12/1055817/inside-a-radical-new-project-to-democratize-ai/>> [accessed 3 December 2023].

standing geographical and historical biases than underpin the existing inequalities seen in the Web today.

In summary, bots and algorithms have become pervasive in the current social era of the World Wide Web, both behind the scenes in determining which information is most easily found and how much is preserved, and increasingly at the forefront of the Web via interactions with human users across social media and on apps on mobile devices. Whilst recent research has focused on the growing concerns around this growing automated influence with regard to global politics, this literature review establishes the importance of each of these areas of online automation for the collection, curation and dissemination of historical information, images and objects. With social bots, chatbots and recommendation systems becoming ever more popular it is crucial to survey how each of these areas are currently influencing history and heritage online and what the future may hold in each case, especially as the heritage sector seeks to make best use of all things 'digital' and explore the potential of AI and algorithms to increase and personalise public engagement. Each of these gaps will be further explored in the case studies that follow.

3. Case Study: Bots Sharing GLAM Collections on Twitter

Social media is an important online platform for engagement and interaction for academic history and heritage, especially official accounts representing popular museums and other GLAMS, with a growing literature dedicated to its role in the study and dissemination of historical information¹. Although primarily a manual, human activity, whether carried out by expert or non-expert users, the rise of automation online and on social media has seen social bots come to play significant roles in disseminating information, of varying degrees of accuracy, across such online communities, especially in the political realm².

The previous chapter introduces key terminology related to bots on social media, highlighting the complexity of the term 'social bot'. For this case study, social bots were defined as automated social media accounts, with Twitter bots specifically referring to those found on the Twitter platform (now called 'X'). The chapter discusses how these bot accounts have influenced the sharing of artworks and artefacts from online GLAM collections, focussing on the impact of this for the different user groups involved, including developers, curators and users viewing the content on social media. Whilst social bots are the subject of existing research in the areas of cybersecurity and political misinformation, this chapter aims to fill a gap by considering their activity in the GLAM sector.

This case study was carried out in 2019, and much has changed between then and the final submission of this thesis in late 2023. In June 2023 the Twitter platform was rebranded and became X, following Elon Musk's takeover of Twitter in October 2022. However, for the sake of accuracy, it will be referred to as 'Twitter' throughout this chapter, given that this shows work done in 2019. Similarly, the 'Off The Easel' community of Twitter bots has since been renamed to 'Bot Frens'³, but again will be referred to throughout by its original name as of 2019. As a result of Twitter's takeover and transition to X, numerous

¹ *Heritage and Social Media: Understanding Heritage in a Participatory Culture*, ed. by Elisa Giaccardi, 1st ed (Routledge, 2012).

² Howard, Woolley, and Calo, 'Algorithms, bots, and political communication in the US 2016 election'.

³ BotFrens, 'BotFrens', 2023 <<https://botfrens.com/>> [accessed 8 December 2023].

changes have occurred on the platform, from stopping free use of the API to another attempted crackdown on automated accounts (while reinstating previously banned human ones) to potential user charges, all of which have already had an impact and seen many users leave the platform. The potential implications of these recent developments are explored in more detail in the broader discussion at the end of this chapter.

This survey sought to explore the landscape (as of 2019) of social bots disseminating history and cultural heritage (including historical artworks), such as sharing part of the British Library's collections⁴ or the digitally-available works of Van Gogh⁵, taking the most popular social media platforms on which bots are found, namely Twitter, Facebook, Tumblr and Reddit as a starting point. Pilot work identified two main types of social bots sharing content related to history and cultural heritage to be explored: [1] individual automated accounts that are not connected to others; and [2] bot accounts that are members of a coordinated community focusing on art and cultural heritage called Off The Easel. The vast majority of individual bots were found solely on Twitter, whilst Off The Easel bots were present and active on Twitter and Tumblr, sharing the same content across both platforms. As Twitter is the most commonly used platform for automated accounts and is the most permissive in terms of data collection, this survey restricts itself just to the groups of bots found on Twitter.

Figure 4 shows four example accounts covering the main different types of Twitter bots identified: @artistmonet, an artist fan account within the Off The Easel botnet (top left); @met_medievalart, a cultural heritage-focussed account also within the Off The Easel botnet (top right); @DeathMedieval, an individual history bot account sharing excerpts from Medieval coroners' rolls (bottom left); @MechCuratorBot, the official bot for the British Library sharing book illustrations.

⁴ '(20) Mechanical Curator (@MechCuratorBot) / Twitter', *Twitter* <<https://twitter.com/mechcuratorbot>> [accessed 28 October 2019].

⁵ 'Vincent van Gogh (@vangoghartist) / Twitter', *Twitter* <<https://twitter.com/vangoghartist>> [accessed 24 May 2020].

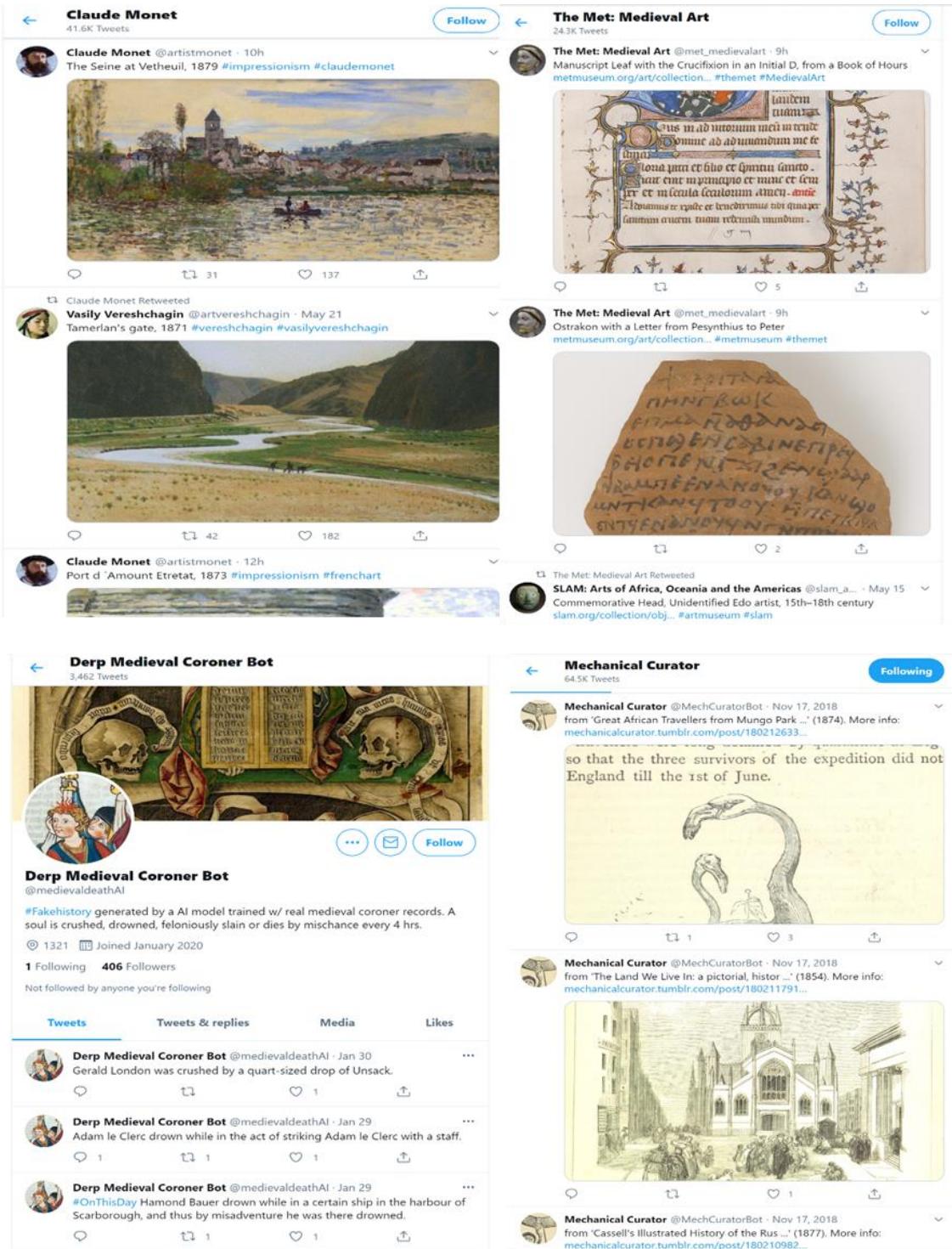


Figure 4: Examples of Twitter bot accounts (from top left to bottom right): @artistmonet (Off The Easel); @met_medievalart (Off The Easel); @DeathMedieval, (individual bot account); @MechCuratorBot (The British Library).

It is important to note that similar dissemination activities have been and are carried out by humans across accounts. These tend to provide a more historical context and detail to their content. Some of the approaches adopted, for

instance tweeting about events that happened ‘on this day’, are almost semi-automatic and so developing bots to fully-automate this activity is a natural progression, despite the loss of nuance and context that full automation often involves. This can prove confusing for the user and some accounts, such as @AZSportsHistory, specifically declare their humanness to avoid confusion. It should also be noted that this study has looked solely at bot accounts related to ‘history’ and ‘cultural heritage’, but there are other bots and humans out there in related fields, such as @MythologyBot, an account born out of a literature project examining styles and motifs in myths and folk tales.

A brief initial survey identified a variety of relevant social bots acting individually, mainly on Twitter, and a coordinated community of social bots, known as the Off The Easel botnet, sharing art and cultural heritage artefacts across Twitter, Facebook and Tumblr. A full survey was then done to identify all relevant bots across social media platforms and cover the full extent of the ‘Off The Easel’ bot community. The survey sought to identify the number and type of bots present, their approaches to sharing information and the type and range of content relating to history and GLAM collections that they shared.

It was important to assess the growing role of AI and automation in sharing historical and cultural heritage content across social media so that academics, heritage professionals and developers are aware of the advantages and disadvantages of such approaches for engaging an online public with the past.

3.1 Survey of Bots Sharing History and Cultural Heritage on Twitter

The research methodology followed a structured search strategy, focussed on the terms ‘history bot’, ‘heritage bot’, ‘cultural heritage bot’, ‘art bot’ and ‘museum bot’, which was carried out on Twitter. Data collection was achieved programmatically through the Twitter API after bots had been manually identified. Having employed the above systematic search strategy, it was also possible on Twitter, thanks to its lists function which allows users to create lists of accounts which they follow for others to view, to find groups of bots belonging to a list. By far the largest, and most easily verifiable, lists of relevant bots are

managed by a platform called Off The Easel which hosts and controls a large and growing number of bots disseminating art ('art bots'⁶) and items from museum collections ('museum bots'⁷). All of these bots are described as such in their user profiles, however not all automated Twitter accounts are so forthcoming. In some cases it was very difficult to verify whether an account was automated, run by a human or a mixture of the two (a 'cyborg'), even when following guidelines developed for this very purpose⁸. In these ambiguous instances the accounts were not included in the dataset. Bots that tweeted in languages other than English (for example, a bot tweeting out images with descriptions from the collection of the National Museum of China) were included in the dataset and the textual content analysis, as it is still possible to see how many times the content of a tweet is repeated over time, even though its meaning is hidden. Automatic translation of such tweets was not attempted as it was still too inaccurate for use in academic research, despite the impressive progress made in this area⁹.

Whenever a bot was identified its Twitter username was recorded. The list of collected usernames was used as the starting point for automated scraping, through the Twitter API, of the most recent sample of tweets (in the range of 3200-3250 tweets as determined by the Twitter API limits) posted by each bot from its creation up until 06/02/2021 (the date of data collection). Although collecting a sample in this way does not reflect the full tweet corpus for each bot, the most recent sample is still representative of a bot's activity which will not significantly differ over time in terms of content given its automated nature. Along with the text of each tweet and its number of associated retweets, hashtags and any media (e.g. images), the bot's location (as textually described, not geolocated), number of followers, date of creation and

⁶ '@andreitr/Art Bots on Twitter / Twitter', *Twitter* <<https://twitter.com/andreitr/lists/art-bots>> [accessed 1 October 2019].

⁷ '@andreitr/Museum Bots on Twitter / Twitter', *Twitter* <<https://twitter.com/andreitr/lists/museum-bots>> [accessed 1 October 2019].

⁸ Zafar Gilani and others, 'Of Bots and Humans (on Twitter)', in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17* (presented at the the 2017 IEEE/ACM International Conference, ACM Press, 2017), pp. 349–54, doi:10.1145/3110025.3110090.

⁹ Michael Groves and Klaus Mundt, 'Friend or Foe? Google Translate in Language for Academic Purposes', *English for Specific Purposes*, 37 (2015), pp. 112–21, doi:10.1016/j.esp.2014.09.001.

description were all recorded. All the usernames of followers for each bot were also collected so that network analysis could be undertaken to reveal audience demographic, interaction patterns and engagement.

To gain a rough idea of audience demographic, followers' profile descriptions were read and categorised into groups by profession or interest, where this was possible to decipher (some were left blank). This was only done for highly engaged followers of the Off The Easel bots, that is those who followed 10% or more of the bots in total. This ensures that these followers actually engage with Off The Easel in a meaningful way, rather than just following one or two bots and never actually engaging with their content. It also reduced the number of profiles to be manually parsed to a manageable number in the hundreds rather than the hundreds of thousands, making this a far more feasible and ethically acceptable approach. Such demographic profiling was not possible for the individual Twitter bots given their lack of coordinated community, which made identifying a representative sample of truly engaged followers across all bots unfeasible.

The data collected on all identifiably relevant bots was first analysed using descriptive statistics to compare the aforementioned Twitter metrics for each bot. Then, content analysis was carried out on the textual data from all the tweets before the overall system of connections between bots and humans was explored through network analysis techniques (see Figure 5 for an overview of the whole process). This methodology received ethical approval from the FASS and LUMS Research Ethics Committee at Lancaster University.

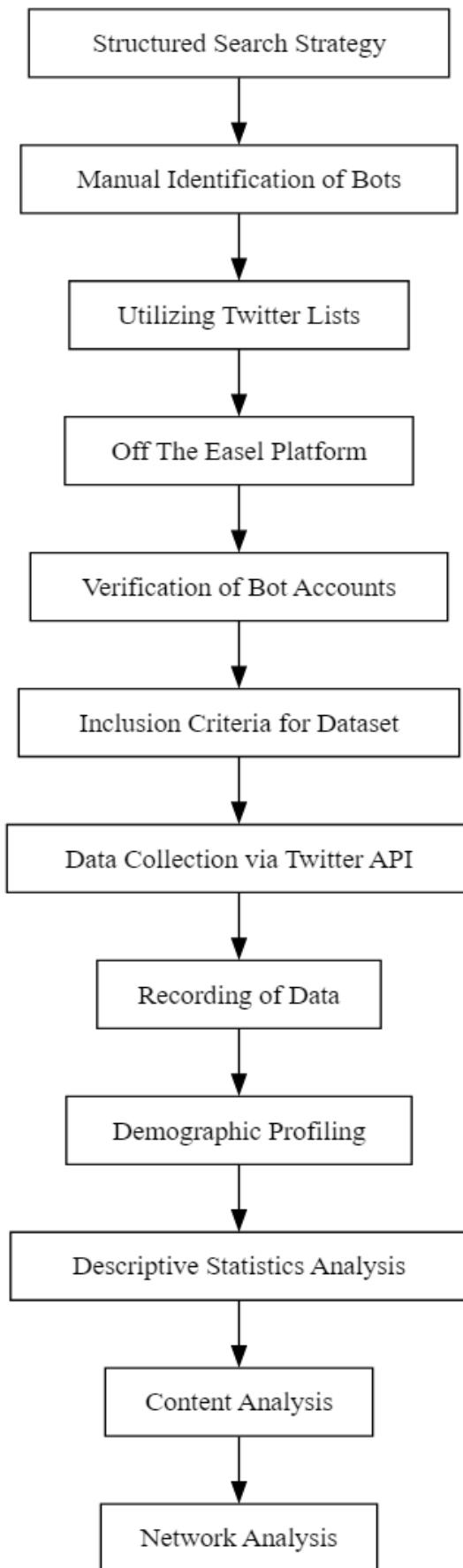


Figure 5: Overview of the Twitter bot survey data collection and analysis process

Given the highly repetitive nature of automated tweets, content analysis was primarily done via close reading informed by broader corpus linguistics techniques, such as word frequency and collocation analysis, performed using Voyant Tools¹⁰ on all tweets to see if any patterns emerged in the content itself. Encompassing a set of techniques growing in popularity within the digital humanities, corpus linguistics “is an area which focuses upon a set of procedures, or methods, for studying language”¹¹. The power of corpus linguistics lies in its ability to analyse very large bodies of text (‘corpora’) using both quantitative and qualitative approaches to discern patterns that would be impossible for a human reader to mentally process, such as in this case where the corpus consists of nearly one million tweets. The patterns and details that emerge may then be subjected to more detailed content analysis through close-reading. This mixture of close and distant reading is a hallmark of corpus linguistics methodologies and is effective in picking out details and trends from large amounts of textual data¹². For this case study, the aim is to see what the art bots tweet about, but also whether the textual style they use influences how users engage with their content.

To initially explore the broader patterns metrics such as frequency and dispersion are used alongside collocation analysis. The entire corpus of tweets collected from all bots was arranged into a series of documents corresponding to each bot so that all the tweets for a given bot were stored, line by line, in one text file. This allows for comparison of how many times a word appears both overall, its frequency, and in each document, its dispersion, to highlight common themes across all the bots and topics specific to individual bots. For instance, it would be expected that the word ‘impressionism’ would appear far more frequently in the texts of art bots sharing impressionist paintings. Collocation analysis takes this approach a step further by showing how

¹⁰ Stefan Sinclair and Geoffrey Rockwell, ‘Voyant Tools’ (Voyant Tools, 2016) <<https://voyant-tools.org/>> [accessed 22 March 2020].

¹¹ Tony McEnery and Andrew Hardie, *Corpus Linguistics: Method, Theory and Practice*, Cambridge Textbooks in Linguistics (Cambridge University Press, 2012), p. 1.

¹² McEnery and Hardie, ‘Corpus linguistics’.

commonly words are associated with each other across the whole corpus¹³, with the aim of highlighting patterns in content shared.

Network analysis is used across a wide variety of disciplines, from biological science (to explore gene interactions) to economics (to explore trade networks) and sociology (to explore connections between individuals), and is therefore flexible as an approach for finding, analysing and visualising patterns in all kinds of data, including that of networks of users on social media platforms¹⁴. This approach is becoming more widely used within the field of Digital Humanities as it enables interrogation of extremely complex systems of relationships and connections between entities both physical and virtual, contemporary and historical¹⁵. The main aim of network analysis in the context of this study is to describe the structure and function of the art, history and heritage social bot community on Twitter.

It is important to note that the statistical analyses carried out and the visualisation of these analyses, and of the network itself, are both complex, interrelated research processes that change perceptions of both researchers and readers; visualisations can help make sense of complex sets of datasets, but are also subject to the visual choices of their designer and/or analyst, meaning that they are not neutral or objective ways of seeing information¹⁶. Cytoscape¹⁷, and Gephi, two freely available network analysis and visualisation tools primarily used in the sciences and humanities respectively, were used because of their in-built range of statistical analysis options and substantial catalogue of add-in applications to aid with visualisation of complex networks.

¹³ Vaclav Brezina, Tony McEnery, and Stephen Wattam, 'Collocations in Context: A New Perspective on Collocation Networks', *International Journal of Corpus Linguistics*, 20.2 (2015), pp. 139–73, doi:10.1075/ijcl.20.2.01bre.

¹⁴ Ulrik Brandes, *Network Analysis: Methodological Foundations* (Springer Science & Business Media, 2005).

¹⁵ Deryc T. Painter, Bryan C. Daniels, and Jürgen Jost, 'Network Analysis for the Digital Humanities: Principles, Problems, Extensions', *Isis*, 110.3 (2019), pp. 538–54, doi:10.1086/705532.

¹⁶ Lev Manovich, 'What Is Visualisation?', *Visual Studies*, 26.1 (2011), pp. 36–49, doi:10.1080/1472586X.2011.548488; D'Ignazio and Klein, 'Data Feminism', p. 75.

¹⁷ Paul Shannon and others, 'Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks', *Genome Research*, 13.11 (2003), pp. 2498–2504.

The network analysis of the art, history and heritage social bot community on Twitter took the following approach. Each bot account in the network is known as a 'node', viewed as a circular point on the network graph, and a connection between two bot accounts is called an 'edge', shown as a line connecting points on the graph (see Figure 19). With hundreds of bots and thousands of connections, this network is large and complex. Therefore, exploring its structure first requires examining the extent to which the whole bot community, or network, is made up of smaller communities, or 'subnetworks'. This entails the use of a statistical measure called 'modularity' which determines how well connected a node is to its neighbours, and then to their neighbours and so on, giving an overall number of how many communities exist within in the whole¹⁸. This structure can then be visualised through colour-coding of the communities to show clusters of similar bots over the whole network. The overall aim is to highlight similarities between nodes in a given community which will be useful to test if there are enough similarities between art bots sharing the same style of art to represent some of the historical, real-world artistic networks of the past.

One of the aims of examining the Off The Easel botnet within a wider social bot community on Twitter is to see if its algorithmically determined and coordinated network allows the individual art bots within it to wield greater influence in exposing users to art and artefacts than the individual Twitter bots. Measuring influence in networks is primarily achieved through centrality measures, statistical tests which quantify how densely, diversely and deeply connected a given node is¹⁹. In this case three centrality measures were used: betweenness centrality, which measures the ability of a node (in this case, bot) to connect disparate parts of the network, i.e. very different communities of art bots; closeness centrality, which quantifies how closely connected a bot is to others in the network; and eigenvector centrality, which measures how well-connected a bot is to other well-connected, i.e. influential, bots. Eigenvector centrality is most useful single measure of influence, but combining the three statistics

¹⁸ Vincent D. Blondel and others, 'Fast Unfolding of Communities in Large Networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2008.10 (2008), p. P10008, doi:10.1088/1742-5468/2008/10/P10008.

¹⁹ Francis Bloch, Matthew O. Jackson, and Pietro Tebaldi, 'Centrality Measures in Networks', 2016 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2749124> [accessed 9 June 2017].

shows different aspects of influence across the network covering how able a social bot is to share its content with users who follow other bots in the network, and conversely how likely it is to highlight the content of other bots to its followers.²⁰ When considering the ethics of scraping vast amounts of information from social media platforms like Twitter, the last few years have seen a greater awareness and identification of ethical issues around researching Twitter and its users, along with more comprehensive sets of practical guidelines for approaching this potentially sensitive and grey area²¹. Compared to Facebook, Twitter is a primarily public platform its privacy policy stating that:

Twitter is public and Tweets are immediately viewable and searchable by anyone around the world. We give you non-public ways to communicate on Twitter too, through protected Tweets and Direct Messages. You can also use Twitter under a pseudonym if you prefer not to use your name.²²

There have been issues raised around tweets posted by vulnerable users or those in other situations where the context of the tweet may be lost or (unintentionally) misrepresented²³, however the only content collected originating from individual users in this study comes in the form of retweets of automated accounts, usually without any added text, and the cultural and artistic content being shared is not that of a potentially sensitive topic area, such as health, that demands particular ethical scrutiny²⁴. The only user data collected were usernames (usually pseudonyms) and locations (entered by the user and not necessarily accurate or truthful). Usernames are pseudonymised upon collection and never associated with a location. This work was approved by the FASS and LUMS Research Ethics Committee at Lancaster University.

²⁰ Himansu Sekhar Behera and others, 'Finding Correlation Between Twitter Influence Metrics and Centrality Measures for Detection of Influential Users', in *Computational Intelligence in Data Mining: Proceedings of the International Conference on ICCIDM 2018* (Springer, 2019).

²¹ Wasim Ahmed, Peter A. Bath, and Gianluca Demartini, 'Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges', 2017, doi:10.1108/S2398-601820180000002004.

²² 'Privacy Policy' <<https://twitter.com/content/twitter-com/legal/en/privacy.html>> [accessed 14 October 2019].

²³ Ahmed, Bath, and Demartini, 'Using Twitter as a Data Source'.

²⁴ Association of Internet Researchers, *Internet Research: Ethical Guidelines 3.0 Association of Internet Researchers*, 2019 <<https://aoir.org/reports/ethics3.pdf>> [accessed 5 March 2020].

3.2 Individual Twitter Bots and Their Approach to Sharing Content

This section focusses on the 139 currently active Twitter bots that are not part of the Off The Easel bot community. The results presented here relate to the content shared and its sources, the bots' audiences and levels of engagement, the developers and their aims and the ways in which the bots operate both in comparison to each other and standard human accounts.

The analysis shows that the majority of individual Twitter bots share content available from Open Access online collections of GLAMs, primarily larger, better-known institutions situated in Western Europe and the USA. The bot accounts are often created by lone developers, or occasionally on behalf of organisations like the British Library by in-house staff, but a significant proportion are then poorly maintained or completely abandoned. In contrast to the Off The Easel bot community approach, although individual bots can have large numbers of followers and high engagement with their audiences, overall these bots rarely, if ever, interact with each other, so the combined sharing power of all the individual bots is lower than that seen in the coordinated approach of the Off The Easel bots community.

Individual Twitter bots share a wide range of content and are more varied in their approach than the coordinated Off The Easel bots. Categorising them is therefore less straightforward, but as Figure 2 shows they broadly fall into three areas: bots that share content from GLAMs, bots that share content from other online sources and bots that generate or remix content from GLAMs or other sources. Bots sharing and/or remixing content from GLAMs, particularly museums, make up the bulk of the dataset (see Figure 6), whilst those that share textual content, such as AI-generated causes of death derived from Medieval coroners' rolls, are in the minority. The sources of content echo this trend with the majority coming from open access GLAM collections.

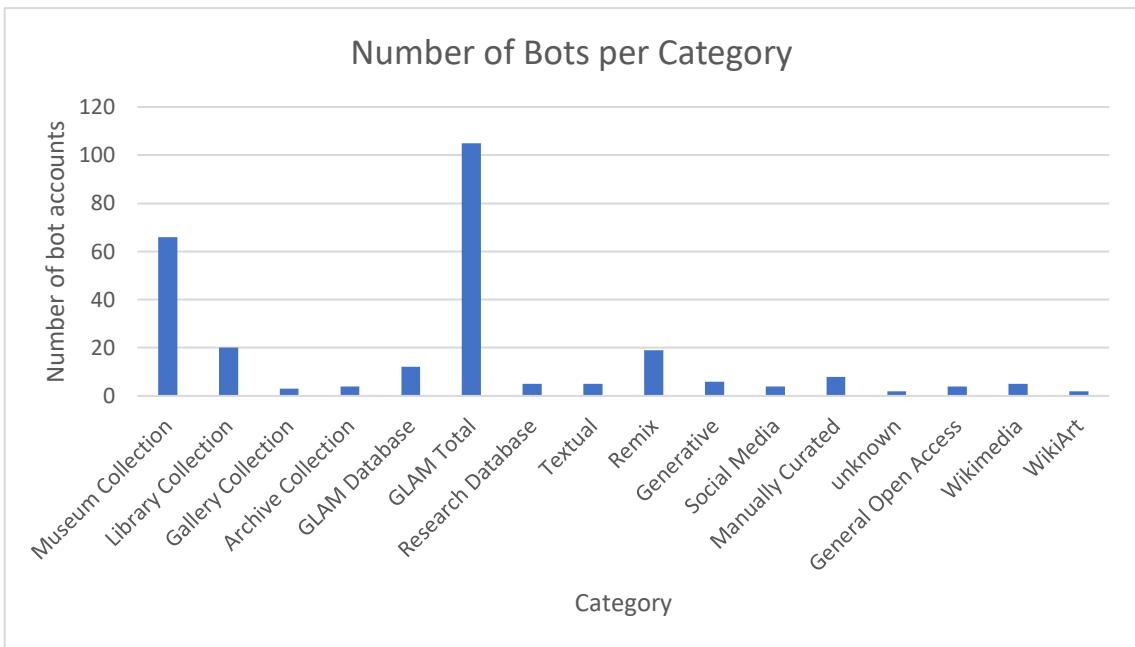


Figure 6: Bar chart showing categorisation of Twitter bots by sharing approach and content source.

To assess the historical and cultural diversity of the content being shared, bots were categorised in three ways: firstly, where the institutions, collections or historical content they represent is physically located (most pertinent for those sharing content from GLAMs – Figure 7); secondly, the regions represented in the content shared (Figure 8); and thirdly, the historical time period represented in the content shared (this does not always match up to time periods as defined in the art world, for instance ‘Modern’ covers anything post 18th-Century, not just ‘modern art’ – Figure 9).

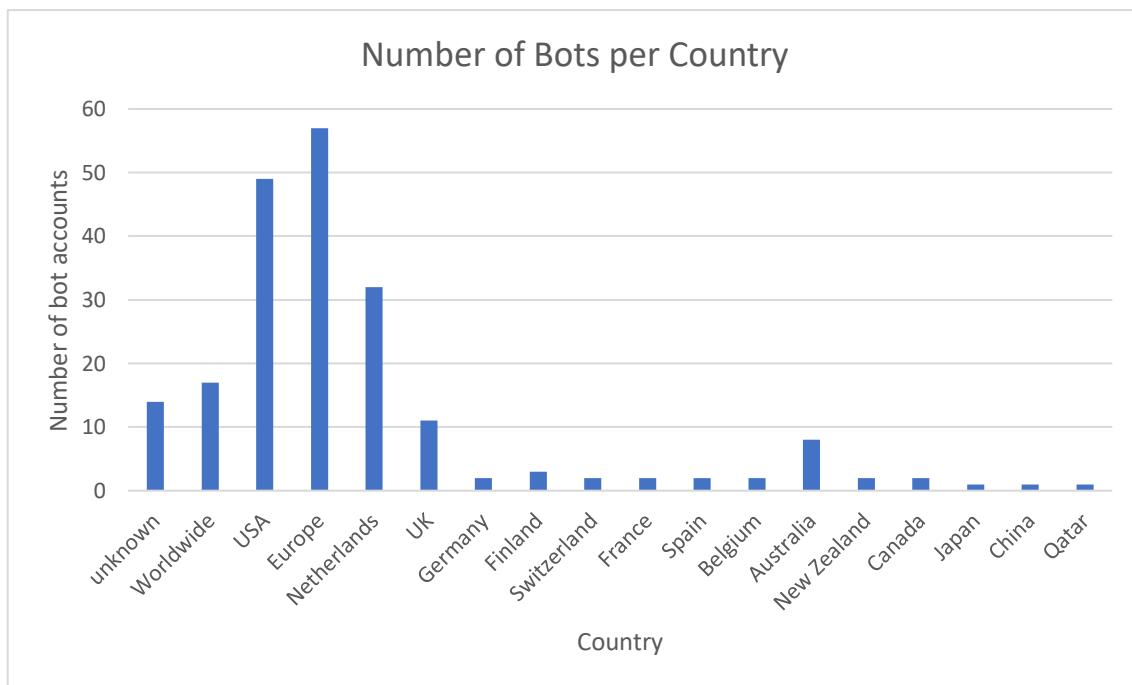


Figure 7: Bar chart showing locations of institutions, collections or textual content shared by Twitter bots.

The dominance of Western GLAM institutions as the collectors and sources of content shared by the individual Twitter bots is clear (see Figure 7). In a similar vein, Figure 8 shows the regional spread of the content shared by the bots; this is an important distinction as many GLAMs are based in Western countries but house collections from other parts of the world. This is demonstrated by the largest proportion of content covered by 'Worldwide' and the two highest specific regions being North America and Europe, highlighting that open access data primarily comes from Western institutions and represents Western art, cultural heritage and history. This is only amplified when considered in the context of the problematic colonial aspects of museum collections, which digitisation and online sharing only serve to complicate and potentially enhance (see Wallace and Pavis (2019) for a thorough discussion)²⁵.

²⁵ Mathilde Pavis and Andrea Wallace, *Response to the 2018 Sarr-Savoy Report: Statement on Intellectual Property Rights and Open Access Relevant to the Digitisation and Restitution of African Cultural Heritage and Associated Materials* (Rochester, NY: Social Science Research Network, 25 March 2019) <<https://papers.ssrn.com/abstract=3378200>> [accessed 28 October 2019].

As an example, one of the main sharers of non-Western art is Ukiyo-e bot²⁶ which shares Japanese woodblock prints from an online platform²⁷ that developed out of a research project to make these prints findable. However, examining the sources used to compile the database reveals the fact that most of these woodblock images are held in Western museums. The bot tweets out each image with its title, artist and a link to the database listing, so that anyone sufficiently interested can click and find out where an image is held, but for those merely scrolling this wider context of colonial collecting is lost. This applies to the vast majority of GLAM accounts across both the individual and Off The Easel datasets.

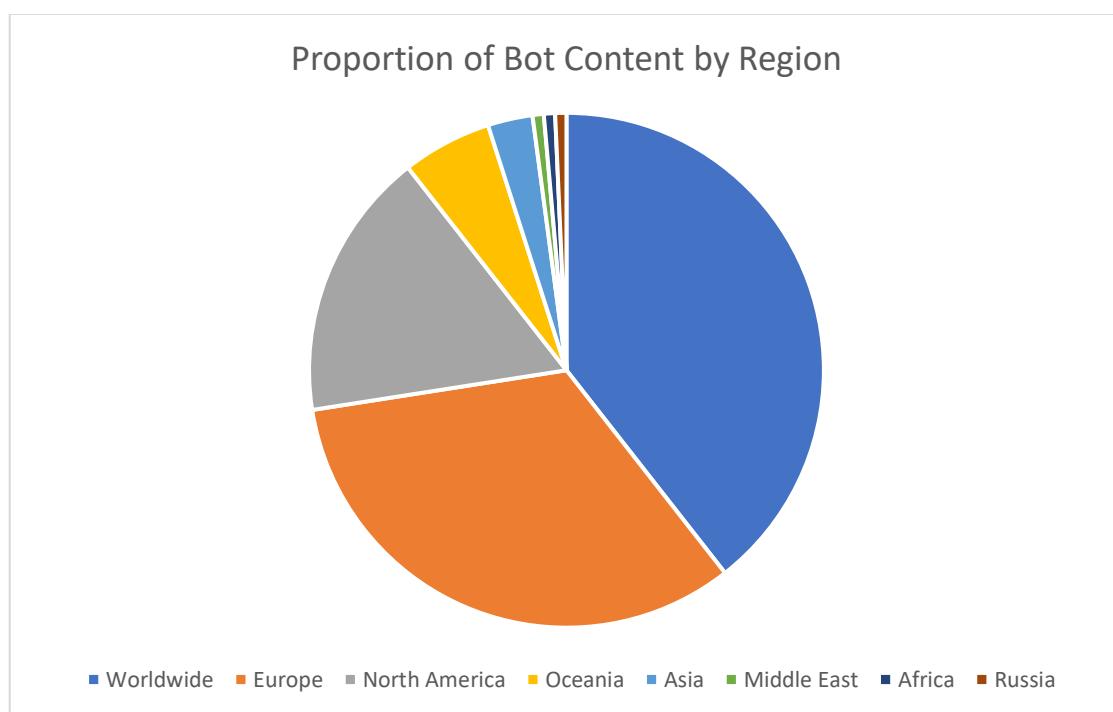


Figure 8: Pie chart showing regions covered by the content shared by bots (Worldwide means that a bot account shares content from multiple regions).

Representativeness of content in terms of historical time periods covered is skewed towards the modern end of the spectrum, although many bots sharing content from GLAM collections cover a range of time periods, as seen in the

²⁶ '(20) Ukiyo-e Bot (@UkiyoeBot) / Twitter', *Twitter* <<https://twitter.com/ukiyoebot>> [accessed 28 October 2019].

²⁷ 'Japanese Woodblock Print Search - Ukiyo-e Search' <<https://ukiyo-e.org/>> [accessed 2 March 2021].

large proportion of 'Mixed' period content in Figure 9. 'Modern' covers content shared from specific GLAMs, some specialising in Modern Art, but also most content from North American and Australian libraries and archives, such as postcards from the New York Public Library or newspaper reports from Trove, Australia's primary GLAM database site. This kind of more modern historical content is often more plentiful, easily accessible and easier to digitise. Couple this with drives to increase the re-use of Open Access content, such as the New York Public Library's call for users to remix their digitally available collections²⁸, and the prevalence of such modern historical content being shared by Twitter bots, seen as novel media that sit comfortably within the remit of digital sharing and remixing, is hardly surprising.

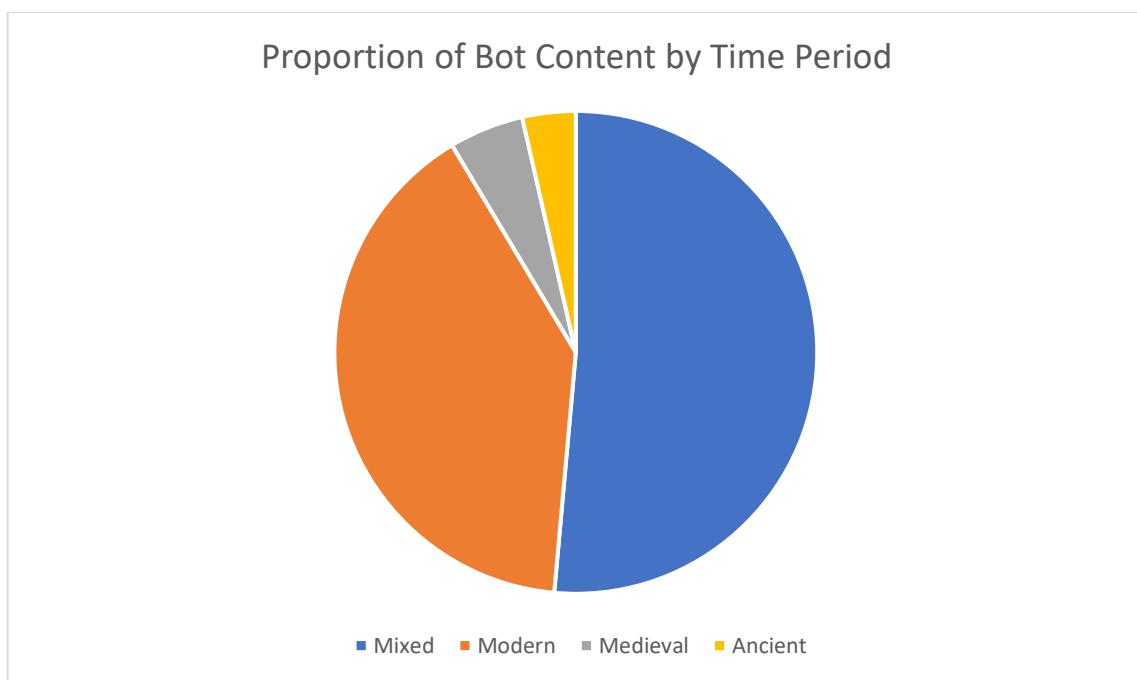


Figure 9: Pie chart showing the historical time period covered by the content shared by bots (Mixed means that a bot account shares content from multiple time periods).

Content analysis using corpus linguistics methods gives a clear sense of the overall mix of content shared by the individual Twitter bots, with markers of historicity and art-related terms being prevalent (see Figure 10). Most pieces of text will serve as captions for images, either of artworks, artefacts or historical

²⁸ '#nyplremix: Get Creative With the Public Domain', *The New York Public Library*, 2016 <<https://www.nypl.org/blog/2015/01/06/creativity-public-domain>> [accessed 2 March 2021].

documents, but a few are the sole source of information, especially for those bots that remix textual sources or generate their own tweet texts algorithmically. The terms with the highest relative frequency across all bots - 'century', 'years', 'ago' and 'today' – reflect the historical nature of the content shared, whether artworks, cultural artefacts or pieces of text. The strong links between 'century', 'early' and 'late' also shows consideration for dating and periodisation of material. The overall skew towards artworks shared from GLAM collections is shown by the prevalence of the terms 'nga' and 'portrait', 'nga' representing the National Gallery of Art in Washington. The art-related content tends towards the same general trends as that seen in the 'Off The Easel' corpus (see below), with 'woman', 'landscape' and 'drawing' featuring as terms commonly mentioned across the corpus (see Figure 10 blackberry plot (bottom right)).

Audience engagement can help to understand how successful Twitter accounts are in getting users to read and interact with the content they share, which applies to human and automated accounts. For the individual Twitter bot accounts, engagement was measured using the standard metrics of followers and impressions, i.e. the number of retweets and likes bots' tweets receive. These metrics cover different periods because of differences in posting frequency across a given sample of ~3200 tweets and the varying active lifespans of these bots, 44% of which are currently inactive (have not tweeted in the last two months). This wider issue is discussed in more detail in the following section, but often boils down to developers being unable or unwilling to maintain bots for prolonged periods of time. Account bans or suspensions are also common problems; DeathMedieval, an extremely popular Twitter bot that had more than 60,000 followers, was recently suspended and is not included in the dataset. Table 3 shows the number of followers and impressions standardised as a monthly average.

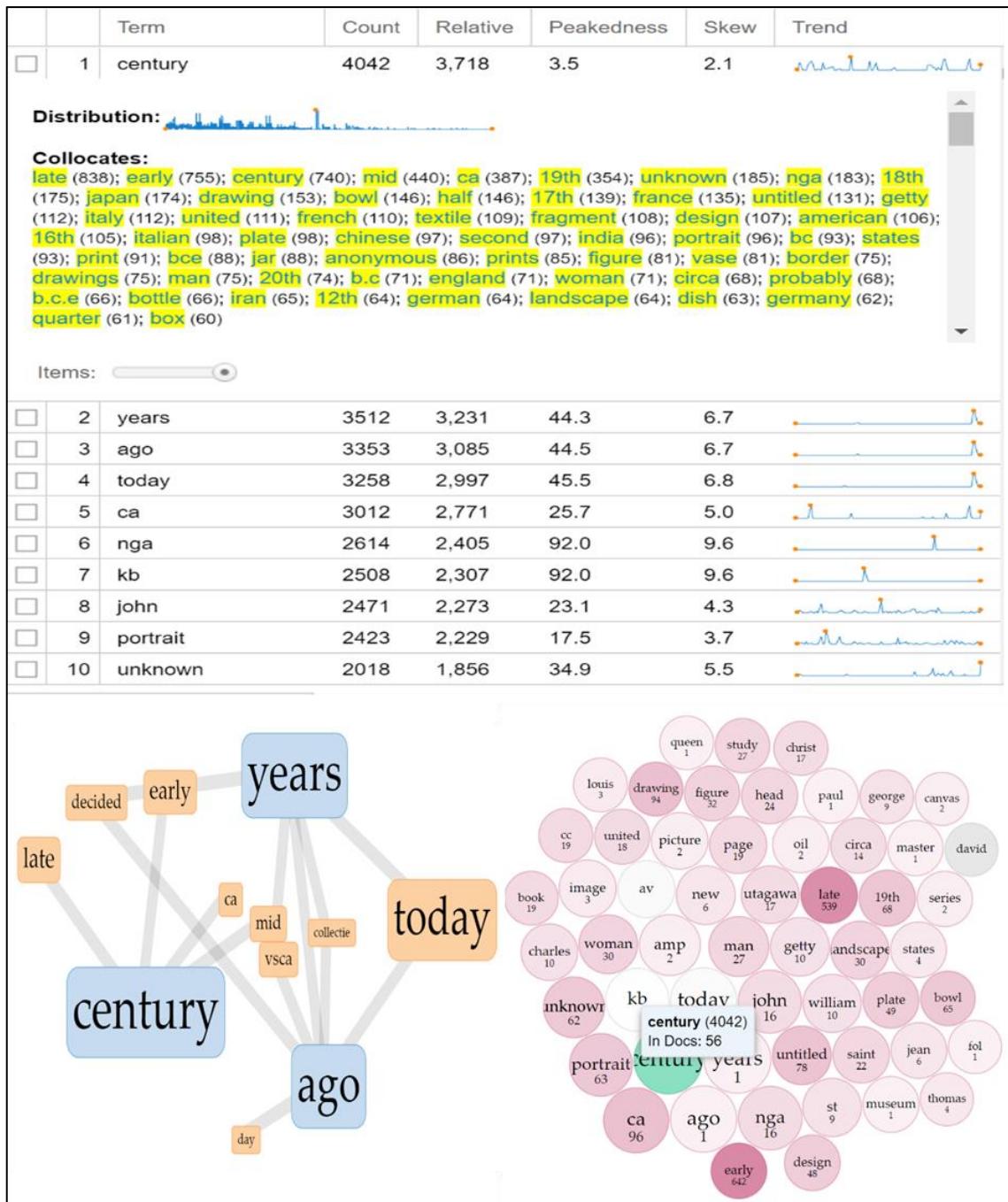


Figure 10: Corpus linguistics analysis for all individual bot accounts. Table showing top ten terms in order of relative frequency. Measures of dispersion across the corpus are shown for each term. The keyword 'century' is highlighted and its collocates displayed.

Name	Followers	Active?	Retweeted per month	Likes per month
r_HistoryMemes	43420	y	37107	365923
wayback_exe	15541	y	381	1333
SovietArtBot	12466	n	785	3889
Geocitiesgifbot	10512	y	2528	10116
MuseumBot	8326	n	322	354
UnwobblingPivot	7579	y	976	3408
MoMARobot	4494	y	93	380
Oldschoolflyers	4059	y	309	752
NYPLEmoji	3704	y	14	55
Openaccessart	3192	n	148	406

Table 3: Table showing Twitter engagement metrics for top ten individual bot accounts with the most followers.

The top performing bots are those dedicated to very specific, more recent aspects of history, or particular social media communities based around history, for instance r_HistoryMemes which automatically retweets history memes from a popular Reddit community. Wayback_exe and geocitiesgifbot make use of the Internet Archive as a source from which to share examples of old websites and GeoCities pages, celebrating the very recent history of the World Wide Web. At the other end of the timeline, UnwobblingPivot tweets out sections from the works of Confucius. Art and GLAMs feature too, with SovietArtBot sharing artworks from WikiArt and MoMARobot disseminating items from the Museum of Modern Art's open access collections. NYPLEmoji, which shares items from the NYPL corresponding to an emoji tweeted by a user, is an example of a bot created in response to the New York Public Library's call to remix its collection; despite its relatively strong following, monthly impressions are low. The

presence of these eclectic bots, often not using GLAMs as content sources, shows an audience with broad interests and not one solely focussed on art or GLAM collections.

Assessing the aims of developers of the individual bots is difficult, but it is clear that even for those that share GLAM content this is in an unofficial capacity and most developers do not seem to maintain their bots or create them for a specific project and then forget about them. The only bot developed 'officially' was @MechCuratorBot, made by British library staff as an experimental project, but this is no longer maintained and is currently inactive. 61/139 (44%) individual bot accounts are currently inactive, i.e. they have not tweeted for at least two months before data collection occurred, and the average active lifespan of an individual bot is 4 years and 8 months with the average time since last tweet for inactive bots being 2 years and 6 months. This shows the temporary nature of these individual bots, a phenomenon only exacerbated by ever-changing platform automation policies, bans and purges. @DeathMedieval is a prime example of an extremely popular bot that has somehow fallen foul of the rules and is currently suspended, although it has inspired two other generative bots that have quite possibly been trained on its data that is no longer available.

Some developers have created a few bots, for example one developer manages 15 of the bots within this dataset, whilst another manages 28 bots as part of a project called 'OpenArchief' (Open Archive) based in the Netherlands. Yet, all of these individual bots are true loners, forming no connections to other similar bot accounts through retweets, follows or hashtags. None of the bots retweet each other, not even those set up as a group by a single developer (e.g. OpenArchief bots), and only 41/139 (29%) of the bots use hashtags at all! This seems like a missed opportunity to increase engagement and help to create a wider community around the bots, which is exactly what the Off The Easel bot community achieves.

3.3 The ‘Off The Easel’ Twitter Bot Community and the Sharing of GLAM Collections

The Off The Easel Twitter bots (n=455) form a coordinated community of accounts designed to share Open Access art and cultural heritage artefacts from Open Access online museum collections. There are two types of account within the network: artist bot accounts (81% of bots), which share artworks for a given artist, primarily from WikiArt (a collaborative, community-based platform of Open Access artworks dubbed the Wikipedia of art); and GLAM bot accounts (19% of bots), which share artworks and artefacts from GLAM collections.

Originally, Off The Easel began with the aim of sharing art, but as it has developed the GLAM accounts have been added, so their role within the network is slightly different as will be discussed. A few accounts have recently been added that highlight contemporary art from different countries, but these have been excluded from analysis as the focus here is on art that is also part of the cultural heritage sector – think of these as bot accounts for dead artists, but not dead art.

The artist accounts within Off The Easel primarily represent European artists, with American and Russian artists making up the bulk of the rest (see Figure 11). This breakdown follows the familiar Eurocentric narrative of art history and its most revered figures, such as DaVinci, Van Gogh, Monet and Picasso, whilst also reflecting the collections of the Western GLAMs where their artworks are most commonly found. However, Russian artists gain a boost thanks to @andreitr’s personal desire to highlight Russian and Soviet art, showing that human curatorial choice still plays a major part in shaping a largely automated network.

Similarly to those present in the individual Twitter bots dataset, the GLAM accounts within Off The Easel are predominantly Western institutions, particularly North American (see Figure 12). Again, old colonial collecting habits are plain to see with all non-Western content being shared via accounts for Western GLAM institutions. The majority of GLAM bot accounts share content representing the cultural heritage of numerous regions, the ‘Worldwide’ category in Figure 13, with European content not far behind. As Figure 15

shows, most of this cultural heritage content is in the form of artworks rather than artefacts, which is not surprising given Off The Easel's focus on art and artists. The time periods covered for content shared by the GLAM bot accounts also follow historical definitions rather than art periods. As for the individual Twitter bots, the majority of content shared from the GLAM bots cuts across multiple periods or is modern (i.e. post-18th Century). It should be noted that content in the modern category was not just modern art, but a range of objects including, for example, bots sharing collections of photographs from various institutions. Some of these photos might be viewed more as art, such as portrait or landscape images, but others are records of events and better viewed as historical sources than artworks; this is even more apparent for the medieval category where illuminated manuscript pages from Books of Hours make up significant proportions of the content shared (although this comes with its own confusions, see below).

The sources of content for all the Off The Easel bots are primarily Open Access GLAM collections. Although the main direct source for artist bots' content is WikiArt, many of the images collated there by its community originate from Open Access GLAM collections. The same images may therefore appear in artist bot accounts and GLAM ones, with respective accounts often retweeting each other. The human community behind WikiArt and created by those following the Off The Easel bots is therefore central in determining the content that can be shared by these bots, along with the GLAMs that make their collections publicly available. For an artist bot to be created there needs to be at least 100 images available for use, leading to calls from @andreitr for the community to help source more images in order to create a bot, or sometimes have bots tweet out requests to enlarge the scale of the collections that they share²⁹. For some artist accounts that have little more than 100 images to share, content feeds can soon become repetitive, and so these calls engage their followers to helping to improve their own art browsing experience.

²⁹ Artemisia Gentileschi, '@andreitr Please Help Me to Make This Account Better - Add More Art to Gentileschi's Wikiart Profile [Https://T.Co/HK1eMdi0k7](https://T.Co/HK1eMdi0k7)', @gentileschi_art, 2019
<https://twitter.com/gentileschi_art/status/1184320197806542848> [accessed 3 March 2021].

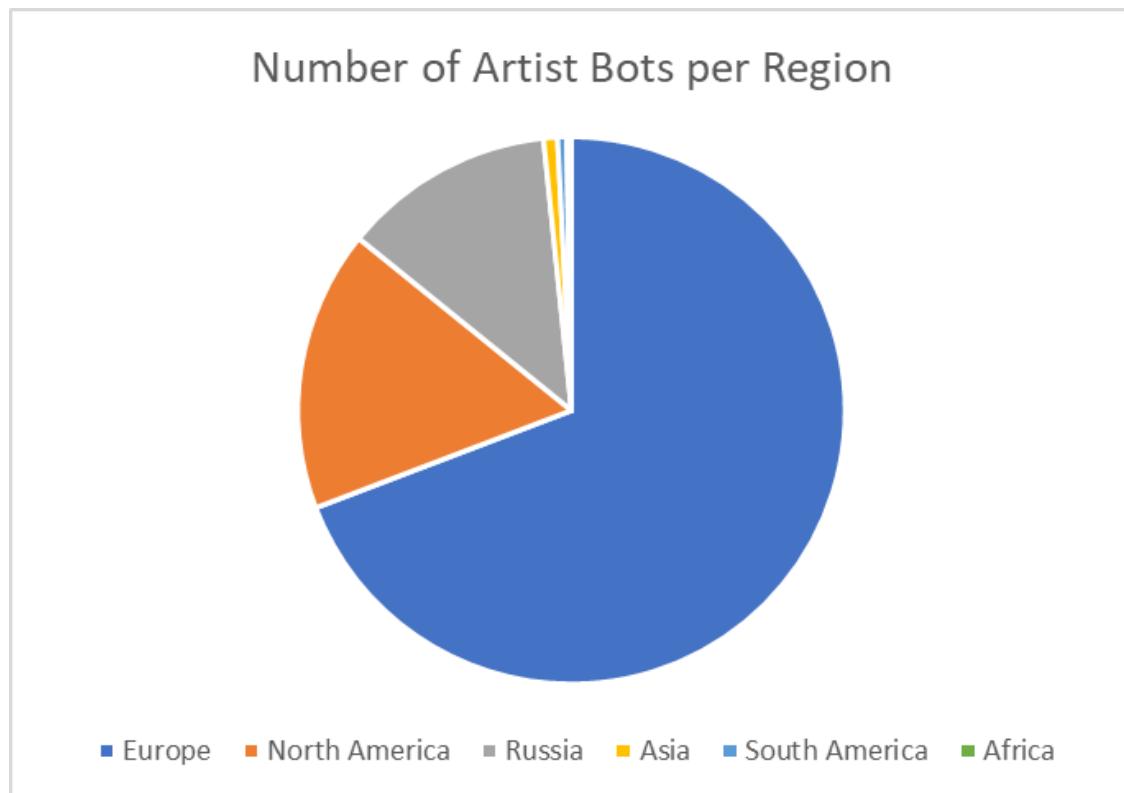


Figure 11: Pie chart showing regional spread of artist Twitter bot accounts in the 'Off The Easel' network.

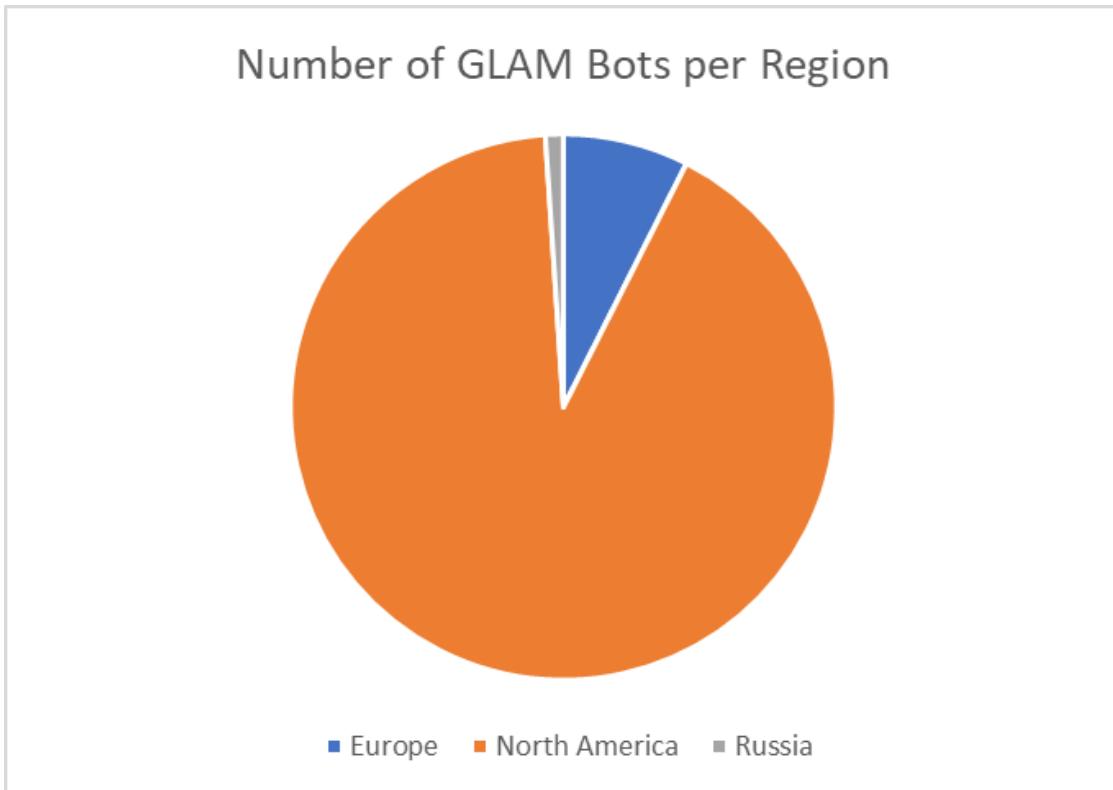


Figure 12: Pie chart showing regional spread of GLAM Twitter bot accounts in the 'Off The Easel' network.

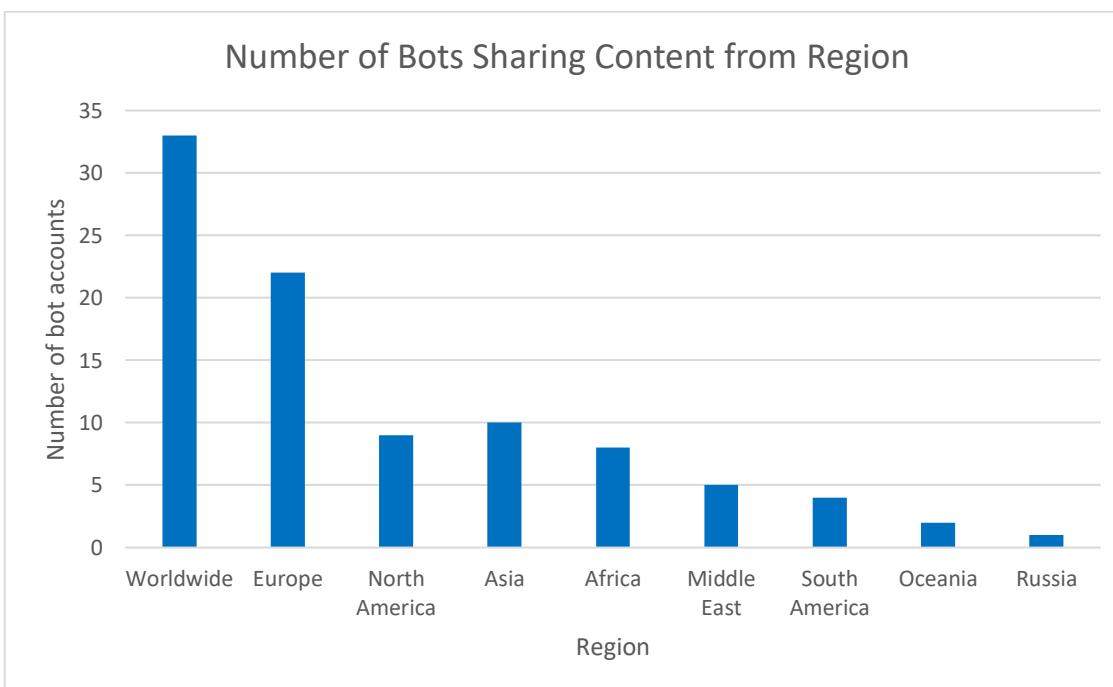


Figure 13: Bar chart showing proportion of content representing different regions for GLAM bots in the 'Off The Easel' network; worldwide represents bots sharing content covering multiple regions.

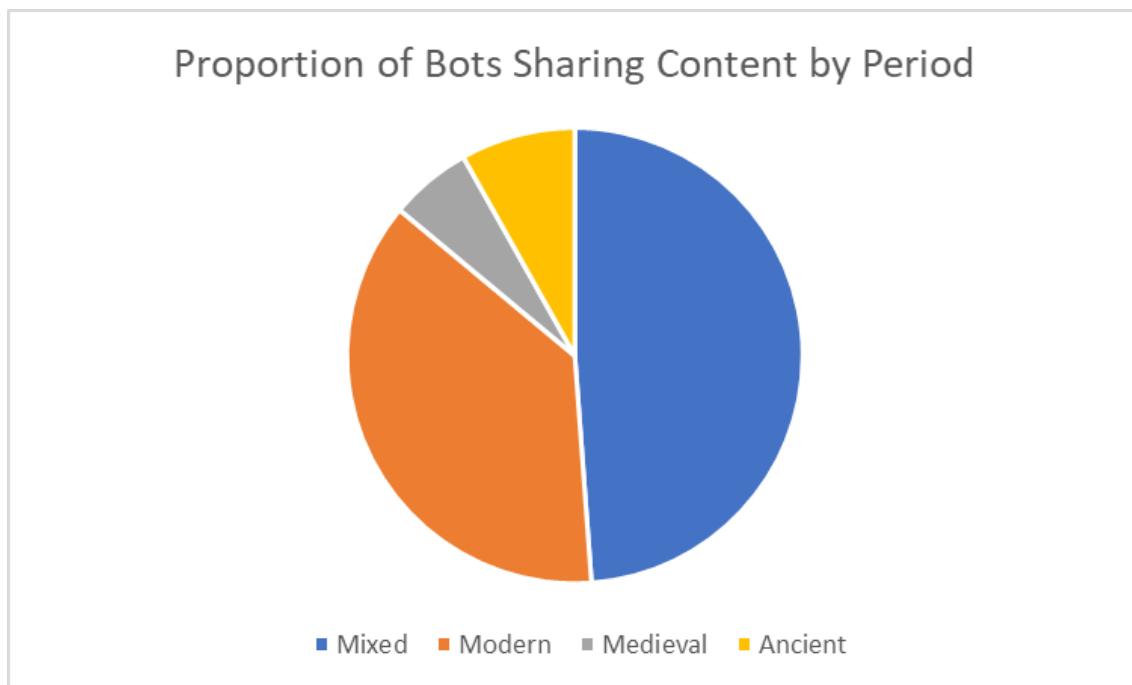


Figure 14: Pie chart showing proportion of content representing different time periods for GLAM bots in the 'Off The Easel' network; mixed represents bots sharing content covering multiple periods.

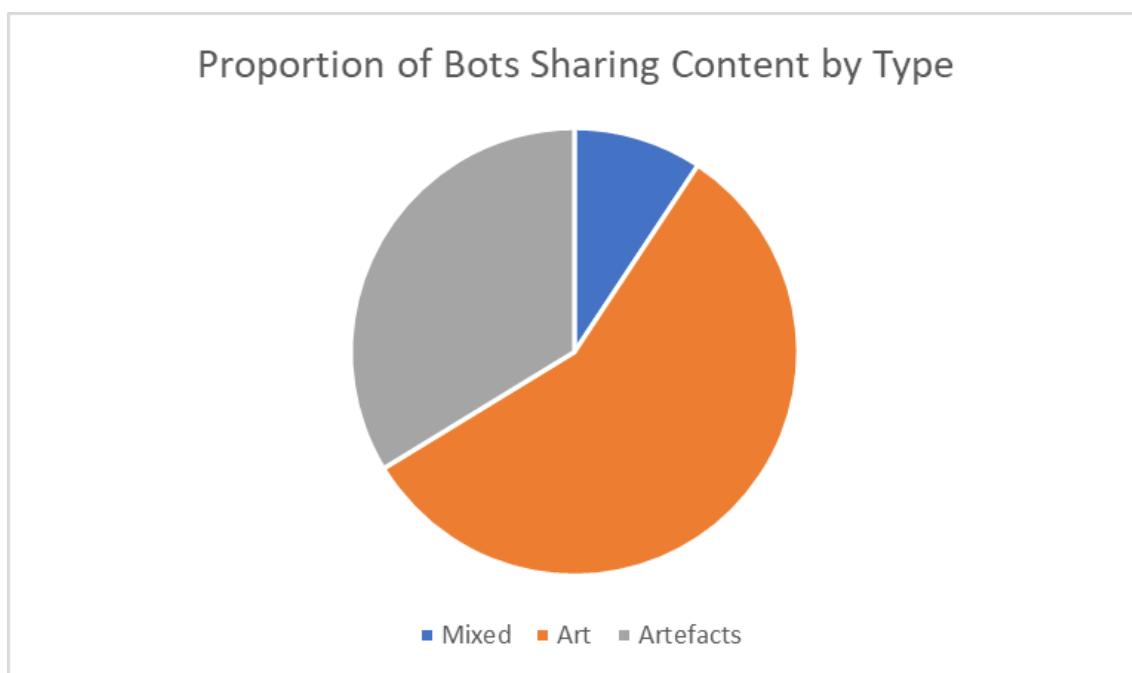


Figure 15: Pie chart showing proportion of content representing different categories for GLAM bots in the 'Off The Easel' network; mixed represents bots sharing both art and artefacts in roughly equal measure.

To examine the content of the tweets themselves more deeply, corpus analysis and close reading were performed for all the tweets posted by artist account bots and GLAM account bots as distinct corpora. Retweets, mentions and hashtags were removed from the GLAM bots corpus, but hashtags were left in for the artist bots corpus as these are the main indicators of the artistic style being shared. Most of the remaining tweet texts read much like a list of artwork or object collection titles due to the nature of content tweeted by automated accounts, particularly textual information accompanying images or other media, that is often repetitive and of a fixed style and format. Calculating the relative frequencies, dispersions and collocations (associations) between words in the tweets gives a general idea of content topic and style across the corpus.

The corpus of textual content tweeted by bots in the art bots corpus, nearly all representing artists via Off The Easel, the majority sourced from WikiArt, is dominated by the word 'portrait', the relative frequency of which is more than twice that of the next most popular term 'woman'; 'saint' (or 'st') and 'landscape' follow. This is what one might expect from a corpus of artwork titles from predominantly European painters, and naturally terms like 'portrait' occur far more often for certain artists than others, although it also has the highest dispersion measure (occurring across different accounts) of any term, with 'woman' a close second (the two are often associated, see Figure 16). Given that the overall aim of 'Off The Easel' is to promote all fine art, and there are plenty of accounts for modernists and 20th Century artists, these results still reflect the reality that much of the digitised fine art held in openly accessible online museum collections is European paintings of people, particularly women.

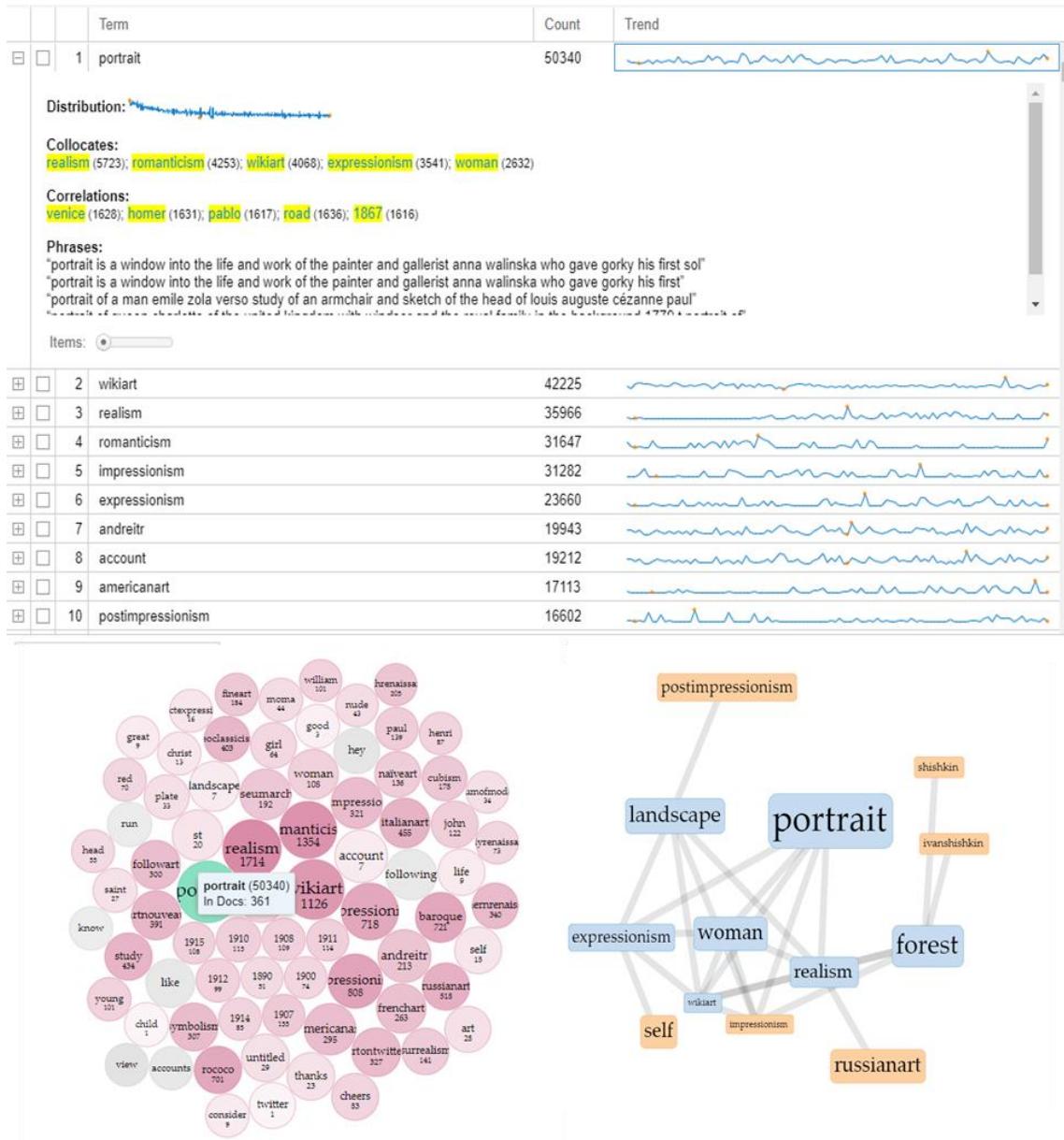


Figure 16: Corpus linguistics analysis for the ‘Off The Easel’ art bot accounts. Table showing top ten terms in order of relative frequency. Measures of dispersion across the corpus are shown for each term. ‘Woman’ is highlighted and its collocates displayed.

The GLAM bots corpus (see Figure 17) is a mixture of the themes present in the art and history corpora. The terms ‘man’, ‘woman’, ‘portrait’ and ‘landscape’ having high relative frequencies, echoing both the historicity and importance of place in the content disseminated and the characteristically concise style of curatorial titling and description. This is evident across all corpora, and is perfectly suited to Twitter, but works best if a link is included in the Tweet to a source where the user can find more information. The Off The Easel bots

covering various departments of the Cleveland Museum of Art and the Met Museum provide such links, taking the user to a full description from the original, authoritative source, but for many of the artworks tweeted a title is all that exists on WikiArt; instead art bot account profiles contain links to Wikipedia pages about the artist. This lack of deeper context may be frustrating for some users, but for the vast majority the image is what matters along with the opportunity to see art and cultural heritage as they scroll through a social media feed.

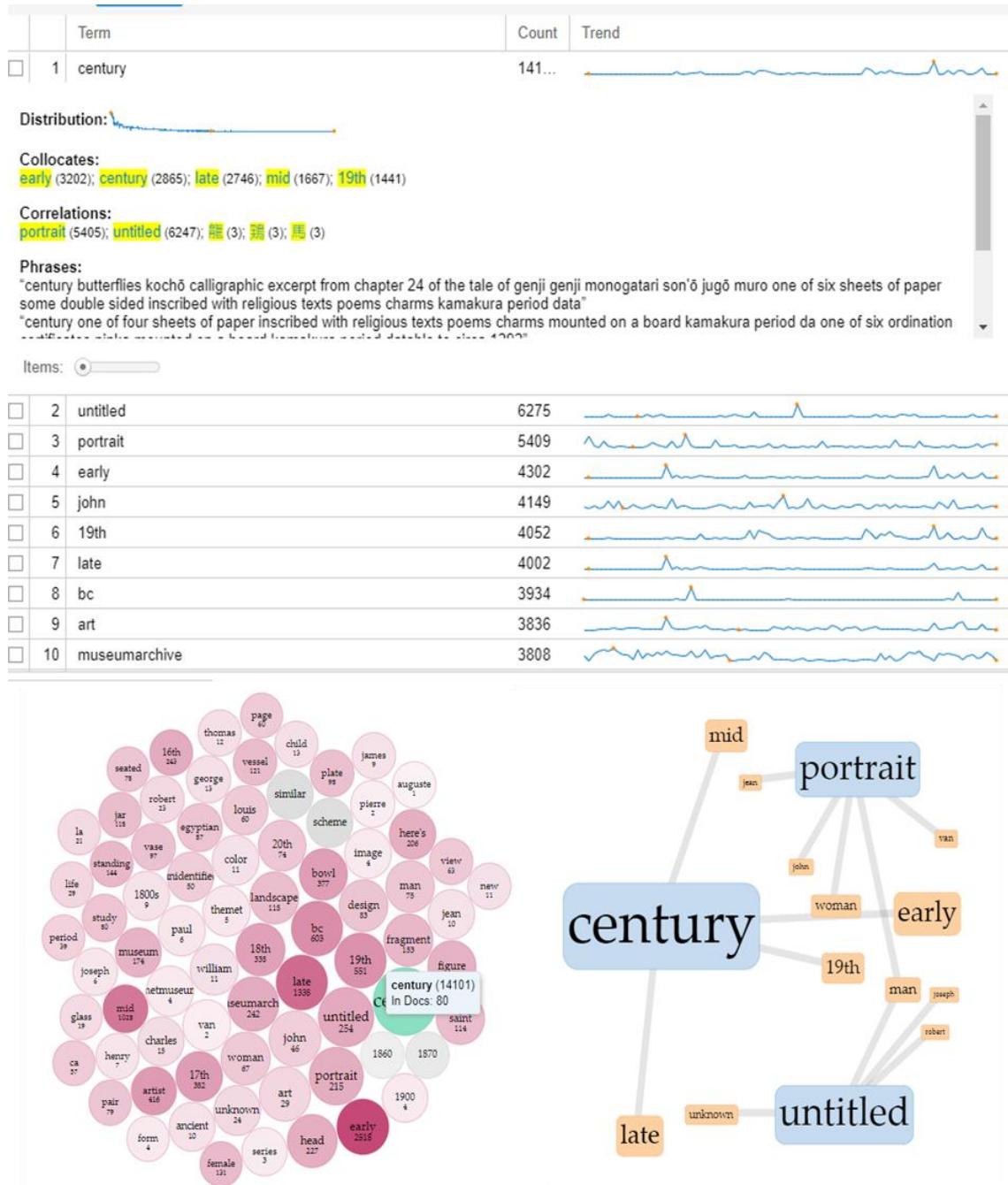


Figure 17: Corpus linguistics analysis for Off The Easel GLAM bot accounts. Table showing top ten terms in order of relative frequency. Measures of dispersion across the corpus are shown for each term. ‘century’ is highlighted and its collocates displayed.

Close reading of the history and heritage content being disseminated also reveals the nuances involved in automatically sharing different types of historical object. This becomes clearer when comparing the Off The Easel bots for the Cleveland Museum of Art and the Met Museum. For instance, @cma medieval tweets out images of objects from the Cleveland Museum of

Art's (CMA) Medieval Department, of which there are more than 4,000. However, contained within these are some objects like the book of 'Hours of Isabella the Catholic, Queen of Spain', which contains more than 275 illuminated manuscript pages. The choice of classification here is also crucial; 4,000 physical objects actually resembles tens of thousands of digital images, the automated dissemination of which will not necessarily happen together or in order.

Confusions over object context, a desire for more information and attempts to recombine objects with sets of related images all appear in users' replies to tweets posted from the GLAM bots. These bot tweets all contain links to the original listing on the GLAM's website, so sometimes these comments are directed at the bot (or indirectly at @andreitr) and other times indirectly at the GLAM itself. For example, one user (seemingly an academic working for a different GLAM) replied to a tweet from met_medievalart bot wanting the museum's collection entry for some glass fragments to better engage with its potential audience:

"So, what's going on with your fragments? What would you like your audiences to take home about them (or, what are they looking for)?"³⁰

Replies to other tweets also wanted more information from the GLAM collection entry about the object being shared, including a response by a user to an image of a textile fragment:

"Is that a face on the torso? I really wish the linked page could have given a bit of context and background."³¹

In an effort to regain some of the context lost by automated tweeting of individual images, some users replied to single images of objects actually present in sets in the original GLAM collection, for example an image of a pottery bowl from one angle received a reply showing the other two images of

³⁰ The Met: Medieval Art, 'Glass Fragment [Https://T.Co/Owx45OJ0Qg](https://t.co/Owx45OJ0Qg) #MedievalArt #metmuseum [Https://T.Co/SHTuXAK1Ng](https://t.co/SHTuXAK1Ng)', @met_medievalart, 2021
<https://twitter.com/met_medievalart/status/1364062121181540352> [accessed 3 March 2021].

³¹ The Met: Islamic Art, 'Fragment [Https://t.co/eBsTx2aCdr](https://t.co/eBsTx2aCdr) #metmuseum #themet [Https://t.co/g5Eud5tJ6H](https://t.co/g5Eud5tJ6H)', @met_islamicart, 2021
<https://twitter.com/met_islamicart/status/1364204331843649539> [accessed 3 March 2021].

the bowl from the linked GLAM set³². However, the case of CMA_medieval bot and its numerous tweets of manuscripts of Books of Hours is not amenable to this approach and its original organisation by page causes confusion for users, for example in this instance where a blank page in a book has been scanned and the trace of an image on the next page shows through³³. This makes sense when viewing the manuscripts in the context of their set on the GLAM collection website, but individually it is confusing (although that tweet still received four likes). In response to such occurrences, @andreitr manually trawls through some GLAM image datasets to remove blanks, faint or poorly-digitised images from collections, leading to debates with the community about what should or should not be included³⁴.

Audience engagement occurs on various levels, from people following a couple of bots to highly engaged users who follow more than 10% of all Off The Easel bots and users who are involved members of the community, helping to shape and discuss the content shared as much as consuming it. Assessing the contribution of the latter is challenging, although an idea of this can be gained from close reading of tweet threads and replies, but the wider metrics of engagement for Off The Easel bots give a general picture of content reach and popularity.

Table 4 shows the engagement metrics for the Off The Easel bots with the most followers. Unsurprisingly, these are accounts for world-famous, extremely popular artists including Monet, Van Gogh and Picasso, with many other artist bots appearing next in the list showing an audience focus on artist bots over GLAM bots. The numbers of likes a bot receives and the number of times it is retweeted are key markers of engagement and it is interesting to note that some bots, such as those for Dali and Magritte, have fewer followers but higher levels

³² CMA: Chinese Art, 'Tripod with Cover, Late 1700s [Https://T.Co/6mg3pX3R43](https://T.Co/6mg3pX3R43) #museumarchive #cmaopenaccess [Https://T.Co/1DfI0I7P1A](https://T.Co/1DfI0I7P1A)', @cma_chinese, 2021 <https://twitter.com/cma_chinese/status/1363670251595042818> [accessed 3 March 2021].

³³ CMA: Medieval Art, 'Hours of Queen Isabella the Catholic, Queen of Spain: Fol. 24r, c. 1500 [Https://T.Co/lIz1FTIUi1](https://T.Co/lIz1FTIUi1) #museumarchive #clevelandartmuseum [Https://T.Co/ev1BrwYeau](https://T.Co/ev1BrwYeau)', @cma_medieval, 2021 <https://twitter.com/cma_medieval/status/1362514454869663758> [accessed 3 March 2021].

³⁴ Andrei Taraschuk, 'Tate's Collection Includes Hundreds of Sketchbook Pages like This One. I Am Not Sure Whether It Makes Sense to Include Those into the #artbot. What Do You Think?', @andreitr, 2020 <<https://twitter.com/andreitr/status/1266030314108801025>> [accessed 3 March 2021].

of engagement, showing a smaller but more involved audience. Monet and Van Gogh top both lists, showing their overall popularity, with Van Gogh topping the list for the number of unique hashtags across the sample of tweets, a rough measure of content diversity and wider engagement, such as retweeting official GLAM accounts that are exhibiting his work. This is one way in which the whole bot community attracts users to view and share content.

Name	Followers	Retweeted per month	Likes per month	Unique Hashtags
artistmonet	55599	23716	88008	86
vangoghartist	48889	16315	71459	103
cezanneart	30159	6402	21013	82
artisthopper	25726	6120	26493	76
artistgklimt	25045	3799	16378	76
artistgauguin	23565	4974	15757	92
artist_dali	22970	10670	23551	79
artistmagritte	22421	11491	27299	73
artpissarro	22363	10225	20990	91
pablocubist	22236	7102	13343	73

Table 4: Table showing Twitter engagement metrics for top ten Off The Easel bot accounts with the most followers.

In order to try and find out more about the highly engaged followers of the bots, seemingly with an interest more in art than cultural heritage, user account descriptions were manually read and classified into professional groups: [1] historian or archaeologist, [2] GLAM professional, [3] academic in another field (including students), [4] artist or art market professional, and [5] other. 'Other' included those who did not fit into any of these categories or who did not

provide enough information for this to be clarified and was further subdivided into users from other professions, those stating an interest in art or history and everyone else. Reading the location descriptions of followers (where possible) also showed that the vast majority were from Europe or North America, mirroring the regional spread seen for the content shared by the bots.

The results (see Figure 18) show that the 'other' category dominates, suggesting that a general audience is the main consumer of content shared by the social bots. Within this group, a far higher proportion were interested in art than history, which reflects the prevalence of the Off The Easel botnet. Unsurprisingly, 'artist / art market professional' was the most common of the four professional categories, but perhaps more interesting is the relative lack of GLAM professionals following the bots, especially given the fact that the bots are sharing the collections of many GLAM institutions. This could be because this sample covers only highly engaged followers of the bots and perhaps GLAM professionals may be more likely to follow only a few bots relevant to their institution or area, but it also seems plausible that they would prioritise following official, human-run GLAM accounts in favour of unofficial, automated ones.

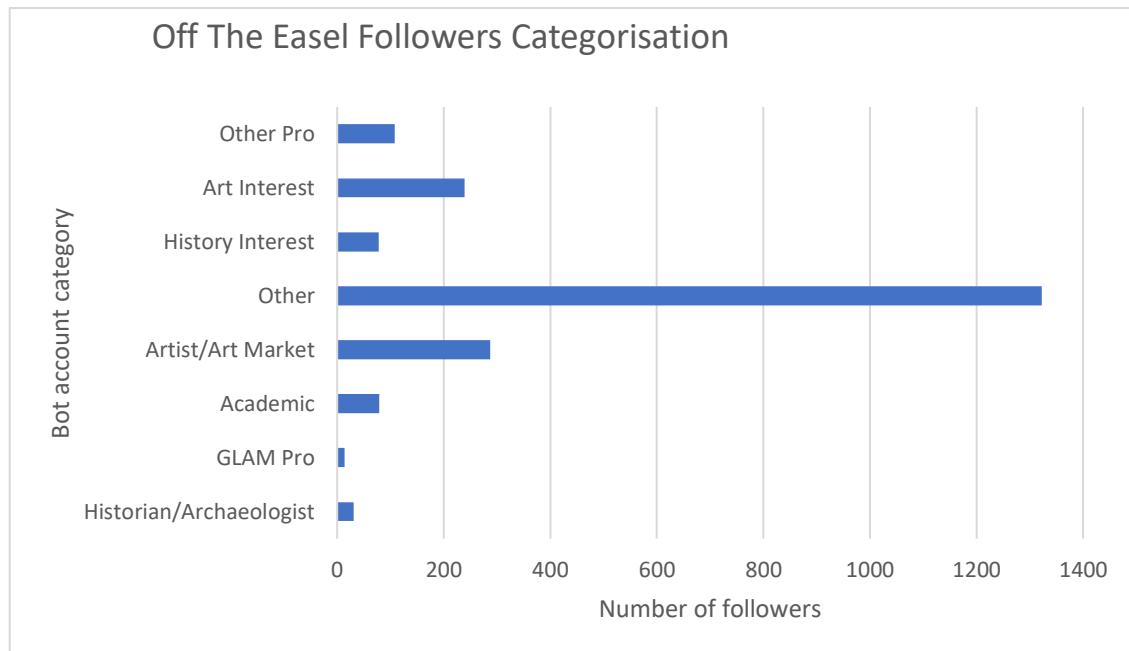


Figure 18: Bar chart showing the distribution of highly engaged followers by profession or area of interest.

The perceived authority of official accounts, and conversely the suspicions around automated accounts, may also play a part; when designing the Off The Easel botnet, the developer was keen for his bots to follow authoritative human accounts to lend a sense of authenticity and trustworthiness to the bots themselves. When created, each bot selects a few human accounts to follow from a whitelist of accounts chosen by the developer that represent “authoritative sources”³⁵ of information for art and GLAMs, for instance official museum accounts like @metmuseum and respected resources and publications such as @artnet and @nytimesarts. Given the current reputation of social bots on Twitter and the public perceptions around fake news, the developer thought it important that the art bots follow (and interact with) such authoritative, trustworthy human-run accounts.

The pattern in following across all the different groups was very similar and copied the overall trend with popular art bots like @vangoghartist and @artistmonet being the most followed. The fact that this general trend did not change across each demographic suggests that entertainment, relaxation and an enjoyment of art is the main motivation for following the art bots rather than education or research. The sheer number of art bots present also makes this far more likely overall, but many within the professional categories exclusively followed art bots, probably because the design and extent of the botnet made them much easier to find.

The nuances of content and context discussed in the previous section highlight the fact that audience engagement is about more than just metrics and demographics. Replies to tweets are rare compared to the enormous amount of content that is tweeted, but they offer valuable insight into the ways in which more involved users interact with the bots and the wider community. For some artist bots, a few users will comment frequently saying how much they like a shared artwork, whilst some comment far less frequently but want to know more

³⁵ Andrei Taraschuk, ‘Bots Going Artsy – Strategies for Sharing Large Art Collections with Social Media Bots and AI’ (presented at the SharingisCaring X Stockholm, 2019). ‘Bots Going Artsy – Strategies for Sharing Large Art Collections with Social Media Bots and AI’.

about certain artworks, often unusual ones³⁶. Other users will have conversations about shared content³⁷, although these are rarer. The overall effect of this more in-depth engagement is to create a sense of human community around the bot community with which followers can participate, even if only briefly or infrequently, allowing the bots to reach and maintain a wide audience. This is central to Off The Easel's approach and is enabled by the coordinated suggestion and sharing of content from accounts across the network.

The aim of @andreitr in setting up Off The Easel is to share as much art content as possible across social media, democratising access to art and making it easily accessible³⁸. A feature of this goal was not only to share the work of already popular artists, and further amplify the content bubbles that can come to define social media accounts, but to harness their popularity to highlight the work of other, lesser-known artists. Although not an artist or GLAM professional, @andreitr has a passion for art and a background in the tech sector. In a general sense, the overall community of Off The Easel bots functions as both an art content sharing platform and a basic recommendation network³⁹. This connectedness is what helps to drive engagement and makes Off The Easel more effective at sharing a wide range of content than individual Twitter bots. Connections between bots (and groups of bots) are achieved and maintained through three main routes: retweets, recommendation mentions and hashtags. These mechanisms and their effects are examined along with the manual time and effort required by @andreitr to maintain the overall function of the bot community and keep expanding its reach.

Network connections via retweets allow bots to highlight content shared by others, often those of a similar style. For artist bots sourcing content from

³⁶ Marc Chagall, 'Lid: "Meeting of Isaac and Rebecca" #naïveart #chagall [Https://T.Co/HttCYTP2hh](https://T.Co/HttCYTP2hh)', @artistchagall, 2021 <<https://twitter.com/artistchagall/status/1361849925961711618>> [accessed 3 March 2021].

³⁷ CMA: Greek and Roman Art, 'Torso of a Youth, 400-375 BC [Https://T.Co/IY2Elg4NF2](https://T.Co/IY2Elg4NF2) #museumarchive #clevelandartmuseum [Https://T.Co/cPcPEy3Tng](https://T.Co/cPcPEy3Tng)', @cma_greekroman, 2021 <https://twitter.com/cma_greekroman/status/1363598191376162819> [accessed 3 March 2021].

³⁸ Taraschuk, 'Bots Going Artsy – Strategies for Sharing Large Art Collections with Social Media Bots and AI'.

³⁹ ibid.

WikiArt, sets of tags for artistic style manually assigned on WikiArt are used as one of the determining factors when deciding which bots retweet each other, meaning that bot accounts for stylistically similar artists will create small sub-networks through retweeted content to engage followers with related bot accounts. Artist bots also retweet human-run accounts, such as popular auction houses and contemporary art magazines, if the work they share is featured on these platforms⁴⁰, for instance if a particular artist's work has just been sold. These create flashpoints of engagement for existing followers but are designed more to add a human authenticity and authority to the network as discussed above.

As Table 5 shows, retweeting is frequently used to highlight the content of GLAM bots across the bot community. As these accounts usually share artworks from a range of artists who have bot accounts within the network, they are frequently retweeted by artist accounts highlighting the presence of their content in the relevant GLAM collections. The network centrality statistics show that this approach makes the GLAM bots very influential in terms of how visible content becomes across the whole network, and so increasing the chances of users finding it from very different artists accounts, but this activity is passive and all of the GLAM bots use many more of the tweets sharing content than they do directly suggesting other accounts to users via mentions.

This strategy is favoured by some artist bots, with the most prolific recommenders listed in Table 6, that are far less engaged in terms of retweets and being retweeted but use as many as half of their tweets to directly suggest other Off The Easel accounts, both artist and GLAM bots (for example a tweet from artfridakahlo bot which highlights the bot account for artist Juan Gris: "Juan Gris @artist_gris #followart #juangris")⁴¹. This recommendation tweet received 28 likes and was retweeted 5 times, showing a reasonable level of follower engagement with this approach.

⁴⁰ My Modern Met, 'Frida Kahlo Is an Iconic Artist Whose Legacy Lives on. Learn More about This Legend. [Https://T.Co/f2vR9OZiHk](https://T.Co/f2vR9OZiHk)', @mymodernmet, 2021 <<https://twitter.com/mymodernmet/status/1364805996196069376>> [accessed 3 March 2021].

⁴¹ Frida Kahlo, 'Juan Gris @artist_gris #followart #juangris <https://t.co/3Ik37J3C4y>', @artfridakahlo, 2021 <<https://twitter.com/artfridakahlo/status/1361009060427018249>> [accessed 3 March 2021].

Name	Times Retweeted	Retweets	Betweenness Centrality	Stress Centrality	Suggestions Percentage	Unique Hash tags
the_barnes_bot	1289	262	0.11	777063	1%	119
cma_eupaintings	1330	231	0.03	482380	2%	28
guggenheimbot	2978	300	0.06	452274	2%	308
met_eupaintings	1084	1052	0.04	330793	1%	453
met_lehman	1018	1030	0.04	299651	1%	214
TateArtBot	1076	929	0.03	188606	1%	490
met_ampainting	1282	1012	0.02	186171	1%	322
126garvard_artbot	889	968	0.02	177411	1%	460
vangoghartist	974	271	0.02	163176	1%	103
slam_americana	816	495	0.02	151011	6%	44

Table 5: Table showing network statistics for top ten most influential Off The Easel Twitter bot accounts in terms of highlighting content across the botnet. Note these bots focus on retweets for influence, shown by their low percentages of suggestion tweets.

Name	Times Retweeted	Retweets	Betweenness Centrality	Stress Centrality	Suggestion Percentage	Unique hashtags
artfridaka_hlo	709	284	0.00373	13097	51%	77
artcarava_ggio	684	255	0.00006	2462	50%	46
artistvaro	410	322	0.00013	1317	49%	61
lauraknighrtart	531	147	0.0022	17114	47%	72
artistkollwitz	342	231	0.0001	1489	44%	90
waterhouse_art	539	402	0.0006	5184	42%	65
artistpollo_ck	359	272	0.00264	13701	41%	70
gentilesc_hi_art	310	157	0.00015	226	41%	26
artistfriedrich	563	409	0.0015	12071	41%	80
artistbasquiat	208	320	0.0014	2201	41%	37

Table 6: Table showing network statistics for top ten recommenders within the Off The Easel Twitter bots network in terms of pointing followers towards content shared by other bot accounts. Note these bots focus on tweets suggesting other art bot content.

Hashtags are a staple of Twitter communication, allowing tweets to be tagged as part of a theme or topic that can be broadly searched across the platform. Within the Off The Easel botnet hashtags play an important role in linking together artist bots of the same style (as tagged on WikiArt) and help to bring the art bot to the attention of Twitter users searching for information about an artist or art style more broadly. GLAM accounts obviously have a far more diverse range of content than artist accounts which is reflected in the higher number of unique hashtags they tweet out, adding to connections across the network and bringing their content to the attention of a broad audience of users both in and outside the network. For GLAMs that have multiple collections represented by bot accounts, for example the eighteen Metropolitan Museum of Art bots that share content from different collections (or similar collections from different GLAMs), hashtags can form common connections between the different collections with very different content (see Figure 19). GLAM bots sharing art content can thus act as a bridge allowing GLAM bots primarily sharing artefacts to indirectly connect with followers of artist accounts elsewhere in the network.

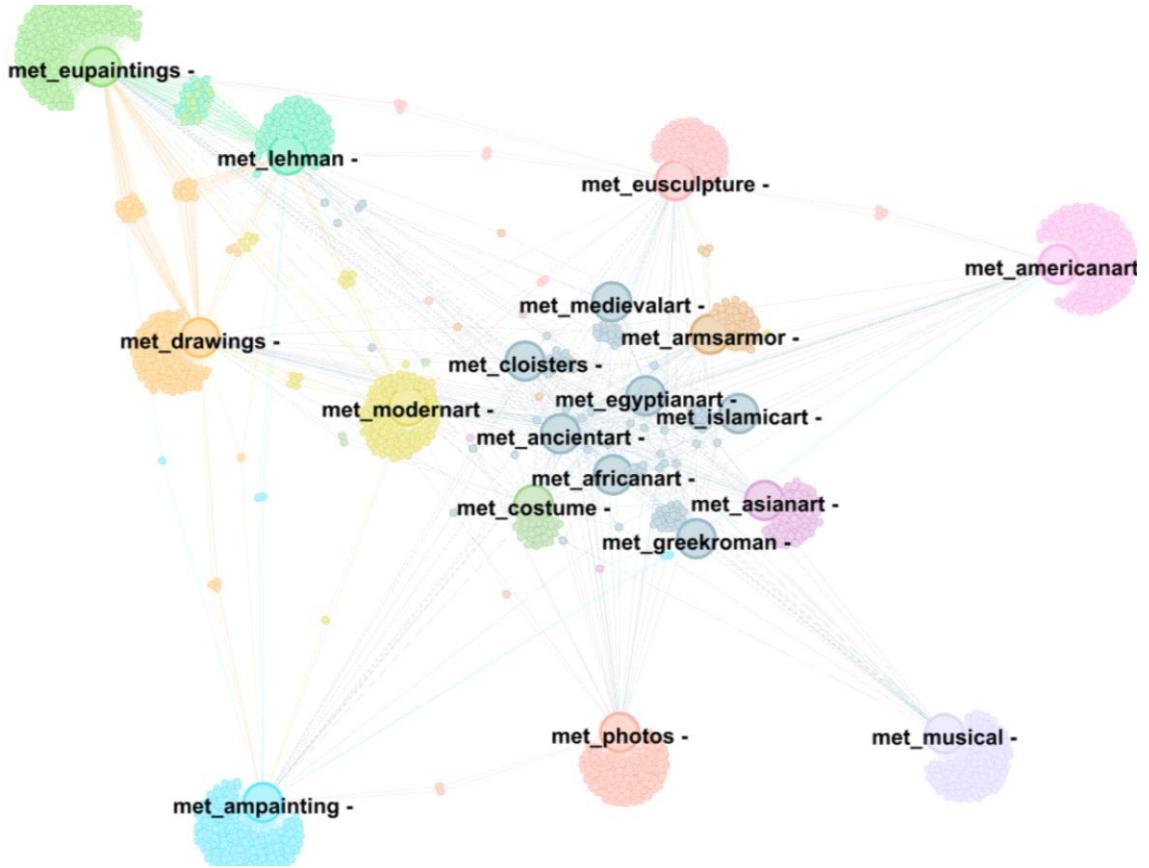


Figure 19: Visualisation showing how the eighteen Off The Easel Metropolitan Museum of Art bot accounts form a connected and coordinated network via the hashtags that are included in some of their tweets.

The network visualisation in Figure 19 shows the eighteen Off The Easel Metropolitan Museum of Art bot accounts, the large circles, and the hashtags they tweet, the clusters of tiny circles. These clusters contain hashtags that form a link to relevant artist accounts in the rest of the network and tags more widely in use, such as '#metmuseum'. The lines link a bot to its hashtags, with common hashtags showing similarities in content shared between bots.

This visualisation demonstrates that the combination of retweets and hashtags to highlight shared content and diversify followers' browsing habits is what forms the core of a semi-automated, semi-algorithmic coordination of content sharing that engages millions of users. However, these approaches are reliant upon a collaborative digital infrastructure that is largely based on human effort, primarily WikiArt and open access GLAM collections, along with the large amounts of time, effort and money expended by @andreitr to manage it all.

Whilst this existing infrastructure enables such a bot community to exist and thrive, the coordinated approach can begin to suffer when content sources diversify away from it.

For example, there are some newer artist accounts like [artdellabella](#)⁴² that do not source their content from WikiArt, but instead have had their collection brought together from different GLAM account collections by @andreitr. These accounts do not have the tags from WikiArt and so tweet hashtags that link to GLAM bots and institutions rather than related artists and they cannot make use of the semi-algorithmic retweeting patterns reliant on WikiArt tags, so do not retweet other accounts. This is something @andreitr may develop over time, and indeed Off The Easel is already experimenting with an altered approach for semi-algorithmic content coordination not reliant on existing tags but based on visual similarity. What is clear is that automated, algorithmic aspects of Off The Easel enable its coordinated approach and help achieve its high levels of engagement, but central to all of that work is manual, human curation and oversight.

3.4 Discussion: Influence is Key to Engaging a Twitter Audience with Online GLAM Collections

Across the population of Twitter bots surveyed, both individual and Off The Easel members, common themes have emerged when considering the main research question of what role social bots on Twitter can and might play in influencing the sharing and consumption of history and cultural heritage. These factors cover users, developers, GLAM institutions and the platform itself.

Automating the dissemination of digitised historical content, particularly books of medieval manuscripts, prints or illustrations contained within larger texts, ultimately means changing its context, both in terms of splitting physical objects into individual pages or images and displaying them in a scrolling feed sandwiched between images of unrelated objects loosely within a similar

⁴² 'Stefano Della Bella (@artdellabella) / Twitter', *Twitter* <<https://twitter.com/artdellabella>> [accessed 3 March 2021].

collection. Hyperlinks in tweets seek to remedy this situation by offering the interested user a chance to find out more and piece together an extended historical picture, but it is likely that most will admire the images and scroll on. This is yet another drastic change for objects already contextually altered through physical curation, digitisation, digitalisation and uploading to the Web⁴³. Plus, such links are also only of use if the digitised, Open Access collections of GLAMs contain enough information, as borne out by various frustrated tweet replies from users engaging with GLAM bots who want more context or find its lack confusing. This speaks to a wider challenge for GLAMs, which has come to the fore in the light of increasing digital outreach due to the Covid-19 pandemic⁴⁴, that throwing open the doors to digital collections also means exposing the fact that little is known about many of these items that have not been subject to detailed research.

Such changes of historical and physical context brought about by the ‘feeds’ approach of social media effectively turns one object into several hundred, treating each manuscript folio as a piece of historical art in its own right. The links in the tweets back to the original image on the CMA collections website play a vital role here in allowing a user to regain the lost historical context and see the other manuscripts from the book in the context of each other, but how many users actually do this is likely far fewer than the number who simply look at the images. This phenomenon has precursors in the physical art world; Books of Hours have suffered similar fates with manuscript pages being cut out of Books of Hours to be distributed as standalone artworks⁴⁵, which seems even more alienating and permanent than its digital equivalent.

Similar bots disseminating manuscripts and book illustrations, such as the British Library’s @MechCuratorBot, choose to deal with changes of context

⁴³ Melissa Terras, ‘Opening Access to Collections: The Making and Using of Open Digitised Cultural Content’, *Online Information Review*, 39.5 (2015), pp. 733–52, doi:10.1108/OIR-06-2015-0193.

⁴⁴ Myrsini Samaroudi, Karina Rodriguez Echavarria, and Lara Perry, ‘Heritage in Lockdown: Digital Provision of Memory Institutions in the UK and US of America during the COVID-19 Pandemic’, *Museum Management and Curatorship*, 35.4 (2020), pp. 337–61, doi:10.1080/09647775.2020.1810483.

⁴⁵ University of Oxford Ashmolean Museum, ‘Ashmolean – The Elements of Drawing, John Ruskin’s Teaching Collection at Oxford’ (Ashmolean Museum, University of Oxford), World <http://ruskin.ashmolean.org/collection/8995/per_page/25/offset/0/sort_by/relevance/object/13559> [accessed 3 March 2021].

through hyperlinks to pages containing more information and an explanation of the original context. Overall, these effects of digitisation and then automation have an impact on how the user comes to understand, or misunderstand, both the artwork or artefact and more likely its context. Even if users are simply viewing these images as pieces of art that are enjoyable to look at in their own right, the knowledge they gain from viewing a manuscript page in isolation, and not as part of a book, could create a misunderstanding about the physical nature of these historical artefacts and that their function went beyond just being ornate. The algorithmic, semi-random process behind sharing such content seems to make such misunderstandings more likely to occur.

This fundamental point, which speaks to the broader issue of mass digitalisation of collections resulting in thousands of objects being accessible about which very little information is available, does not seem to have changed much since Donovan pointed it out in 1997. His statement that 'museums add so little value to the data they provide'⁴⁶ seems incredibly apt in these cases, with the sheer scale of digitalisation outstripping the ability for human curators and researchers to keep up. His associated point about the importance of narratives rather than objects also rings true when automated approaches result in piecemeal interactions with digitally-dissociated objects like manuscript pages, which is only highlighted further when many of the manuscripts to which they belong are also lacking in engaging description or narrative. This might be one area where a narrative-driven implementation of an AI-powered chatbot might be able to bridge the engagement gap between a wealth of digitised information and a broader narrative context.

The importance of social bots for these differing audiences in different situations can be summarised using the framework of the 'Digital Economy' which describes functionalities, systems that enable changes to occur, such as new modes of communication, and affordances, a set of actions that a particular group can now perform thanks to the new technologies or infrastructures⁴⁷. It is

⁴⁶ Donovan, 'The Best of Intentions: Public Access, the Web & the Evolution of Museum Automation'.

⁴⁷ Rumana Bukht and Richard Heeks, *Defining, Conceptualising and Measuring the Digital Economy* (Social Science Research Network, 3 August 2017), doi:10.2139/ssrn.3431732.

simpler, therefore, to view the social bots, and in particular the Off The Easel botnet, as a set of functionalities that create affordances for different audiences; namely affording all users the opportunity to consume history and cultural heritage content for entertainment purposes, while affording the opportunity of education only to those users for whom a lack of contextual information does not preclude gaining knowledge (i.e. most likely non-experts). Linked to this is the way that subtle biases in content are not just dependent upon available sources but also the influence of the developers, their expertise and curatorial activity.

The coordinated botnet approach employed by Off The Easel drives high levels of engagement with content, but more importantly, especially in comparison to individual bots, creates a wider community of users around the bot community that enjoy the content whilst also helping to shape and discuss the developmental choices influencing the botnet itself. While official human-run GLAM accounts may command large followings and foster social media communities, it seems unlikely that these would cut across different GLAMs and be engaging with the same volume of content that the Off The Easel botnet provides.

The interactions between the wider community of both human followers and social bots enriches the whole network, just as the ability of the bots to create influence is dependent on humans unwittingly influencing their algorithmic choice of which content to share. This complex two-way relationship means that value emerges from the actions of both bots and humans in such a way that by simply following social bots out of interest in art, cultural heritage or history, a user is actually encouraging its wider dissemination in ways they could not foresee and most probably affording far greater opportunities for others like them to come across such content online.

Social media platforms ultimately have control over any automated activity occurring on them and are becoming increasingly hostile towards it. Twitter's strict and often-evolving automation policies often make an impact with @andreitr often battling with algorithmic policing of automation that sees Off The Easel bot accounts suspended from time to time. Individual bots, without

this level of oversight or maintenance, are far more prone to disappearing from this kind of over-zealous regulation; the most notable casualty thus far has been DeathMedieval bot, an extremely popular account that shared excerpts from medieval coroners' rolls, the banning of which has left users baffled and increased calls for human oversight of algorithmic content policing⁴⁸. @andreitr has created a specific account, nude_art_bot, to share all the artworks depicting nudes across all Off The Easel art bot accounts to stop them regularly being banned and so that users know what content to expect when following these accounts. Ultimately, the sort of historical content likely to be shared by bots is influenced by Twitter.

From the perspective of content shared, developers also exert levels of control that are less obvious than determining tweeting approach but are far more important. In terms of development and content curation, most developers come with their own curatorial biases. @andreitr is not a professional artist or involved in the GLAM sector, but is an art lover with a passion for Russian art⁴⁹, hence there is a slight underlying curatorial bias in the creation of bots which promote Russian artists and Soviet artworks, a preference which the automated botnet then gradually amplifies. The far greater biases that exist come from the GLAM collections themselves and the way that these institutions organise, curate and present on their own sites and through the APIs connecting their collections to bots. Copyright and legal issues can also limit the content which is able to be shared by GLAMs and thus impact the bots sharing content from them⁵⁰.

Finally, there is the role of the recommendation algorithms built in to the social media platform themselves that exert their own level of control over what content is made most visible to whom. As Herman and Arora note when

⁴⁸ Viscount Jonkeer. , 'Okay Twitter Is Getting Truly Weird. Apparently They Have Permabanned Medieval Death Bot???? Twitter Really Needs Some Humans to Keep an Eye on Their AI before It Becomes All Powerful...', [@carlclare](https://twitter.com/carlclare/status/1351831333870444552), 2021. <<https://twitter.com/carlclare/status/1351831333870444552>> [accessed 3 March 2021].

⁴⁹ Taraschuk, 'Bots Going Artsy – Strategies for Sharing Large Art Collections with Social Media Bots and AI'.

⁵⁰ openarchief, 'The Image Collection of the @MuseumRotterdam Is Offline Due to Ruling of a Court about Copyrights. @oa_mrotterdam Stream Is Modified.', [@openarchief](https://twitter.com/openarchief/status/615099232807190528), 2015. <<https://twitter.com/openarchief/status/615099232807190528>> [accessed 10 February 2021].

discussing creative activity on the social platform TikTok more recently, the recommender system becomes another gatekeeper:

Just as previous generations of creators sought the approval of art critics, museum curators, gallery owners, or art collectors, the current generation must add another stakeholder to the mix: the algorithm. Algorithmic curators are gatekeepers that mediate the creative's access to an audience.⁵¹

In the case of the art bots, the coordinated approach to sharing content across the community of bots, as purposefully developed by Andrei, is a deliberate attempt to game the algorithmic system by having more popular bots suggest less popular ones. Overall, the analysis showed this approach was relatively effective, especially when compared against other individual bots not part of such a community. This means that for GLAMs seeking to explore such automated methods of sharing collections on social media or other algorithmically-mediated platforms, it is crucial to understand how these systems work and what that means for the bot, those who interact with it and those within the GLAM who maintain it. Human skill and resource will always be needed to achieve this, but perhaps the next generation of AI-powered bots will be able to help lower these barriers, potentially by writing their own code, analysing their own engagement metrics and automating their own schedules.

The technical constraints imposed by automation are seen in the ways that social bots source, curate and spread their content. Wikipedia and similar online, collaborative, community-based sites, like WikiArt, are the dominant information sources for automated apps and accounts and those who develop and manage them. The vast majority of Off The Easel bots share images from WikiArt and most of the automated artist accounts contain a link to Wikipedia for users to find out more about the artist in question. Amongst the other Twitter bots, many relied upon APIs to Open Access collections hosted on GLAM websites, such as the Rijksmuseum's collections. It is often such well-funded, digitally literate institutions, many of which are European, that have the resources and skills to do this, further increasing the dominance of their content

⁵¹ Herman, 'For who page?'

and above that of other smaller, non-digitised collections. Disseminating diverse interpretations, arguments and perspectives requires a larger number of trustworthy, easily automatable information sources to be available, but it also needs those with expert knowledge to be involved in the process of developing automated content dissemination tools, such as social media bots, in the first place.

The Off The Easel project also demonstrates the tough balancing act between democratising artworks and cultural heritage and paying for the infrastructure to enable this – servers, data storage and time are not free (and at this scale are very expensive). A business model could sustain Off The Easel, and Andrei has shown support for Twitter's new 'super follower' option which monetises tweets to an audience that is willing to pay (much to many users' dismay), but how to balance this with making all art freely available to everyone? In reality, the rapidly changing nature of technology on the Web threw up new Web 3 enabled products like NFT ('non-fungible token') art, artworks whose provenance was stored using Blockchain technology to ensure and define their 'uniqueness', and Andrei rebranded 'Off The Easel' to 'Bot Frens' and moved commercial efforts into this initially lucrative area. However, in spite of this transition, the 'Off The Easel' bots still remain sharing art and artefacts on X as of December 2023, although Andrei is keen to point out that given the end of free API access on the platform, it may not be financially viable to keep them up and running on X for much longer⁵².

Both the case study and the wider social bot survey have their limitations, especially in terms of data collection where finding and confirming the identity of bots was often a challenge, even on Twitter with its strict automation policies and relative transparency (at least compared to other social media platforms). For the wider bot survey, and with some of the Twitter bots not part of the Off The Easel botnet, a lack of detailed information provided by bot creators and/or managers about their bot accounts also made data collection and subsequent

⁵² Rhea Nayyar, 'Is This the End of Twitter's Beloved "Art Bots"?' , *Hyperallergic*, 8 February 2023 <<http://hyperallergic.com/799103/twitter-api-beloved-art-bots/>> [accessed 1 December 2023].

analysis harder and more subjective, with assumptions about activity or mechanism being necessarily more common than would be ideal.

The big data approach required for the Twitter case study came with its own methodological and ethical challenges, as others have observed in similar work involving scraping of social media⁵³. For example, the limits placed upon a researcher using the non-commercial Twitter API, such as returning only a sample of tweets over a variable time period, mean that such datasets will never be as complete as desired, and that was also true for this study. In addition, the data collected represents the picture available from a static time point (February 2021) for different periods of time; the samples of tweets returned from the API are based on a given amount of tweets (up to 3250) which can span two weeks or two years depending upon the activity of the bot accounts. Although this was mitigated to an extent during the analysis, it still makes comparison, contextualisation and visualisation of the data a more subjective task, reducing replicability. As of 2023, with an end to free API access following the transition from Twitter to X, future research work would have to be done through a commercial version of the API.

The context of the wider sociotechnical environment of a social bot's activity alters all the time, therefore trying to compare the influence of such external factors for data samples spanning vastly different time periods relies on mitigation measures such as averaging, normalisation or extrapolation, all of which risk obscuring the nuances in the data which big data techniques are supposed to identify and explore⁵⁴. Social bots are appearing and disappearing all the time, the algorithms available to bot developers are changing and their numbers growing, and the social media platforms, along with their rules and regulations, are changing too. The banning of one of the most popular bot accounts sharing historical content, Medieval Death Bot, prior to data collection exemplifies this potential for rapid and drastic change that academic research

⁵³ Marta Krzyzanska and Chiara Bonacchi, 'Digital Heritage Research Re-Theorised: Ontologies and Epistemologies in a World of Big Data (2019)', *International Journal of Heritage Studies*, 2019 <https://www.academia.edu/38273188/Digital_heritage_re-theorised_Ontologies_and_epistemologies_in_a_world_of_big_data_2019_> [accessed 4 December 2019]; Association of Internet Researchers, 'Internet Research: Ethical Guidelines 3.0 Association of Internet Researchers'.

⁵⁴ *ibid.*

can struggle to contextualise. This is exemplified by the recent takeover of Twitter and its transition to X which has resulted in the end of free access to its API, something that will disproportionately affect academic researchers and likely decrease the amount of research being done in this area⁵⁵. Many of these issues, whilst often recognised as challenges and limitations by researchers working with social media and big data approaches, are yet to be properly theorised and early work is underway to create stronger frameworks to critically reflect upon much of this work⁵⁶.

Along with the future of social platforms themselves, as highlighted by the rise of TikTok in recent years, it is also important to consider the arrival of ChatGPT and similar AI models. It is hard to predict much in the field of generative AI given the rapid pace of technological change, but it is already possible for a user to have a conversation with a bot during which images of artworks or objects can be shared, along with much greater description and context for each object and the wider collection to which it belongs, which may improve the experience of users who wanted more information than just images and titles. However, ongoing issues around not being able to guarantee the accuracy of text generated through these models need to be considered, along with what it would mean from a curatorial and user perspective for descriptions of collection items to potentially be generated from scratch by the model if no detailed information had been included during digitalisation. It would be easy to see this as a technical solution to the point raised by Donovan that existing collection information is often lacking or not presented in a sufficiently engaging format⁵⁷. However, applying a sociotechnical lens to the problem means considering this from the perspectives of all users, which highlights greater complexity - including the point that some of the Twitter users were perfectly happy just to receive images in the social media feeds with very little context in the first place.

In summary, this chapter has shown that a significant amount of automated (and semi-automated) dissemination of historical and cultural heritage content is

⁵⁵ Nayyar, 'Is This the End of Twitter's Beloved "Art Bots"?'.

⁵⁶ Krzyzanska and Bonacchi, 'Digital heritage research re-theorised'.

⁵⁷ Donovan, 'The Best of Intentions: Public Access, the Web & the Evolution of Museum Automation'.

occurring online, especially on social media. The algorithmic curation, recommendation and networking of bots and the content they disseminate influences the representativeness of history and heritage that users are exposed to and changes its context, sometimes drastically, creating different understandings of the past. However, automation is not passive and objective as is often perceived, human input, curatorially from bot developers, online communities that source content and from the users who consume it, shape what is disseminated, how and why, from prioritising certain artists to unconsciously biasing against certain cultures and time periods.

The research challenges of this work have highlighted the subjective nature of big data methodologies, in contrast to their supposed objectivity, as exemplified by the vagaries of human and bot agency in the complex, chaotic world of social media and its attempts at self-regulation. Ever-evolving algorithms and new approaches, such as the image recognition techniques being trialled by Off The Easel, will offer new opportunities and challenges, undoubtedly altering the contexts in which historical content finds itself when shared on social media. The potential power of automation for sharing content is clear to see, but so are the challenges, both intellectual and practical, which have reinforced the need for proper theorisation of research in this area.

4. Case study: Talking to Bots about History via Amazon Alexa

The use of chatbots is one of the fastest-growing online phenomena, with a recent report into global chatbot usage trends predicting that commercial enterprises will spend a collective \$77.6 billion on chatbots and other AI-assisted conversational technologies in 2022 alone¹. Such growth in the sector is starting to filter into the GLAM (Galleries, Libraries, Archives, Museums) sector, with some museums experimenting with chatbots as tools for visitor engagement², both physically and online³. Whilst there is a small amount of recent discussion about the potential roles of chatbots in GLAMs from a heritage perspective⁴, there is a distinct lack of information available around chatbots and historical content. As discussed in previous chapters, defining a chatbot is not always straightforward, but for this chapter the term ‘chatbot’ refers to an automated agent that interacts with user via a conversational interface. This definition is necessarily broad but still differentiates between the usually passive social bots described in Chapter 3.

A starting point for cultural heritage comes from Tzouganatou’s work surveying the use of chatbots across the GLAM sector⁵. Aside from a couple of innovative examples used to engage with visitors physically in the museum, the majority of the few chatbots discussed in the survey were not designed for interaction using free text input or speech, instead acting as condensed, guided versions of the information or contact pages already found on many museum websites. Tzouganatou aptly refers to them as ‘info bots’ rather than chatbots and suggests that for chatbots to thrive in the heritage sector they need to be

¹ BRAIN [BRN AI] CODE FOR EQUITY, ‘Chatbot Trends Report 2021’, *Medium*, 2021 <<https://chatbotsjournal.com/chatbot-trends-report-2021-b15479c404e4>> [accessed 2 June 2021].

² Ann Borda, ‘Engaging Museum Visitors with AI: The Case of Chatbots’, *Museums and Digital Culture, Springer Series on Cultural Computing*, 2018 <https://www.academia.edu/40437562/Engaging_Museum_Visitors_with_AI_The_Case_of_Chatbots> [accessed 21 February 2020].

³ Stefania Boiano and others, ‘Chatbots and New Audience Opportunities for Museums and Heritage Organisations’, 2018, doi:10.14236/ewic/EVA2018.33.

⁴ Tzouganatou, ‘Can Heritage Bots Thrive?’.

⁵ *ibid.*

designed to ‘chat’ and engage users with content from museum collections, not merely inform them about the museum⁶.

This prompted Tzouganatou to develop ChatÇat⁷, a chatbot that is part of the EMOTIVE project which seeks to engage users with some of the deeper archaeological and historical questions surrounding the Neolithic site of Çatalhöyük in modern-day Turkey. Designed as a “Bot of Conviction”⁸, this chatbot takes a rules-based approach to provoke the user with questions rather than simply provide information in the usual manner, the aim being to create a bot that truly engages its audience in archaeological, cultural and social debates.

This work forms the basis of the wider exploration of chatbots undertaken in this chapter, expanding out into history as well as heritage and considering virtual conversational assistants (VCAs) such as Alexa, and the ‘Skills’ available via the Alexa platform. Amazon Alexa itself is a service that allows users to verbally ask questions to find information, schedule tasks, play media and more⁹. When asked to find information, particularly when asked knowledge-based questions, Alexa often gets its information from Wikipedia, and if a question is too specific or complex Alexa may not provide an answer. Alexa Skills are like smartphone apps for the Alexa platform and allow developers to add custom functions and knowledge to Alexa via the Skill¹⁰, for example making a quiz that users can play which may contain answers sourced from somewhere other than Wikipedia.

Pilot work found a significant number of Alexa Skills potentially relevant to history. This is a very broad definition which covers Skills that aim to educate

⁶ Tzouganatou, ‘Can Heritage Bots Thrive?’

⁷ ‘Museum Bot - About’ <<https://www.facebook.com/pg/catalhoyukbot/about/>> [accessed 31 March 2020].

⁸ Maria Roussou and others, ‘Transformation through Provocation?’, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19 (Association for Computing Machinery, 2019), pp. 1–13, doi:10.1145/3290605.3300857; Mark Sample, ‘A Protest Bot Is a Bot so Specific You Can’t Mistake It for Bullshit’, *Medium*, 2015 <<https://medium.com/@samplereality/a-protest-bot-is-a-bot-so-specific-you-cant-mistake-it-for-bullshit-90fe10b7fbaa>> [accessed 27 March 2020].

⁹ ‘Amazon Alexa Voice AI | Alexa Developer Official Site’, *Amazon (Alexa)* <<https://developer.amazon.com/en-US/alexa.html>> [accessed 24 May 2021].

¹⁰ Amazon Alexa Developer Documentation, ‘What Is the Alexa Skills Kit?’, *Amazon (Alexa)* <<https://developer.amazon.com/en-US/alexa/techdoc-template.html>> [accessed 24 May 2021].

users by disseminating historical information (e.g. sharing historical facts for a given date), test their knowledge and entertain them via quizzes about different historical events, periods or people, and raise public awareness about overlooked histories (e.g. Black History Facts).

This chapter aimed to survey the current landscape on chatbots related to history and GLAMs. The surveys of chatbots and Alexa Skills were carried out in 2019. An initial chatbot survey was undertaken to get a broad view of the field, followed by a focussed exploration of Amazon Alexa Skills, small chatbot-like apps that work within the virtual conversational assistant Amazon Alexa, commonly accessed through the Amazon Echo smart speaker the likes of which are becoming more popular in homes across the developed world, but particularly the USA¹¹. These were chosen because Alexa is the dominant VCA on the market, its usage far outstripping the likes of Google Assistant, Microsoft Cortana or Apple's Siri;¹² it was indicated by pilot work to be the most popular way for users to access online historical information in a conversational manner, and its Skills apps have not been previously explored with regard to content about history and GLAMs. Assessing the potential role of such growing technologies in engaging a wide audience with historical content is an important step in determining how and why chatbots can come to meaningfully assist individuals and institutions in sharing historical content in novel ways.

4.1 Survey of Chatbots Relating to History and Cultural Heritage

Social messaging platforms, such as Facebook Messenger and WhatsApp, have become a feature of modern communication and their use dominates smartphone activity across all global demographics¹³. Social chat bots are flourishing on these platforms, especially Facebook Messenger, and the various

¹¹ 'Amazon Echo & Alexa Stats', *Voicebot.Ai*, 2017 <<https://voicebot.ai/amazon-echo-alexa-stats/>> [accessed 22 October 2019].

¹² Matthew B. Hoy, 'Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants', *Medical Reference Services Quarterly*, 37.1 (2018), pp. 81–88, doi:10.1080/02763869.2018.1404391.

¹³ Lorenz Klopfenstein and Alessandro Bogliolo, 'THE QUIZ-MASTER BOT: A PERSISTENT AUGMENTED QUIZ DELIVERED THROUGH ONLINE MESSAGING' (presented at the International Technology, Education and Development Conference, 2017), pp. 9806–11, doi:10.21125/inted.2017.2328.

bots available perform a wide variety of tasks, particularly in relation to business¹⁴, but no work has been done to investigate the presence of bots on these platforms related to history or heritage.

The approach taken to finding and identifying relevant bots is not dissimilar to a literature review. Two popular bot directory websites, Chatbottle¹⁵ and Botlist¹⁶, acted as the starting point for searches, just as a researcher may start a general search for articles on Google Scholar. Chatbottle only contains bots on some messaging platforms, namely Facebook Messenger, Amazon Alexa (virtual conversational assistant), Skype, Telegram, Slack and Kik, whereas Botlist covers those platforms and a wider range including Discord, LINE, Viber and WhatsApp. There are tens of thousands of bots listed across both directories, and although there will be some chatbots not listed on these sites (bots have to be submitted by developers), they still give a comprehensive picture of the bots currently available across social messaging platforms.

A combination of search terms was used to cover anything potentially related to history, historical and/or archaeological knowledge, heritage or GLAMs. The following broad terms were run across both sites in the order listed. For 'history' and 'museums', which returned a wide range of results, some more specific terms for each category were added to ensure that no relevant bots were missed from the initial broad search. The more specific search terms related to history were determined from the initial broad searches and were chosen based on the relative numbers of potentially relevant Skills identified at this stage. This explains the small number of terms and their lack of historical diversity.

Search terms used:

- history
 - roman

¹⁴ D. R. Vukovic and I. M. Dujlović, 'Facebook Messenger Bots and Their Application for Business', in *2016 24th Telecommunications Forum (TELFOR)*, 2016, pp. 1–4, doi:10.1109/TELFOR.2016.7818926.

¹⁵ '10,000+ Bots for Facebook Messenger', *ChatBottle* <<https://chatbottle.co/>> [accessed 25 September 2019].

¹⁶ 'BotList' <<https://botlist.co>> [accessed 25 September 2019].

- viking
- medieval
- heritage
- museum
 - gallery
 - library
 - archive
- archaeology

Although the more specific search terms were included, none of them returned any additional relevant results across all platforms, showing that the broader terms did indeed cover all relevant results (see 4.3.1 for results). This search strategy aimed to return as many results as possible so that relevance could be determined manually across a wide sample, decreasing the possibility that potentially related bots might be missed.

The total number of search results returned across both directories was recorded and divided into relevant and irrelevant results (see Table 7). Relevance was determined by interacting with the bots on their respective platforms, either via voice commands for Amazon Alexa or typing / navigating menus for the others. A bot was considered 'relevant' if it fulfilled three criteria:

1. conversational content was primarily aimed at presenting information about history, archaeology or museum collections
2. chat functions were fully automated without any real-time human input - if this was ambiguous (especially for Facebook Messenger bots) the bot was deemed irrelevant
3. the chatbot functioned properly, responding to initial queries with a coherent reply that was not simply a set template answer.

This definition of relevance was still subject to some interpretation where functionality was only partial. For example, with some chatbots that could carry out multiple functions, an initial interaction might show that some of these functions worked whilst others did not. In these cases such bots were considered relevant so that the usable functions could be analysed.

The task of identifying chatbots proved particularly difficult on Facebook as some chatbots are embedded within official, human-run pages, such as those for museums, which may not show up in searches from bot directories. For this reason, it is expected that the numbers returned for the 'museum' search term will be lower than is actually the case, but this demonstrates the inherent difficulty of both finding and identifying automated activity.

Searches were carried out as detailed above in the methodology section. Overall, there was a far higher percentage of relevant results for Alexa (see Table 7). The numbers of total and relevant results found are probably underestimates as not all Skills will be added to bot directories by developers and, despite the broad search strategy employed, some creatively named bots will doubtless have slipped through the net. Across other available messaging apps, like Telegram, WhatsApp, Slack, Discord etc., only one relevant result was returned and that was for the 'Imperium Romanum' bot on Telegram that also appears in the Facebook Messenger (this is the only relevant Messenger bot for the search term 'archaeology').

The results in Table 7 show that there were very few relevant chatbots found across FB Messenger. For the few relevant FB Messenger bots there was more of a tangential relevance to historical content, for example chatbots for historical tour agencies, academic research centres and promoting upcoming historical films. Again, this emphasises the point that chatbots on Facebook are primarily seen as commercial entities aimed at boosting visitor engagement with a company, brand or institution rather than its content.

The term 'heritage' returned only one relevant result for FB Messenger, a chatbot for a heritage trail that followed the usual customer-service style approach providing practical information but very little in terms of historical content. The vast majority of the total results returned were companies or

brands with ‘heritage’ in the name, often to do with fashion, which says something about the idea of history and heritage in branding and again demonstrates the commercial nature of FB Messenger chatbots.

Facebook Messenger			
Search term	Total results	Relevant results	Percentage of relevant results (%)
history	865	14	1.6
heritage	260	1	0.4
museum	139	33	23.7
archaeology	15	2	13.3
Totals	1279	50	3.9

Table 7: Table showing results of the chatbot survey across social messenger platforms; relevance determined as above in the methodology.

The high total number of ‘museum’ results for FB Messenger highlights the point about many results being pages about museums or human-run communities, not actual chatbots, which led to a high number of irrelevant results. It also shows that Facebook is the primary platform for GLAM organisations trying to reach as many potential visitors as possible. Nearly all of the relevant ‘museum’ bots, both on FB Messenger and Alexa, were of the closed, information-based, customer service style chatbots commonly used by commercial businesses and described in detail in the heritage sector by Tzouganatou¹⁷. Given this and the lack of relevant results, Alexa Skills were prioritised as the main focus of this case study.

4.2 Amazon Alexa Skills Survey

¹⁷ Tzouganatou, ‘Can Heritage Bots Thrive?’.

As described in the introduction, Amazon Alexa itself is a service that allows users to verbally ask questions to find information¹⁸. Skills, which are like smartphone apps for the Alexa platform, allow users to access extra content and so have more focussed interactions with Alexa¹⁹. The brief, prescriptive and app-contained user interactions with Alexa and Alexa Skills take slightly different forms, but all demonstrate the relatively straightforward approach taken by Skills to engaging the user with the historical information (see Figure 20). The screenshots in Figure 20 come from the smartphone app for Alexa, for which the displayed textual summary is read aloud; but for users accessing Alexa just via a smart speaker, the narrated text is all the information they will receive.

The large number of Skills available for Alexa relating to history make it by far the most relevant platform for this study (see section 4.3.1, Table 1). Google Assistant also supports its own chatbot-like apps, but searches across Botlist and Chatbottle returned only one result relevant to history, heritage, museums or archaeology, compared to 141 for Alexa.

¹⁸ 'Amazon Alexa Voice AI | Alexa Developer Official Site'.

¹⁹ Amazon Alexa Developer Documentation, 'What is the Alexa Skills Kit?'

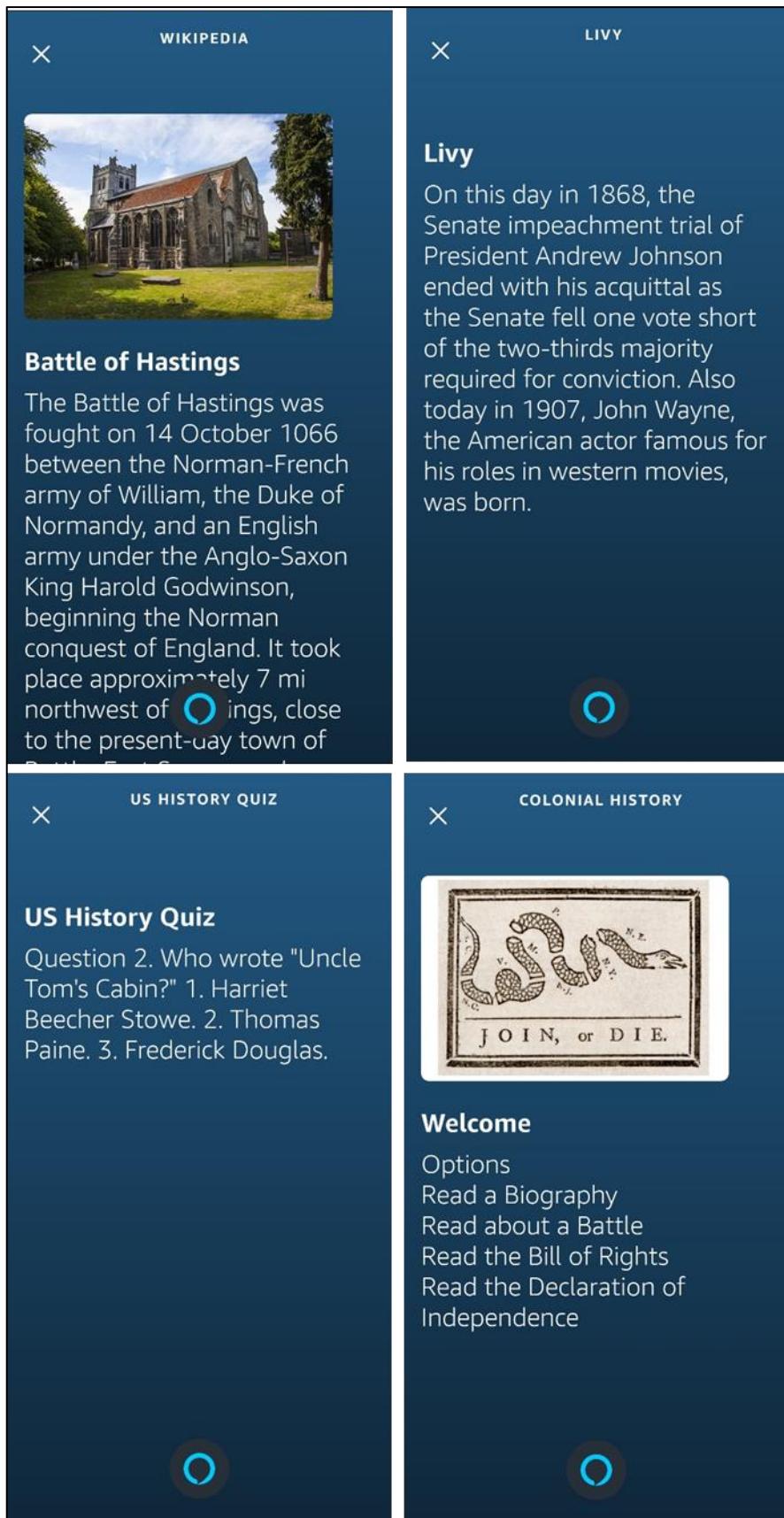


Figure 20: Screenshots of interactions with Alexa on a smartphone. The top left request was carried out by Alexa itself and not an additional Alexa Skill. Top right to bottom left show three screenshots of interactions with different Alexa Skills.

For each Alexa Skill identified from the search strategy, the Skill's name, description (as provided by its developer), number of ratings, content source and category were collected. Occasionally, probably as a result of a continuing development process, Skills with the same content appeared twice under different names and in such cases only the first occurrence of the Skill was recorded. In the opposite case, there were a few skills with the same name but different content; for instance, there were five Skills collected called 'Today in History', each of which was collected individually.

Number of ratings was used as a proxy measure to give an idea of the number of users of each Skill. As Alexa Skills are found and enabled through the Amazon store they come with the usual star rating and product review options; although the ratings and reviews themselves often have little to do with the content of the bot and may not reflect its quality²⁰, the number of ratings does give an idea of relative usage. Granted, most users will not bother to leave a rating or review, but if it is assumed that the proportion who do so is consistent across the population of Alexa users then some general inferences about audience size may be possible.

Information about the content source for the Skill (e.g. Wikipedia) was collected where possible, although often this was not specified in the description and the source was still unclear even after interacting with the Skill. In these cases, the content source was listed as 'unknown'. For the majority of the quiz Skills where the content source was not specified it was assumed that content had been 'individually curated' by the developer as the compiling of question and answers is very likely to be done manually. Such individual curation may well have involved other sources like Wikipedia, but this was never provable from the information given on the Skill page.

Categorising the different types of content and interaction offered by each Skill through talking to Alexa required a mixed methods approach, quantitatively categorising Skills based on existing tags and through the creation of new groupings based upon content type and delivery. Amazon lists category tags for

²⁰ Tobias H. Engler, Patrick Winter, and Michael Schulz, 'Understanding Online Product Ratings: A Customer Satisfaction Model', *Journal of Retailing and Consumer Services*, 27 (2015), pp. 113–20, doi:10.1016/j.jretconser.2015.07.010.

available Skills such as 'games' or 'education'; all of the Skills collected came under one of four category tags: 'education', 'trivia', 'games' or 'podcasts'. These tags are assigned by developers when listing Skills on the Alexa store, based on their decisions about which existing category their Skill fits into best. These categories are specified by Amazon and relate to types of interaction rather than content, so there is no 'history' category, but history content may fall under 'education', 'trivia', 'quiz' or 'game'. Therefore, there is little consistency in defining what differentiates an 'education' Skill from a 'trivia' one and the two categories are often used interchangeably for Skills that relate facts to users. Given this ambiguity, each Skill was manually assigned to one of five 'content categories' based on my interpretation of the Skill's description and initial interaction with the Skill:

- 'Quiz' – this covers any content delivered as a quiz or a series of questions (sometimes referred to as trivia rather than a quiz)
- 'Daily' – this includes any Skills which provide content in the set format of births, deaths, events or other facts occurring 'on this day in history', sometimes delivered automatically by Alexa as part of a daily Flash Briefing
- 'Facts' - this is the most general, covering any Skill which presents historical information to the user, or allows them to find it, in a non-quiz or open-ended format not restricted by date
- 'Game' – this mirrors the developer-assigned tag and covers any Skill that is an interactive game other than a quiz
- 'Podcast' – again this mirrors the developer-assigned tags and covers any Skill listed as a podcast

These content categories were used in all future analyses.

The actual historical content provided by the Alexa Skills was captured through voice interactions with each Alexa Skill and then analysed through content analysis of transcripts. Reflections upon the quality of the interactions, especially noting if any errors occurred, were also noted down for each Skill.

When used on a device with a screen (smartphone, TV or PC), Alexa always replies to a request with speech and sometimes a textual transcription, which may be accompanied by an image (see Figure 20). All speech interactions between the user and Alexa are automatically recorded and transcribed by Amazon (note: transcription here means simply saving the text that Alexa generates from the speech-to-text algorithm that powers its voice recognition). Whilst this has rightly been the focus of much criticism and discussion around Amazon's privacy policies and use of users' data without their full knowledge²¹ (although Amazon states that all data are saved to improve the quality of the voice recognition system²²), it does mean that all text transcripts, which can be accessed by users after trawling through Amazon Alexa's privacy pages, were available for analysis. Again, Amazon does not make these easy to access and users cannot simply download them; instead a laborious manual process of copy and paste was required to collect every single interaction into usable text documents (see Appendix 1 for examples of these transcripts). It should be noted that this only occurs when Skills use Alexa's voice to respond to the user; for narrated content, especially podcasts, Alexa simply plays pre-recorded content and so this is not transcribed, therefore the three podcast Skills were excluded from content analysis. In terms of interacting with Alexa, mindful that everything was automatically recorded (although this did not prevent curse-filled outbursts born of frustration), I chose to formalise the process and approach each interaction as a structured interview with an extremely formulaic participant. This adaptation of an 'interview' method from qualitative research remained grounded in the fact that I was speaking to an algorithm which, in the case of nearly all the Skills, followed predictable patterns usually hard-coded and completely devoid of anything resembling 'intelligence'. This choice was in no part an attempt to anthropomorphise the chatbot / algorithm with which I, the human, was interacting; quite the opposite – its highly structured, yet

²¹ Anne Pfeifle, 'Alexa, What Should We Do about Privacy: Protecting Privacy for Users of Voice-Activated Devices', *Washington Law Review*, 93 (2018), p. 421.

²² 'Amazon.Co.Uk Help: Alexa and Your Privacy' <<https://www.amazon.co.uk/gp/help/customer/display.html?nodeId=GVP69FUJ48X9DK8V>> [accessed 3 January 2021].

subjectively flexible framework was developed with my reasoning and decision-making abilities in mind and not those of the algorithm.

The flow chart in Figure 21, loosely described as an interview schema, was developed from this mindset and followed principles commonly used in the testing of chatbots (including Alexa Skills) in industry where such work comes under the heading of ‘user experience’ or UX testing²³. The focus on consistency of process can make it appear as if the human ‘interviewer’ is being reduced to something little better than a bot themselves, but despite this rigid structure, designed to try and capture three interactions with each Skill to see whether content was quickly repeated or not, the conditional decisions present at each arrow were ultimately always mine to make.

²³ Nicole Radziwill and Morgan Benton, ‘Evaluating Quality of Chatbots and Intelligent Conversational Agents’, p. 21.

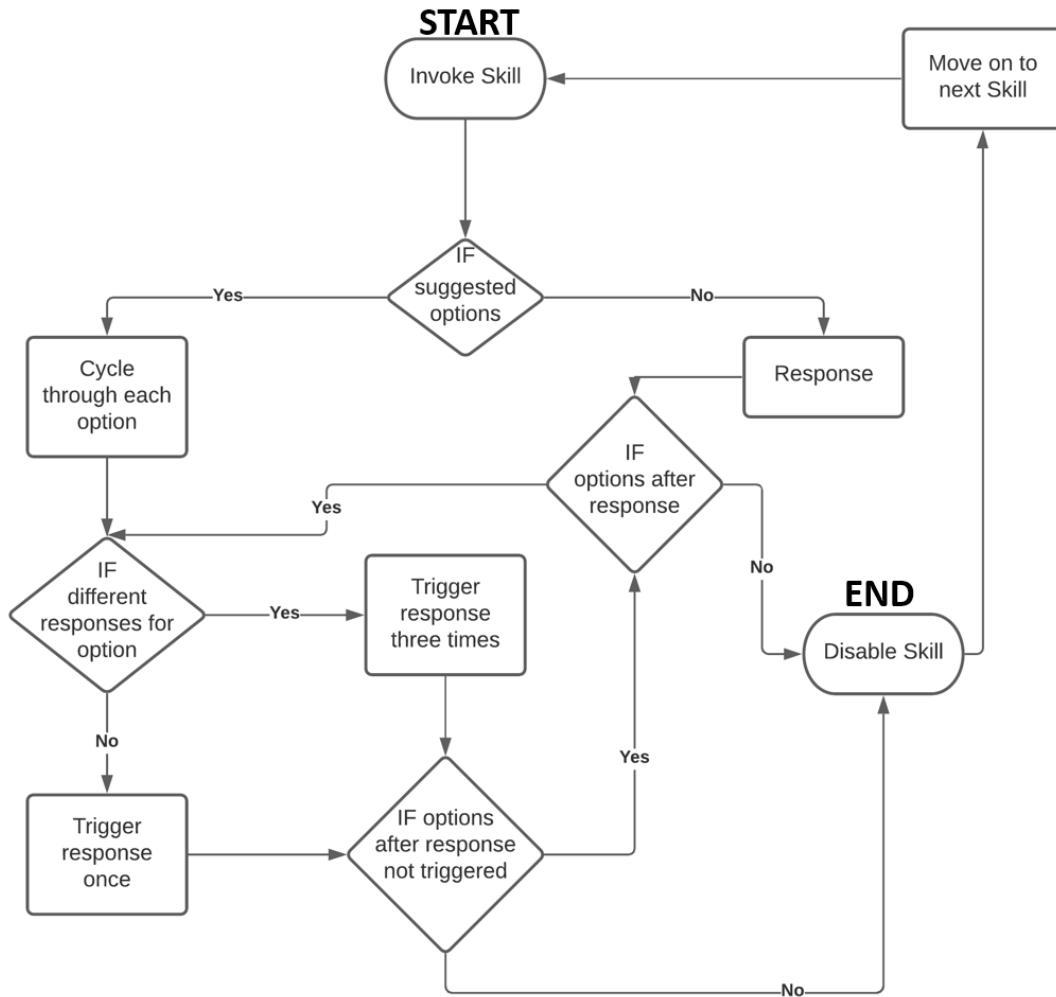


Figure 21: Flow chart detailing the interview schema for interactions with Alexa Skills.

The interview process starts by ‘invoking’ a Skill, this means telling Alexa to open the Skill so that the user can interact with it, e.g. “Alexa, open This Day in History” or “Alexa, ask History Facts to tell me a fact”. These phrases, known as ‘invocations’, trigger the Skill to start and have to be carefully chosen by developers; if a Skill has the same name as another, as happened in a few cases in the list of history Skills, the invocation will need to be subtly different so that Alexa can differentiate between them and open the desired Skill.

Once started, some Skills present users with a list of options whilst others simply jump straight in and start providing information. Therefore, if a Skill did provide options I would go through the rest of the process for each option in turn to ensure that all the Skill’s functions had been fully explored. In effect, this

would mean going through all the possible responses, a response being the utterance that actually contains the relevant content (e.g. “Here’s your fact: Vomiting was not a regular part of Roman dining customs...” – see Appendix 1, Example 2).

As noted above, this process was designed with flexibility in mind, much like a standard interview schema, and this is most apparent when deciding how many responses to record for a given Skill. As a general rule, if there were no options given upon starting the Skill, or the Skill immediately recited a fact, then I would restart the Skill three times to see if the response was different each time (see Appendix 1, Example 2). This was not uncommon for Skills designed to provide facts, although there was some variability as such Skills might ask the user to choose a date or provide more than one fact per response (see Appendix 1, Example 1). When interacting with fact-style Skills that provided information for given dates, the option chosen was always ‘today’. After a response had finished, some Skills asked the user if they wanted more or to hear another fact, whereas others (including most quiz Skills) simply ended or said “Goodbye”. If there was an option to hear more, this would be done until three responses were recorded. After the process had been completed that Skill was disabled and I moved onto the next one.

If an error occurred when trying to get a response, such as the Skill being unable to access an external data source, the next option in the Skill was attempted or the Skill was disabled. Such technical errors were rare, but far more common was misunderstanding and/or mishearing of speech, which had to be worked around by trying similar phrases or just being persistent. This was more of a problem when attempting to enable and invoke Skills by voice, especially with some Skills having the same names, which often meant resorting to a manual, button-pressing approach of enabling and disabling Skills via the Alexa smartphone app Skills store. This hybrid of button-pressing and speech to interact with Alexa is likely representative of the majority of real-world

user interactions given that Alexa is most commonly used via its smartphone app²⁴.

Users find and enable Alexa Skills primarily through the Amazon store page for Skills which is accessible on any compatible device with a screen, most likely using the Alexa app on a smartphone. This means that searching for ‘history’ will be the most common way for users to interact with a history Skill, but once on the Amazon store page for a Skill a list of recommended Skills that “other customers have enabled” will also be present. The aim of this analysis is to explore how likely it is that a user who has found a given history Skill will then be recommended another history Skill. This gives an idea of the role Amazon’s recommender system plays in shaping users’ interactions with history Skills and how this influences the historical content they are likely to experience.

Network analysis is well-suited to studying recommendations like these because clicking through from one suggested item to another soon creates a network of links between dozens of different items that can be quantified. Three measures are used in this approach: the number of times a Skill *recommends* another Skill (its ‘outdegree’), the number of times a Skill is *recommended by* another Skill (its ‘indegree’), and the number of times a Skill in the recommendation network can be clicked through in a given chain of recommendations (its ‘stress centrality’, a marker of how much ‘stress’ it is under in the network). The more ‘stressed’ a Skill, the more influential it is in the overall ability of history Skills to recommend each other to users. It is important to note that the network itself only includes history Skills and not all the Alexa Skills available on Amazon, so the network statistics are showing the influence of history Skills within the community of other history Skills and are not measures of overall popularity across all available Skills.

To collect the data to build the network, the Amazon store page of each collected history Skill was accessed and each of the ten recommended Skills recorded for that Skill. Out of these ten recommended Skills, only those in the list of history Skills were included. Completing that task for every history Skill

²⁴ ‘Twice the Number of U.S. Adults Have Tried In-Car Voice Assistants as Smart Speakers - Voicebot.Ai’ <<https://voicebot.ai/2019/01/15/twice-the-number-of-u-s-adults-have-tried-in-car-voice-assistants-as-smart-speakers/>> [accessed 25 October 2019].

thus creates a network of links between Skills showing the overall picture of how many history Skills recommend how many others (see Figure 22). For example, in Figure 22 the 'History Podcasts' Skill recommends two Skills (outdegree = 2) but is not itself recommended by other Skills (indegree = 0). The network obtained can then be analysed as described above to show which Skills have the greatest influence in recommending other history Skills.

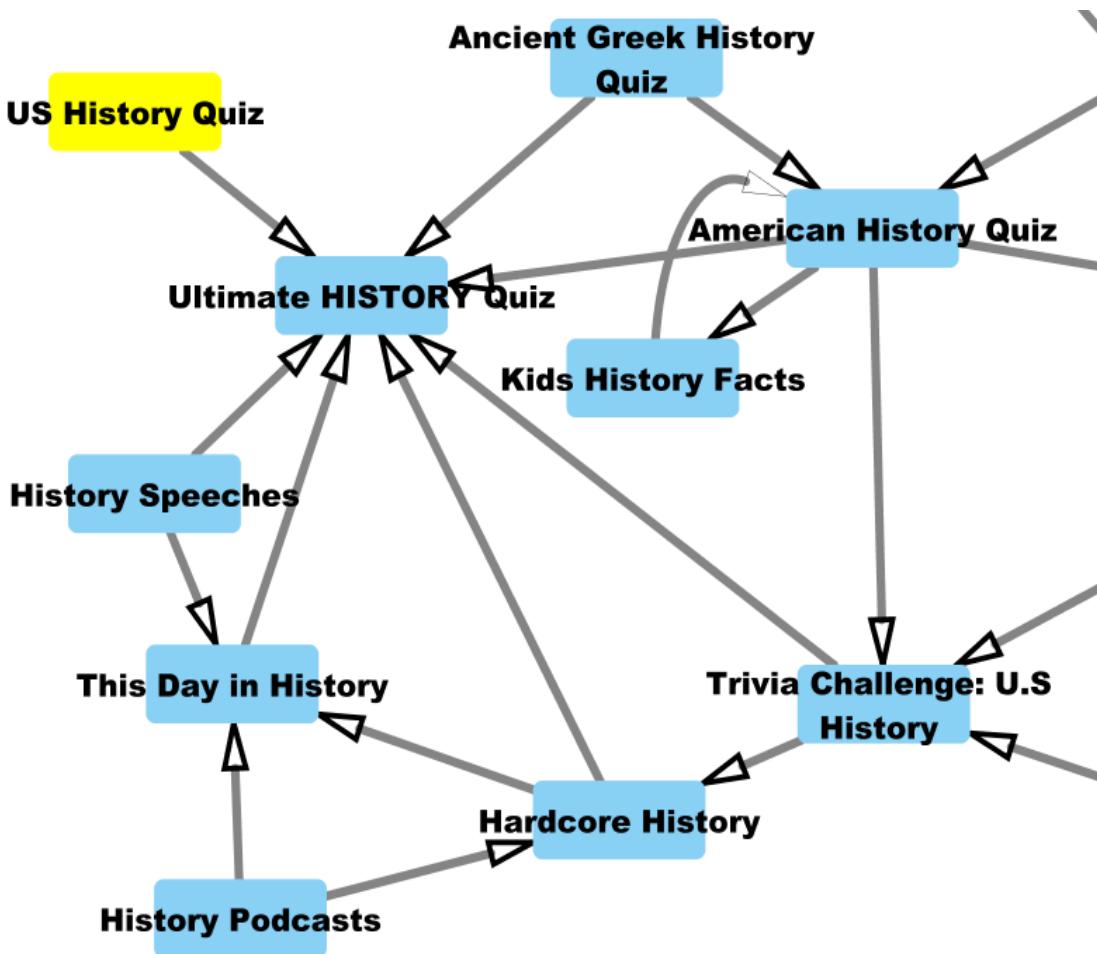


Figure 22: Visualisation showing part of the recommendation network for the history Skills. The direction of the arrows indicates which Skills recommend others (outdegree) and which are recommended by others (indegree).

Amazon Alexa			
Search term	Total results	Relevant results	Percentage of relevant results (%)
history	635	127	19.7
heritage	28	7	25.0
museum	26	10	38.5
archaeology	2	1	50.0
Totals	691	145	20.7

Table 8: Search results returned for bots on Amazon Alexa; relevance determined as above in the methodology.

For the term ‘history’, which returned by far the largest amount of total results across the four search terms, Alexa had a reasonable percentage of relevant results (19.7%). The relevant results for Alexa covered a range of historical topics, from the history of specific countries to quizzes about historical events.

Of the seven relevant results for ‘heritage’ for Alexa, content was only partially relevant to history with the main focus being on world culture and geography. For ‘museum’, Alexa had 38.5% relevant results and ‘archaeology’ returned one relevant result aimed at a more expert audience as it provided daily updates about the latest news in archaeology. Perhaps the relative specificity of this search term contributed to the overall lack of results but greater relevance and focus of content aimed at a specific audience, unlike many of the Skills found for the other, broader search terms.

The number of bots returned for the search term ‘history’ on Alexa (n=127) was by far the largest across all searches (see Table 8). Analysing the different types of content and interaction offered by each Skill through talking to Alexa required a mixed methods approach, quantitatively categorising Skills based on existing tags and through the creation of new groupings based upon content type and delivery. As described in the methodology section, each Skill was

manually assigned a ‘content category’ based upon its description and content. Figure 23 shows the distribution of Skills across these content categories.

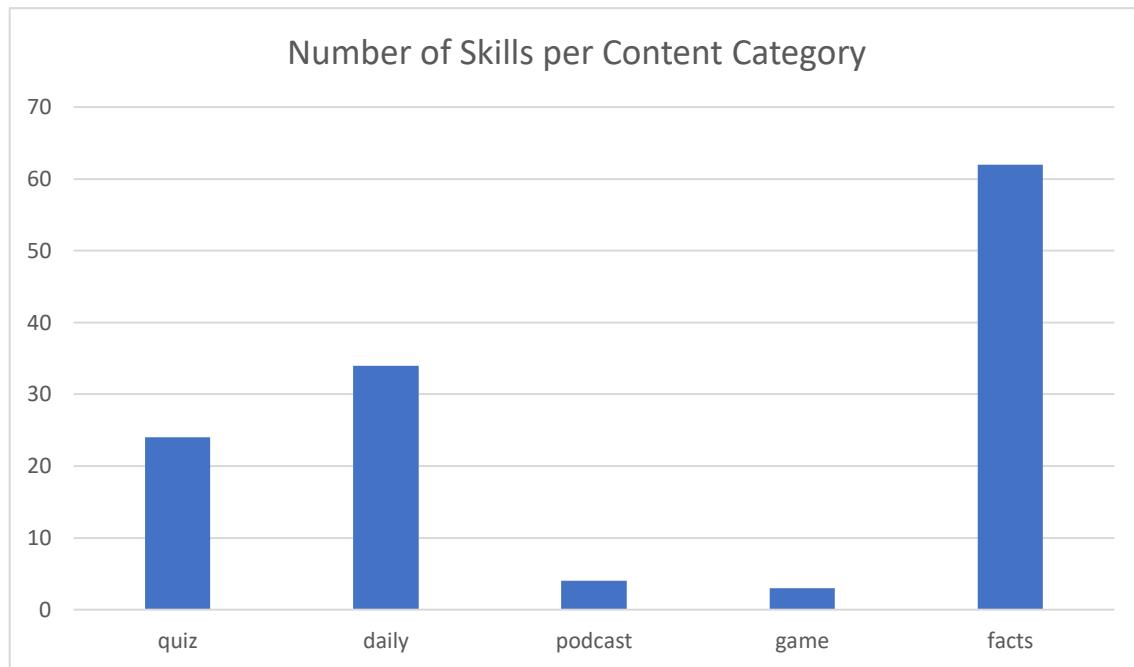


Figure 23: Bar chart showing the number of Alexa Skills in each manually-assigned content category.

Interestingly, although the facts category accounts for almost half of the Skills ($n=62$), the number accounted for by the quiz and daily categories ($n=58$) demonstrates the different types of interaction possible between a user and a VCA like Alexa. There are three main modes of dialogue between Alexa and a user: the predominant one, marketed to death by Amazon, is for the user to ‘Ask Alexa’; similar to this, if a user enables a Skill as part of a daily Flash Briefing (like a brief, verbal newsfeed) or even a wake-up call (like a radio alarm clock), Alexa will ‘tell’ the user; and if the user chooses to enable a Skill that is a quiz, Alexa can ‘ask the user’. This last option accounts for all Skills in the quiz content category, whereas for the majority of the Skills found in the facts and daily content categories the emphasis is still on the user to ask Alexa for information they did not know, or for general historical enlightenment about a given date, event or person (usually just dates). Therefore, the Skills in the daily content category exemplify a passive interaction with Alexa and a willingness on the user’s part to ‘be told’ information rather than actively ‘asking for’ it, which is

required by most of the Skills in the facts content category. This is seen in Figure 24, which shows the most frequently used words in the descriptions of Alexa Skills and the associations between them, where ‘ask’ and ‘tell’ both feature in the word cloud (ubiquitous words not of interest, like ‘Alexa’, were removed before analysis). Interestingly, the graph showing links between words in Figure 24 shows that ‘ask’ is more commonly associated with ‘day’ whilst ‘tell’ links more strongly to ‘events’, a common case being a user being prompted to ask “What happened on this day?” and Alexa responding by telling them about a historical event. This format of the user asking and Alexa telling occurs across many of the fact-style Skills.

The quality and quantity of historical content delivered to the user differed depending on the developer, the mode of interaction they had chosen and the source(s) from which to retrieve the information. As seen in Figure 25, in the majority of cases it was impossible to determine with any certainty where a developer had sourced their information from, as there is no requirement to note this in the Skill’s description. This applies to many of the Skills in the facts content category and ties in with the fact that the majority of developers for all of the collected Skills are unaffiliated individuals (n=119), with ‘Appbly.com’, a small development company in San Francisco²⁵, accounting for five Skills and ‘A&E Television Networks Mobile’, part of the television company that owns the History Channel in the USA, having developed three Skills.

²⁵ ‘Home’, *Appbly.Com* <<https://appbly-com.weebly.com/>> [accessed 25 October 2019].

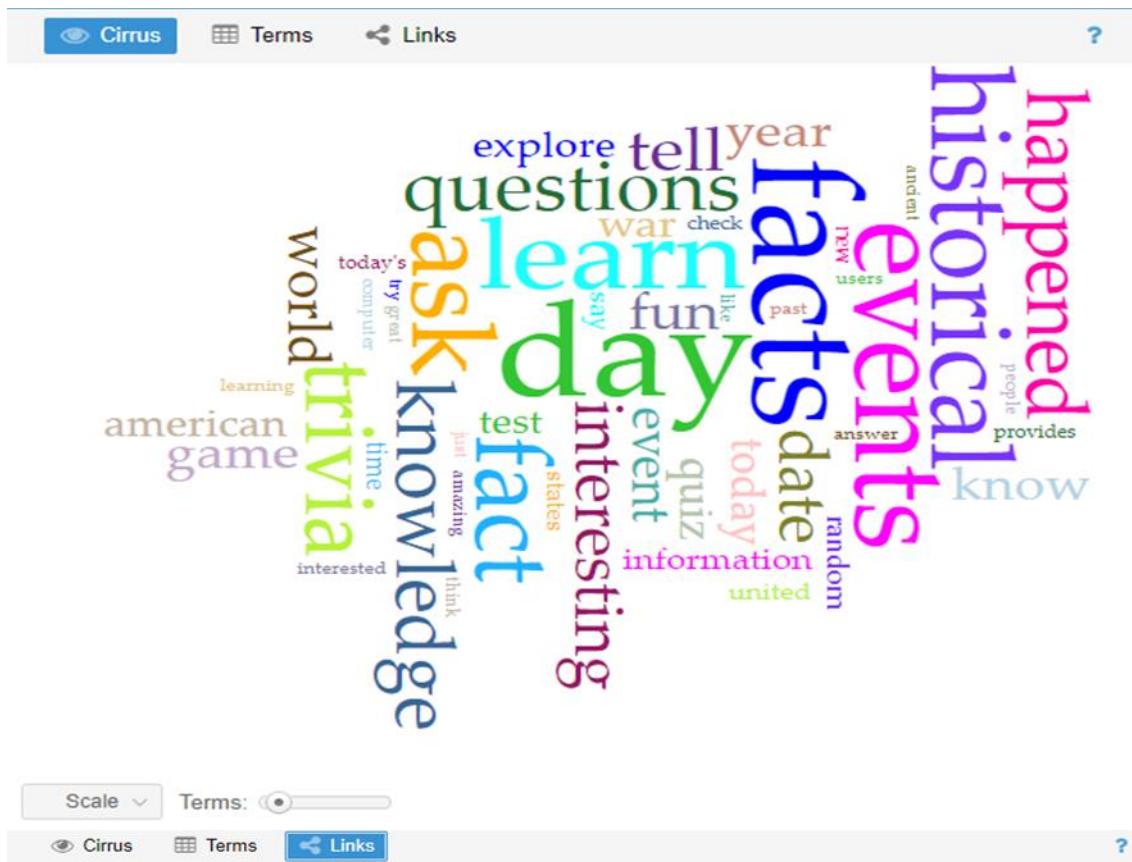


Figure 24: Top: word cloud showing most frequent terms used in Skills' descriptions. Bottom: collocate network showing terms' most frequent associations

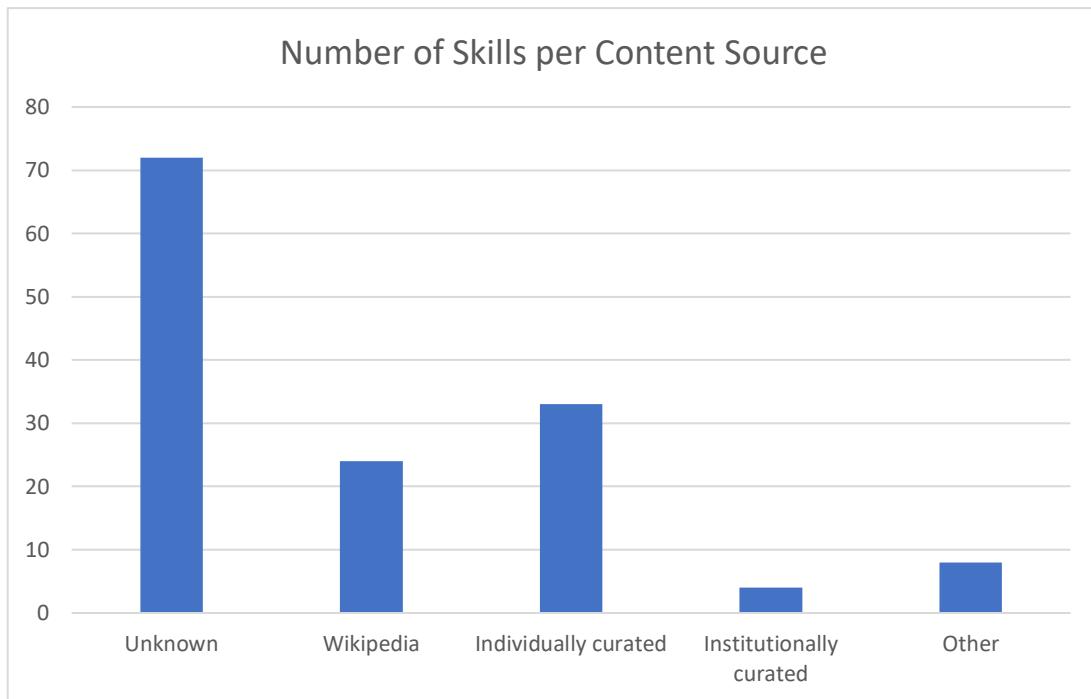


Figure 25: Bar chart showing distribution of content sources across Skills for the search term 'history'.

In the quiz and daily categories, a pattern of content sourcing is apparent. Any quiz, by design, needs sets of questions and answers pre-packaged by developers, and therefore comes under the designation of 'individually curated' as this information is drawn together from different sources, although where from is also, in most cases, unknown. Maybe it is Wikipedia, maybe the developer's head, possibly, in the case of the 'Ultimate HISTORY Quiz' Skill affiliated with the History Channel, this curation may be done by experts, but there is no way of knowing this. Indeed, in the reviews for some quiz Skills developed by individuals, users complained that answers were inaccurate:

Could be a great app if more questions and the answers were accurate²⁶ ('History Trivia' Skill).

²⁶ 'Amazon.Com: History Trivia: Alexa Skills' <<https://www.amazon.com/sawjax-History-trivia/dp/B0754Q8MML?ref=chatbottle>> [accessed 2 October 2019].

Many of this quiz's "correct " answers are wrong!²⁷ ('American History Quiz' Skill).

From my experience of testing each Skill, answers on quiz Skills were correct (for quizzes that functioned properly – some repeated answers for previous questions), but this was just on one run-through of questions. All quiz Skills were multiple choice answers as saying a number as an answer is far easier for Alexa to consistently recognise, but this means that many of the questions and answers are of a similar fact-type style (see Table 8 for examples). User reviews for History Channel's 'Ultimate HISTORY Quiz' rarely questioned accuracy of content, instead complaining about the format of the quiz and its role as entertainment as well as education:

I only have three stars up because I'm also rating the old version, which was much better. The true/false format leaves a ton of room for just luck and simply isn't as fun. Also, get rid of the facts at the end of the questions, I'm playing a history quiz, not reading a history textbook.²⁸ ('Ultimate HISTORY Quiz' Skill).

Another issue with creating quizzes from information sourced individually and manually is that it needs updating over time, otherwise Alexa will start to repeat the same questions. For the History Channel's quiz, with ample resources behind its development, this is not an issue, but for many of the individually developed quiz Skills this is a limiting factor (which also applied for some of the facts Skills too). Repetition was a common user complaint:

The only thing is it needs new History stories as it keeps repeating .
Update please?²⁹ ('Fake History' Skill).

²⁷ 'Amazon.Com: American History Quiz: Alexa Skills'
<<https://www.amazon.com/dp/B01F601P7A/?ref=chatbottle>> [accessed 8 October 2019].

²⁸ 'Amazon.Com: Ultimate HISTORY Quiz: Alexa Skills'
<<https://www.amazon.com/dp/b075zs916k/?ref=chatbottle>> [accessed 8 October 2019].

²⁹ 'Amazon.Com: Fake History: Alexa Skills' <<https://www.amazon.com/Tellables-Fake-History/dp/B074DKP2BK?ref=chatbottle>> [accessed 2 October 2019].

Great the first day. on the second day, the same questions³⁰ ('US History Quiz' Skill).

This is a big problem for individual developers as creating a Skill is time consuming and maintaining it is an even greater commitment, every Skill update needing to pass the same battery of functionality and content tests required by Amazon before a Skill can be published live on the Alexa store³¹. Some developers have noted the issue of users soon finding content repetitive, such as the team behind the 'BBC History Magazine Quiz' Skill which includes this message at the end of the quiz:

Thank you for playing! More quizzes should be added soon.³² ('BBC History Magazine Quiz' Skill).

However, this quiz Skill only offers questions about the Battle of Waterloo and more topics have yet to be added. Across many Skills there appears to be an oft-repeated cycle of poor content leading to low user engagement, and therefore a lack of willingness among developers to maintain and update Skills, which were perhaps only created in the first place as an experimental foray into the new, relatively niche Alexa market (outside of the US at least).

³⁰ 'Amazon.Com: US History Quiz: Alexa Skills' <<https://www.amazon.com/Quyen-US-History-quiz/dp/B07432FYRY?ref=chatbottle>> [accessed 2 October 2019].

³¹ Amazon Alexa Developer Documentation, 'Certify and Publish Your Skill', *Amazon (Alexa)* <<https://developer.amazon.com/en-US/docs/alexa/certify/certify-your-skill.html>> [accessed 18 May 2021].

³² 'BBC History Magazine Quiz: Amazon: Alexa Skills' <https://www.amazon.co.uk/Immediate-Media-Co-History-Magazine/dp/B07D9TXPH7/ref=sr_1_2?dchild=1&keywords=bbc+history+magazine&qid=1622623317&s=digital-skills&sr=1-2> [accessed 2 June 2021].

Skill Name	Example Question	Example Answer Choices
US History Quiz	Which of the following Founding Fathers was never president?	1. George Washington. 2. John Adams. 3. Ben Franklin.
History Quiz	During the Wars of the Roses (1455 - 1487) which Englishman was dubbed 'the Kingmaker'?	1. Richard III. 2. Richard Neville. 3. Henry V. 4. Thomas Warwick.
Persian History Quiz	To govern a far-flung empire consisting of more than seventy distinct ethnic groups, the Achaemenid rulers:	1. Used imperial spies to control the conquered masses. 2. Decentralized their administration. 3. Forced the peoples to speak only Persian and believe only in the Persian religion. 4. Established lines of communication and centralized administration.
Ancient Greek History Quiz	Through what defensive alliance of Greek city-states, established just after the Persian wars, did the Athenian statesmen Pericles divert funds for the beautification of Athens?	1. Delian League. 2. Athenian Alliance. 3. League of Corinth. 4. Peloponnesian League.

Internet History Quiz	Larry Page and Sergey Brin started this company:	<p>1. Google.</p> <p>2. Amazon.</p> <p>3. Facebook.</p> <p>4. Yahoo.</p>
BBC History Magazine Quiz	This is a quiz about the Battle of Waterloo. I will ask you 8 questions, try to get as many right as you can. Just say the number of the answer. Let's begin. Question 1. The fighting in and around the village of Plancenoit was a bitter contest between which two forces?	<p>1. French and Nassauers.</p> <p>2. French and Brunswickers.</p> <p>3. French and Hanoverians.</p> <p>4. French and Prussians.</p>

Table 9: Example quiz Skill set of questions and answers for different historical periods and places.

For Skills in the daily content category, Wikipedia was the most common source of information. There are two main reasons for this: firstly, there are lots of pages on Wikipedia dedicated to historical events that happened on this day, ranging from births and deaths of notable figures to events in the history of aviation; secondly, these pages are easily and automatically accessible through Wikipedia's API (Application Programming Interface) which means that the developer can automate the process of collecting said facts for any given date and returning a randomly-selected one to Alexa to then read out to the user. This means that the developer's job is relatively straightforward and that the user is unlikely to hear the same facts over again. In some cases, where the topic area is too specific to match up to a Wikipedia page, other sources may be used. The Skill 'This Day in US Naval History' has no corresponding Wikipedia page from which to source its information, and it's description gives no clues as to where its daily facts come from, but a quick Google search and interacting

with the Skill reveals the source to be the US Navy's very own 'This Day in History' webpage³³, which the Skill parrots word for word. Some specialist online sources such as this are used for other Skills, including MagicPedia³⁴, a wiki for magic and its history, and Wookiepedia³⁵, a fan-curated wiki for Star Wars and its history. As wikis, with community curation and management, these act in a similar way to Wikipedia but on a smaller scale.

Although Wikipedia, and other wikis, lend themselves to automated retrieval of relevant historical information, it also serves as a starting point for manual, individual curation of 'facts' for dissemination by people or bots. In one example, a Skill entitled 'History Facts'³⁶ which aimed to disseminate "historical facts that are misconceptions from Wikipedia"³⁷, had been populated with content manually by the developer, who had read Wikipedia articles and selected apparently interesting facts, for example:

Despite the terrible nature of and damage caused by the 1666 Great Fire of London, only 8 people were killed. This is despite the fire destroying at least 13,500 houses.³⁸ ('History Facts' Skill).

The Wikipedia article for the Great Fire of London³⁹ contains this information, but it also states that while:

Porter gives the figure as eight... Hanson takes issue with the idea that there were only a few deaths, enumerating known deaths from hunger and exposure among survivors of the fire⁴⁰ (Wikipedia)

³³ 'This Day in Naval History - Oct. 25' <https://www.navy.mil/search/display_history.asp> [accessed 25 October 2019].

³⁴ 'Magicpedia' <https://geniimagazine.com/wiki/index.php?title=Main_Page> [accessed 25 October 2019].

³⁵ 'Wookieepedia | FANDOM Powered by Wikia' <https://starwars.fandom.com/wiki/Main_Page> [accessed 25 October 2019].

³⁶ 'Amazon.Com: History Facts: Alexa Skills' <<https://www.amazon.com/dp/B01NB9Y3C5/?ref=chatbottle>> [accessed 8 October 2019].

³⁷ *ibid.*

³⁸ 'Amazon.Com: History Facts: Alexa Skills'.

³⁹ 'Great Fire of London', *Wikipedia*, 2019 <https://en.wikipedia.org/w/index.php?title=Great_Fire_of_London&oldid=919888587> [accessed 8 October 2019].

⁴⁰ *ibid.*

Here, the nuance is lost. Whilst the Wikipedia pages that concisely state the plain facts of things that happened on this day are perfectly designed for dissemination, more nuanced historical debate that questions the nature of such ‘facts’ is not. Even when Alexa is asked a direct question, outside of a Skill, and finds a Wikipedia page as a possible answer, only the summary of that page is read (and shown) to the user. For nearly all bots, on any platform and whether commercial or not, concision is a design feature. A desire for short, sharp, definite answers, both on the parts of users and developers, does not sit comfortably alongside nuanced academic arguments around history, or any other subject for that matter. In this case, it is also clear to see that Wikipedia, a source often maligned by academia (especially the humanities)⁴¹, is acting as a credible source but is being misrepresented by those using it.

There are a few cases where Skills are used to disseminate longer-form history, usually in podcast form. The most notable in the search results is ‘Hardcore History’, a podcast by Dan Carlin, formerly a US radio show host with a Bachelor’s degree in history, which explores all manner of historical events in detail, debating ‘facts’, posing speculative ‘what if’ questions and taking an approach diametrically opposed to that of short, sharp ‘facts’. The Alexa Skill enables users to play his podcasts through Alexa, although they’re available on other platforms and smartphone apps too, and so the Skill is simply playing the already recorded podcast – it is Dan speaking rather than Alexa, which users appreciate, but Alexa is not designed as a podcast platform, which leads to technical difficulties:

The podcast is great. Fascinating discussions of history made topical. The problem is the usability. If you pause an episode and come back to it later Alexa will start that episode at the beginning.⁴² (‘Hardcore History’ Skill).

So, while Alexa can provide access to detailed, nuanced and entertaining historical content, it is not really designed to do so in this form, faring far better at reciting the dated events automatically plucked from Wikipedia, the history

⁴¹ Lih, ‘Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource’.

⁴² ‘Amazon.Com: Hardcore History: Alexa Skills’ <<https://www.amazon.com/dancarlin-com-Hardcore-History/dp/B01N8TOMI9?ref=chatbottle>> [accessed 2 October 2019].

with little context. In an extreme example, “This Day in History Top Story”, another History Channel Skill, presents a list of events for a given day without including dates, which prompted one reviewer to question whether information without the context of a proper date could even be counted as ‘history’:

Then there's a LIST of events from "today" with NO YEAR THEY HAPPENED. This seems basically POINTLESS - who remembers - WHO CARES? - the actual day "The Pony Express" (or whatever) started. The YEAR is what's interesting for historical context, and memorable.⁴³ ('This Day in History Top Story' Skill).

The next aspect of the content to be considered is its representativeness and breadth. The periods and locations of history covered across all the Skills tended to either be very specific, such as US naval history, or incredibly general; for example, one of the five Skills all called ‘History Facts’ simply recites a ‘fact’, with no source information, that could be an historical event from anytime, anywhere. The lack of source information attached to many such Skills adds to this sense of taking a ‘lucky dip’ into history, just as one does when clicking ‘random article’ on Wikipedia or ‘I’m feeling lucky’ on Google. The more specific Skills might focus on a period or culture, such as ‘Roman History Facts’, or, more commonly, some aspect of US history, by far the most popular topic area for Skills specifying a historical area of interest (see Figure 26). As Figure 26 shows, after the unspecified general Skills, those focussed on US history dominate the rest. When only looking at Skills which received 10 or more ratings, i.e. attracted a significant number of users (relatively speaking), only the US featured alongside the general Skills with no other regions represented at all. This US-centric nature of Alexa Skills, and indeed Amazon Alexa in general, is not surprising. Alexa started out as a US product and dominates the US market, expanding its market share year-on-year; by the end of 2018 Amazon Echo, the speaker that lets user talk to Alexa at home, was in more than 40 million US homes⁴⁴. Consider the fact that in the US the use of Alexa on

⁴³ ‘Amazon.Com: This Day in History Top Story: Alexa Skills’ <<https://www.amazon.com/This-Day-History-Top-Story/dp/B06XH9VWV5?ref=chatbottle>> [accessed 12 October 2019].

⁴⁴ ‘Google Home Added 600,000 More U.S. Users in 2018 Than Amazon Echo, But Amazon Echo Dot Is Still the Most Owned Smart Speaker’, *Voicebot.Ai*, 2019 <<https://voicebot.ai/2019/03/07/google-home-added-600000-more-u-s-users-in-2018-than->>

smartphones was more than double that via Echo speakers and the numbers become enormous, far greater than in any other part of the world⁴⁵.

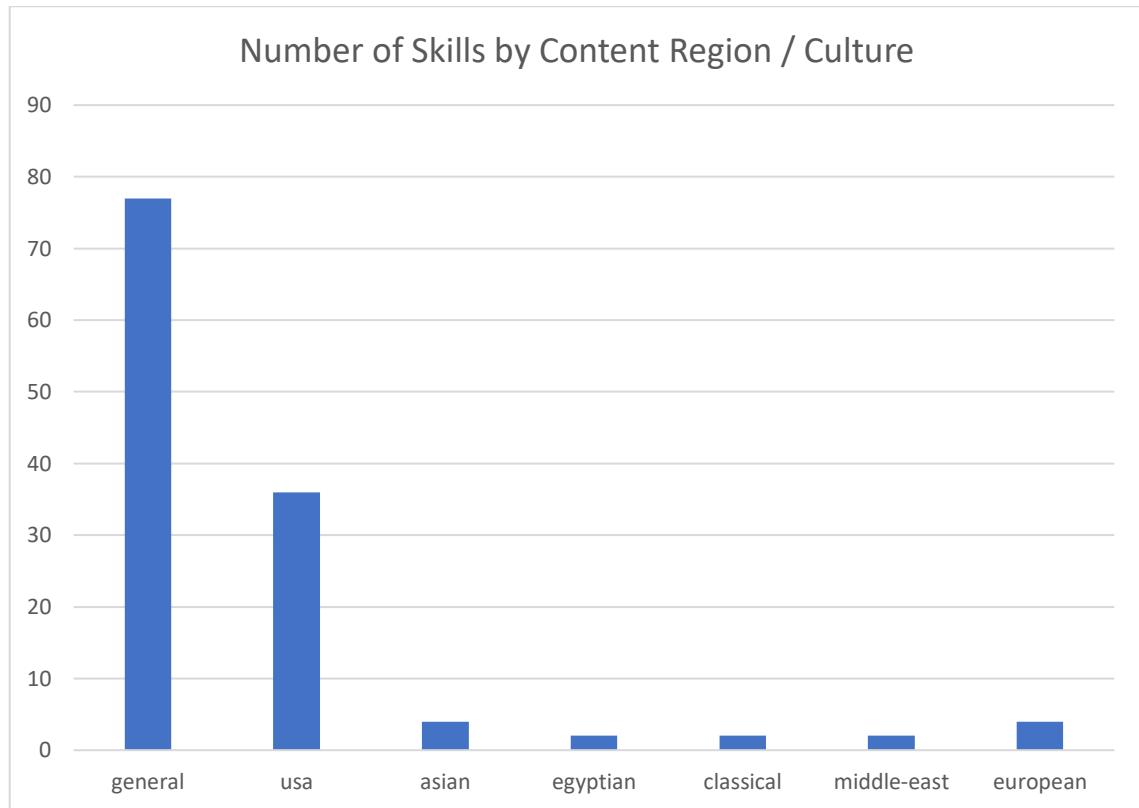


Figure 26: Bar chart showing the distribution of Skills' content per region and/or culture.

With the USA being Alexa's main market and user base, and presumably many of the developers of Alexa Skills therefore coming from, or living in, the USA, it is unsurprising that more than a quarter of all history Skills should focus on American history. For some users, this is exactly what they want:

I'm am going to continue this skill because of my love for American history and the content (I do receive) is factual and fairly indepth. I hope through future development the facts and biographies won't repeat as often⁴⁶ ('Colonial History' Skill).

[amazon-echo-but-amazon-echo-dot-is-still-the-most-owned-smart-speaker/](#) [accessed 25 October 2019].

⁴⁵ 'Twice the Number of U.S. Adults Have Tried In-Car Voice Assistants as Smart Speakers - Voicebot.Ai'.

⁴⁶ 'Amazon.Com: Colonial History: Alexa Skills' <https://www.amazon.com/dp/B01LX57ZYF/?ref=chatbottle> [accessed 2 October 2019].

However, some obviously feel the opposite, especially in cases where Skills which appear general in fact contain significant amounts of US content:

“Mostly USA based history. Disappointing”⁴⁷ (‘Today in History’ Skill).

This last point raises a number of issues that come to light upon close reading of the transcripts of interactions with each Skill. For general fact-style Skills which use Wikipedia as an automatic source of information for events, births or deaths on a given day, users may find that a Skill predominantly seems to focus on earlier time periods. This is the case with ‘World History’ which uses the Wikipedia article for a given date as its source and then runs through the events listed chronologically, relating three events every time it is asked. Unfortunately, this means that for most users, who may only ask once or twice per day, they will receive only those events that happened at the top of the chronologically ordered list, i.e. those earlier in history. This highlights the need for developers to carefully explore their source material (using a function to randomly pick three events from the page would seem a better idea) and how poorly considered use of automated methods and APIs to access content can lead to biased interactions with users even when the original material itself is actually quite representative and balanced.

In some cases, as seen in the last quote above, the unseen biases in knowledge or experience of the developers, whoever they may be in many cases, can come to the fore in the breadth of content available. At least five Skills pertaining to be ‘general’ in their coverage actually focus primarily on US history. The most striking examples are two Skills dedicated to women in history, ‘This Day in Women’s History’ and ‘Historical Women’, for which the interactions recorded included only historical facts about women in America. This may reflect an underlying authorial bias from the developer’s perspective, or an inherent bias in the user’s view of history (in Britain the focus would be on the Suffragettes, but is their non-appearance an oversight or actually an exposure of my lack of wider knowledge?). This is why detailed descriptions of

⁴⁷ ‘Amazon.Com: Today in History: Alexa Skills’
<<https://www.amazon.com/dp/B01EBCW532/?ref=chatbottle>> [accessed 8 October 2019].

Skills, transparency of sources and developer expertise is vital if apps like these are going to become more useful, popular and trustworthy.

Aside from the content itself, every Skill on the list was only available in English, despite some of them being aimed at Middle-Eastern or Chinese history. Even amongst these Skills some were only available in the US, a hangover from the introduction of Alexa and Echo speakers which was strictly US-only. Alexa is now available in 89 countries, but only synthesises its speech in three languages with any competence: English (with changeable UK and US accents), German and Japanese; and the Alexa smartphone app (for those who can't afford or don't want a smart speaker) is not available in many of these 89 countries anyway⁴⁸. The geography of the Internet and the information on the Web is still a predominantly American, let alone Western, one, and the vast information inequalities that exist on mainstream platforms like Google and Facebook are only amplified when applied to new and growing technologies like Alexa⁴⁹. From a gender perspective, one Skill specifically focussed on 'This Day in Women's History'⁵⁰, but the vast majority of Skills about individuals, often the Founding Fathers of the USA, were focussed on men and their historical activities.

The 'Black History Facts' Skill, originally launched in 2016 in the wake of Black History Month and the growing black History movement, explicitly seeks to balance out such historical information equalities that exist within the USA (ironically, the only African country, at present, which may benefit from this Skill via Alexa is South Africa). Created by developers who worked at Amazon, this Skill is the second most popular and highly used of all those returned by the 'history' search term, having received 1867 ratings. Considering that 64 of the 127 Skills received no ratings at all, and that the average rating across all

⁴⁸ 'How to Use Amazon Alexa in Non-Supported Countries', *The Ambient*, 2019 <<https://www.the-ambient.com/how-to/use-alexa-non-supported-countries-855>> [accessed 25 October 2019].

⁴⁹ Andrea Ballatore, Mark Graham, and Shilad Sen, 'Digital Hegemonies: The Localness of Search Engine Results', *Annals of the American Association of Geographers*, 107.5 (2017), pp. 1194–1215, doi:10.1080/24694452.2017.1308240.

⁵⁰ 'This Day in Women's History: Amazon.Co.Uk: Alexa Skills' <https://www.amazon.co.uk/This-Day-in-Womens-History/dp/B07CXFX2N7/ref=sr_1_1?keywords=women+in+history&qid=1572770243&s=digital-skills&sr=1-1> [accessed 3 November 2019].

results was only 64, the relative reach of this Skill is extremely high. The user reviews of this Skill applauded and emphasised the importance of highlighting lesser known historical and cultural information in an easily accessible way; one of the main factors in achieving this was human, culturally relevant narration rather than the ubiquity of Alexa's synthesised voice.

Such a cool surprise that the facts were read by people, and not the voice of Alexa. It feels more personal. I loved that! ⁵¹ ('Black History Facts' Skill).

This is great, for so long African American history was hidden, distorted and not being taught, if people did right by african americans the first time around we wouldn't need apps like this. But they didn't so here we are⁵² ('Black History Facts' Skill).

Examining the effectiveness of such interactions with all of the history Skills, as well as with Alexa itself, was a challenging task as 64 of the 127 Skills did not even receive ratings, let alone reviews. As described in the methodology section, numbers of user ratings were recorded for each Skill to give an impression of the Skill's overall usage – this is not the same as popularity, as these ratings may be positive, negative or indifferent, and even then, these sentiments may refer as much to the Alexa platform as to the Skill itself or its content. However, the most telling statistic from the numbers of ratings collected (n=8160) is that three Skills – Ultimate HISTORY Quiz (n=2804), Black History Facts (n=1867) and This Day in History (n=1402) – accounted for 74% of all ratings across the 63 Skills that received at least 1 rating from a user. As noted above, half the Skills did not receive any ratings at all. Ultimate HISTORY Quiz and This Day in History are both affiliated with the History Channel, the website of which has a 'This Day in History' page⁵³, an organisation with the resources, publicity and audience reach to attract users to these Alexa Skills. The fact that Ultimate HISTORY Quiz is only available to users in the USA also highlights the main target market for these Skills.

⁵¹ 'Amazon.Com: Black History Facts: Alexa Skills' <<https://www.amazon.com/dp/B01MYC6MJG/?ref=chatbottle>> [accessed 2 October 2019].

⁵² ibid.

⁵³ 'This Day in History', HISTORY <<https://www.history.com/this-day-in-history>> [accessed 27 October 2019].

The vast discrepancy in numbers of users engaging with these top three Skills compared to all the rest (the next highest number of ratings is 468), highlights the importance of such social bots having trusted and identifiable developers and information sources, relevant, varied and oft-updated content, and being 'findable'. Especially on Amazon, which makes extensive use of recommendation algorithms across all of its online marketplaces, being 'findable' means being easily searchable and recommendable. Some of the Skills returned had the same name, for example 'History Facts' or 'Today in History', so the user is unlikely to enable them all, instead probably choosing to use the most popular, or highest up in the search list (often the same thing). The recommendation algorithm also favours popular Skills within the broad category tags available on the Alexa store, such as 'education'. This means that Skills with small numbers of users are not likely to be recommended, especially as the pool of Skills within such a broad category will be huge. For example, after using the Skill 'This Day in History', listed in the 'education' category, Alexa's recommender system suggested that I might like "another popular education Skill: 'Animal Sounds"'; even this was designed only to play animal noises as commonly recognised by English speakers and was distinctly lacking in historical information about said vocal expressions.

It is important to explore the algorithmic influence of the Skills' wider setting, the selling infrastructure of Amazon, to assess audience reach as this is key when considering the popularity and likelihood of finding more relevant history Skills. As discussed in Chapter 2, recommender systems like the one used by Amazon across its site usually fall into one of two categories: content-based and collaborative. Collaborative approaches focus on what other users have engaged with when recommending content, and although the details of proprietary algorithms used by companies are not fully known, it is apparent that Amazon takes this broad approach to recommendation of Skills by listing those that 'Other customers have also enabled'. A recommendation network for history Skills was constructed and analysed as described in the methodology, with 58 out of 127 Skills (46%) recommending another history Skill and therefore being present in the network. Measures of indegree (the number of recommendations a history Skill received from other history Skills) outdegree

(the number of recommendations given by a history Skill to other history Skills) and stress (the overall influence of a history Skill in recommending others in the network) were used to quantify patterns of content suggestion across the network (Figure 27 shows the indegree distribution and Figure 28 the outdegree distribution for the network).

When analysing the network it is important not to equate high numbers of recommendations to and from Skills as a measure of *overall* popularity; this might be a good measure of which people who have used just the history-based Alexa Skills are most likely to click on, but ones like Black History Facts with much greater overall popularity have a far wider audience and therefore will be recommending other, less relevant Skills too. This is because Amazon's use of a collaborative approach, especially when limited to suggesting ten Skills at a time, will favour other general Skills that are more popular than any related history ones. Within the history Skills, those rich in American content are also in the majority with six of the ten most influential Skills in the recommendation network focussing on US history (see Table 9). Therefore, the connecting power of the recommendations is amplified for the US market (like everything about these Skills) leading to a self-perpetuating focus on American historical content.

Looking at the user ratings associated with each Skill, it becomes clear that the strong core of the recommendation network here is between Skills that have barely any (or no) ratings; it could be that this lack of popularity makes them more likely to be recommended to each other within the same category, especially as users are only going to interact with these Skills if they have specifically searched for 'history' and scrolled through the results list past the most popular ones at the top. From a researcher's perspective, such recommendation chains between apparently unpopular Skills could be incredibly useful when trying to find related history Skills, given their overall paucity. Yet, this only worked when done on Amazon's website from a computer, when using Alexa as intended through voice interaction recommendations always occurred by category, such as 'education', so these recommendation chains would be missed.

The point about recommendations only being within the collected dataset of history Skills is best shown by ‘Ultimate HISTORY Quiz’, the History Channel-developed Skill which was by far one of the most popular. It has an indegree of 7, showing it has the equal highest tendency to be recommended by other history Skills; the fact that it is only available to users in the USA means that its audience for recommendation is slightly more limited too. However, its stress centrality measure is 0 because it does not recommend any other history Skills in this network; this is due to the point above that the highly popular Skills in this dataset will often be recommending other highly popular Skills in the same broad categories, such as ‘education’, or very popular Skills more generally. ‘Black History Facts’ (in = 1, out = 0, stress = 0) and ‘This Day in History’ (in = 3, out = 1, stress = 1) are the other two very popular Skills in history set and demonstrate the same problem to a greater extent; other popular skills are completely absent from the recommendation network and play no part in suggesting other history Skills to users (see Table 10). As such, in this context, having a high indegree and low stress is actually indicative of wider popularity, but unlike the social bots network this same popularity prevents them from acting as ‘gateway bots’ into the wider collection of history Skills because Amazon’s recommender system, when focussing on which Skills other customers have enabled, primarily rewards high levels of popularity as the most popular Skills will have been used by the highest numbers of people⁵⁴.

⁵⁴ Brent Smith and Greg Linden, ‘Two Decades of Recommender Systems at Amazon.Com’, *IEEE Internet Computing*, 21.3 (2017), pp. 12–18, doi:10.1109/MIC.2017.72.

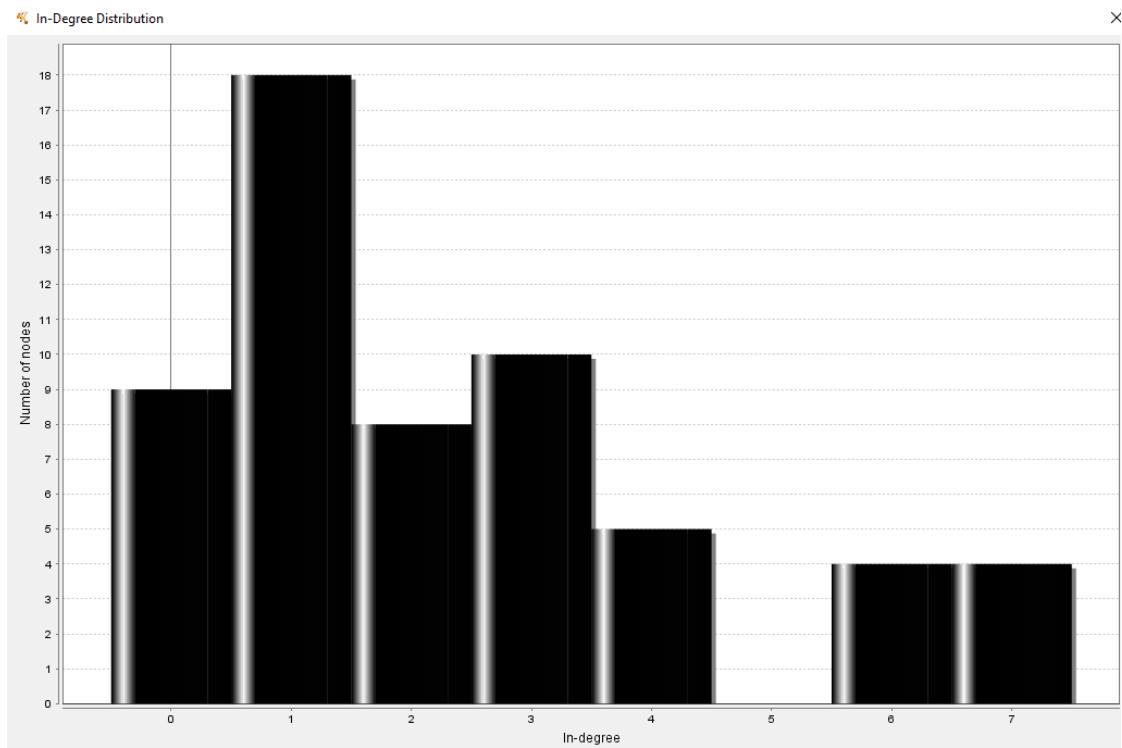


Figure 27: Graph showing in-degree distribution of Skills in recommendation network.

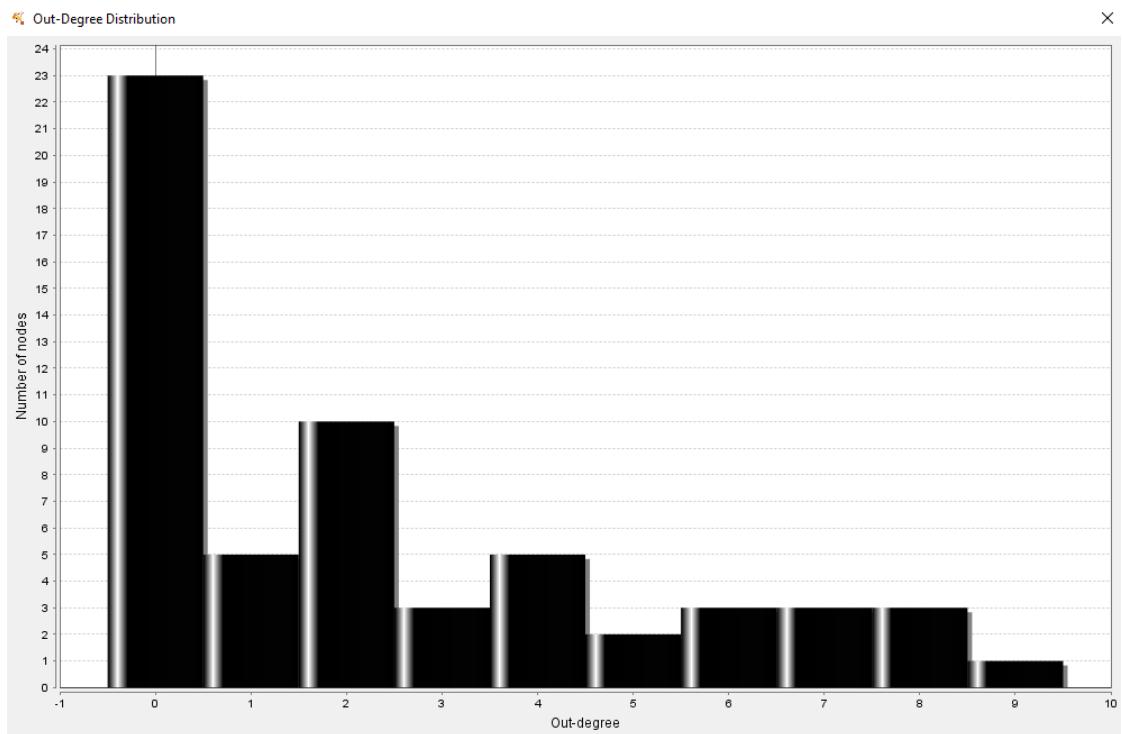


Figure 28: Graph showing out-degree distribution of Skills in recommendation network.

Skill	Edgecount	Indegree	Outdegree	Stress	User Ratings
History Timeline Facts	13	7	6	319	1
History Wizard	10	3	7	259	3
Odd History	11	4	7	222	0
United States History Facts	10	6	4	184	0
Game of American History	9	4	5	130	1
history buff	8	1	7	119	1
American History	9	3	6	114	0
History Buff	10	2	8	104	1
US History Facts	12	7	5	85	0
Trivia Challenge: U.S History	5	3	2	81	32

Table 10: The recommendation network statistics for top ten Skills ranked by stress centrality.

Skill	Edgecount	Indegree	Outdegree	Stress	User Ratings
Ultimate HISTORY Quiz	7	7	0	0	2804
Black History Facts	1	1	0	0	1867
This Day in History	4	3	1	1	1402
Today in History	3	3	0	0	468
Today in History	0	0	0	0	333
US History Quiz	1	0	1	0	179
Livy	0	0	0	0	132
Today in History	0	0	0	0	126
This Day in History Top Story	0	0	0	0	123
History Trivia	4	0	4	0	108

Table 11: The recommendation network statistics for top ten Skills ranked by number of user ratings (popularity measure). N.B. the four Skills with zero in each column play no part in the network, showing that popular Skills often don't recommend each other.

Overall, this means that users are unlikely to be recommended history-related Alexa Skills from those that are already popular; instead they will have to actively search for 'history' or know a certain Skill's name in the first place (probably one of the reasons the History Channel's Skills attract a relatively high number of users). However, such popular Skills are also more likely to appear near the top of searches for 'history' Skills, exacerbating the problem. As the network analysis shows, if users do find a less popular history Skill from a search then they are far more likely to be recommended a relevant, but probably equally unpopular, history Skill from that less popular Skill's page. This

creates the usual kinds of content bubbles where users can find highly popular content from other highly popular content, while getting a far more diverse range of content requires a deeper level of interest and perhaps prior knowledge of the content available. Such a system of recommendation will not encourage a wider audience engagement with a greater range of historical content beyond those already interested users who are knowledgeable, and persistent, enough to find it in the first place. On top of this, the analysis conducted in this study was run without being logged in to an Amazon account, therefore personalisation of recommendations did not occur. In the real-world a user's previous Amazon browsing history, which may for example be focussed on trivia Skills, would further skew the suggestions made available. It should also be remembered that 54% of all history Skills collected did not recommend and were not recommended by any other history Skill, highlighting the overall lack of discoverability of these Skills.

Amazon's collaborative approach to content recommendation creates the kind of 'filter bubbles' common in this area; comparing this with the Off The Easel botnet approach to recommendation which aimed to combat, to some extent, this tendency for popular to recommend popular and unpopular to recommend unpopular (in that case on social media but also more generally), it's clear that something similar is needed here to make these Skills more visible to users and diversify the content they can find beyond that produced by the History Channel.

4.3 Discussion: Early Stages for Conversational Engagement with History and Cultural Heritage

Although chatbots are currently booming in commercial and industry settings, the survey carried out shows that their uptake in academic, educational and GLAM sectors is still very low. In the museums sector, where experimentation with chatbots is most apparent, the influence of the wider commercial developments is clear to see with the majority of bots in this area of the customer-service or tourist-information variety, often little more than website menus compressed into a Facebook Messenger chatbot. The odd exception to

the rule often stems from an academic background and an urge for users to engage with a museum's content, preferably in novel, meaningful ways, rather than simply with the institution itself. However, as described in the introduction, such work coming from academia is extremely limited (if growing) and may end up developing chatbots that play different, perhaps complementary roles to those more functional digital customer assistants common in other commercial sectors. As Tzouganatou⁵⁵, Perry⁵⁶ and Graham⁵⁷ note, chatbots should ideally 'chat', they should have something to say and they should make the user think before wanting to say something meaningful in response. As will be discussed in the next chapter, this vision is unlikely to be fully realised just yet, or at least not solely by a computer program. Chatbots are certainly capable of disseminating historical information, but they have no knowledge or understanding of history; the translation from information to knowledge occurs in the mind of the user, which requires engagement on their part, and as ELIZA shows such engagement need not rely on staggering technological ability; sometimes a change in approach works just as well.

This is borne out by the findings from examining the role of Alexa Skills in disseminating historical content to users and thus engaging them with that content. Compared to the social bots discussed in the previous chapter, the Alexa Skills available for historical content require a level of proactivity on behalf of the user to find them in the first place. The quiz-style Skills also promote interaction and were, in my experience, the most engaging to use. Of course, the general fact-style Skills were still passive, the inevitable monotony of Alexa's synthesised voice decreasing their chances of being remembered, and for users unaware of Skills simply asking Alexa what happened today or about a given event leads to recitals of Wikipedia article summaries, one of the least engaging modes of knowledge consumption imaginable.

This is a marked difference from typical ways of finding historical information, which is most commonly a user actively searching Google or Wikipedia for

⁵⁵ Roussou and others, 'Transformation through Provocation?'

⁵⁶ Sara Perry, 'The Enchantment of the Archaeological Record', *European Journal of Archaeology*, 22.03 (2019), pp. 1–18.

⁵⁷ Graham, 'An Enchantment of Digital Archaeology'.

keywords or a question, or maybe even flicking through a book. Short soundbites, facts for the day, images on a social media feed and a general lack of wider context, historical and / or cultural, is not just a problem for automated means of dissemination - human social media accounts often spread content in the same way (and many with less good will) - but it is an issue which automation naturally amplifies. Both Graham⁵⁸ and Perry⁵⁹ speak of the potential for digital technologies and approaches to 'enchant' researchers and users through greater creative involvement and engagement, yet in the majority of cases here Alexa Skills were simply used as a new medium to disseminate the same old information in the usual ways with the 'names and dates' style of school history dominating. The human-narrated, carefully crafted exceptions, such as Black History Facts, did resonate more strongly with users, but these largely eschewed automation and algorithmic approaches. As discussed below, the constraints put upon developers by the requirements of automation are major factors in the lack of enchantment found in most of the available Alexa Skills.

With the rise of ChatGPT and similar generative AI tools in 2023 bringing greater conversational capabilities to the forefront of developers' minds, Amazon has recently announced a significant update to Alexa built upon a new Alexa LLM (large language model), effectively Amazon's answer to ChatGPT⁶⁰. Although not available yet, this offers developers new ways to create Skills to engage users as these can now make use of the new Alexa LLM's generative AI capabilities; indeed, a new 'Character.ai' Skill will come included in the Alexa update and will feature 25 different characters including historical figures like Einstein and Socrates⁶¹. However, if the new Alexa LLM's capabilities to generalise conversation are akin to those of ChatGPT, this also raises the question of whether specific Skills would be required anymore. As highlighted by the Alexa Skill survey, finding and enabling Skills was time-consuming and difficult, especially when Skills had similar names. Unless a new approach will be taken where the AI model automatically enables the most relevant Skill to

⁵⁸ Graham, 'An Enchantment of Digital Archaeology'.

⁵⁹ Perry, 'The Enchantment of the Archaeological Record'.

⁶⁰ Knight, 'Amazon Upgrades Alexa for the ChatGPT Era'.

⁶¹ *ibid.*

perform a task, or could itself develop such a Skill that can then be used (akin to the new GPTs made via ChatGPT), it's unclear how many users would need or want anything beyond standard conversational interaction with Alexa for history questions or information once this is augmented with full generative AI capabilities that should provide a more engaging experience.

The platforms involved, whether Amazon Alexa or Facebook Messenger, play wider roles in negotiating these user interactions and their influence is pervasive and difficult to measure, both from a research and public perspective. The analysis of how history Skills recommend each other, or do not in most cases, reveals the same patterns created by recommender systems across social media and other e-commerce sites with 'filter bubbles' of highly popular content coming to the fore whilst relatively unpopular content can recommend itself but first has to be found and it is very difficult for users to stumble across it.⁶² Users are very unlikely to find relevant history Skills without actively searching for them, and even then Skills with very similar names, defunct and defective Skills still muddy these waters. Researchers are users too and in this setting where algorithms influence every interaction, it is impossible for researchers' actions not to change the data; I cannot help but wonder what the recommendation network for history Skills would look like if I had not come along and enabled every single history Skill I could find.

Platforms hold the power when it comes to determining how and when automated activity can occur on them, along with what sort of historical content is likely to be shared by bots using their services. The geographical inequalities inherent in Amazon Alexa have already been discussed, and whilst they are not as stark for some social media platforms, a Western-centric perspective still dominates both in terms of who creates bots and the content which is shared through them. The English version of Wikipedia is the information source for many of the Skills, yet it is known that different language versions of Wikipedia can have very different narratives of historical events⁶³, which in itself is yet

⁶² Pearl Pu, Li Chen, and Rong Hu, 'Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art', *User Modeling and User-Adapted Interaction*, 22.4–5 (2012), pp. 317–55, doi:10.1007/s11257-011-9115-7.

⁶³ Gieck and others, 'Cultural Differences in the Understanding of History on Wikipedia'.

another example of how bias inherent in the ecosystem of the Web, especially that which is automatable, seeps into content shared in many different places.

Much like the social bots case study discussed in the previous chapter, the chatbot survey and Alexa Skills case study have unavoidable limitations in terms of data collection, where finding and confirming the identity of social bots in particular is a challenge, and the information available about successfully identified bots is limited. For Alexa Skills, a lack of detailed information provided by bot developers and managers made data collection and subsequent analysis harder and more subjective, with assumptions about activity or mechanism being necessarily more common than would be ideal. Again, the influence of platforms could not be avoided, with Amazon making it as difficult as possible to collect Alexa transcripts and Facebook regularly banning me for ‘harassing’ bots on Facebook Messenger (apparently holding a conversation with a bot for more than five minutes seems to count as harassment). Automation is still viewed negatively by such powerful players, despite the ironically pervasive roles their algorithms play in policing and shaping how those same interactions occur. Although API access did not apply to any of the work carried out in this chapter, the same problems highlighted by Axel Bruns⁶⁴ for researchers studying, analysing or using data from and with automated methods online still apply, and often it is easier (and more efficient) to eschew automation in favour of a manual approach to data collection.

The theme of human input, compared to the perception of bots disseminating content in a hands-off, fully automated manner, is perhaps more apparent with chatbots than it was for social bots. Developers of Alexa Skills have to manually add content to their apps, such as new quiz questions, to keep users engaged and prevent the bot from becoming too repetitive. From the perspective of historical content, the human influence is less obvious but far more important. In terms of development and content curation, non-expert individuals dominate this area rather than official institutions or academics. Most Alexa Skills developers seem to be based in America, hence the propensity for US history

⁶⁴ Axel Bruns, ‘After the “APIcalypse”: Social Media Platforms and Their Fight against Critical Scholarly Research’, *Information, Communication & Society*, 22.11 (2019), pp. 1544–66, doi:10.1080/1369118X.2019.1637447.

or a Western perspective on World history. Automation can be a useful tool for helping to promote lesser-known histories and cultural heritages, as seen in the ‘Black History Facts’ Skill and the ‘Black History Month’ movement’s digital presence as a whole, but the wider structures and system within which it operates can also reduce the representativeness of that same audience. The perception that automation leads to some sort of implicit objectivity, as the human is abstracted away behind the bot, still lingers despite all the evidence to the contrary. This applies as much to the consumption of content as to the content itself. Access to APIs, reliable algorithms for natural language processing and the requirement to take content from primarily one source, rather than assimilate it from several, means that automation often presents content distinctly lacking in context. In this case automation enables a different type of interaction, but in other cases, such as many of the Alexa ‘on this day’ skills, this context-less consumption is forced upon the user.

The technical shortcomings of AI and the constraints imposed by automation are mostly clearly seen in chatbots. As Tzouganatou⁶⁵ notes with regard to museum chatbots on Facebook Messenger, technical obstacles in natural language processing (the ability of the bot’s algorithm to understand and cogently reply to users’ typed queries) and a desire by developers to constrain available choices for the user to reduce complexity, mean that such interactions are often passive, one-way affairs where the bot informs the user to the best of its ability and, if unable to provide an answer, simply does not try⁶⁶. Throw in the added difficulty of processing verbal instructions rather than typed ones, as is required for VCAs like Alexa, and complexity of designing a bot or Skill capable of answering more open-ended user questions rises exponentially. The technical aspects also tend to narrow the potential sources of content. Wikipedia and similar online, collaborative, community-based sites are the dominant information sources for automated apps and accounts and those who develop and manage them. Wikipedia was the most common automated information source for Alexa Skills, and many of those manually created still included content from Wikipedia. Disseminating diverse interpretations,

⁶⁵ Tzouganatou, ‘Can Heritage Bots Thrive?’.

⁶⁶ *ibid.*

arguments and perspectives requires a larger number of trustworthy, easily automatable information sources to be available, but it also needs those with expert knowledge to be involved in the process of developing bots in the first place.

Some of these technical challenges may be overcome, or at least be altered, with the advent of new algorithms and models. The Transformer algorithms explored in the following chapter have shown impressive abilities to produce convincing articles on a huge variety of topics, and much work is underway to try and make these powerful new models the basis for future chatbots. For example, it has already been demonstrated that the most recent iteration of these algorithms, GPT3, can produce quiz questions related to American history simply by being prompted: “Generate 10 questions about early American history”, and being given an example: “Question 1. When was the Declaration of Independence signed?”⁶⁷ From this the algorithm goes on to generate nine more questions that would not be out of place in any of the quiz-style Alexa Skills, before being prompted with all ten of the questions which it proceeds to successfully answer. Such technology would in theory make it possible for questions and answers to be generated on-demand, removing the repetition of manually curated Alexa Skill quizzes which users often complained about. Of course, such use of GPT3 may simply make such Skills redundant, but currently the vast costs of running the algorithm make this unlikely in the near future (see Chapter 5).

In summary, this chapter has shown that the significant potential for chatbots to disseminate historical content online and engage users and visitors in heritage settings is only just starting to be fully explored. The growth of platforms like Amazon Alexa is only likely to see this continue, but the inevitable bias is toward content that is easily accessible, automatable and within the cultural mindset and knowledge of developers, in this case leading to a focus on content about American history. Similarly, the prominence of English Wikipedia as a source of information leads to similar styles of short, sharp bursts of content that rarely carry any context and yet are still the product of much unconscious

⁶⁷ ‘GPT3 Hunt | Prompts, Examples, & Demo’s for OpenAI’s GPT-3 API.’
<<http://www.buildgpt3.com/>> [accessed 28 December 2020].

cultural bias, despite the relative trustworthiness of Wikipedia as a data source (especially when compared to the mass of unknown, individual developers of Alexa Skills). The ability for users to find such Skills in the first place, and their role within the vast algorithmic network of recommendations on Amazon and interactions with Alexa, reinforces the point that using chatbots is about more than the content they share; chatbots need to engage users and the platforms in which they are found can play a major role in this process. The future for chatbots in this area is one full of potential, especially as powerful and potentially revolutionary new algorithms become available for use.

5. Case study: Is the Novel AI Tool ChatGPT the Future of the Past?

The digital landscape of history and heritage dissemination has been gradually transformed by the advent of automation and algorithms, as highlighted in previous chapters. This shift has begun to shape the nature of public engagement with historical content, altering the methods of content curation and distribution and the manner in which users interact with these digital artefacts. Chapter 3 showcased the vast, dynamic realm of social media bots, focusing on the automated and semi-automated distribution of cultural and historical heritage content online. It shed light on the importance of algorithmic curation and the significant role of bot developers, online communities, and end-users in shaping the content that gets disseminated, underlining the inherent subjectivity of big data methodologies.

Chapter 4 shifted the focus onto chatbots, particularly within the Amazon Alexa platform, demonstrating their potential for historical content distribution and engagement in heritage settings. The chapter also highlighted the biases introduced by the available content and cultural knowledge of developers, thereby influencing the type and style of content shared through chatbots. This chapter underscored the importance of user engagement and the influential role of the platforms themselves in shaping user interactions with chatbots.

Building upon these foundations, this chapter delves deeper into the ever-expanding realm of generative AI, with a particular emphasis on the capabilities of the ChatGPT tool which has been making headlines recently¹. In the three years since the previous chapters were written, there has been rapid development of various algorithms that can generate information - whether in the form of text, images, video, audio, code, data and more – along with an increase in the amount of training data that such algorithms use to learn and hone these techniques. The most apparent and publicly lauded changes have come in the abilities of algorithmic language models to produce text and images, with tools like OpenAI's ChatGPT achieving results so spectacular that

¹ OpenAI, 'Introducing ChatGPT', 2022 <<https://openai.com/blog/chatgpt>> [accessed 20 July 2023].

they have justifiably been described as true Artificial Intelligence². Although it will not be discussed in detail in the scope of this chapter, the contribution of generative AI tools like OpenAI's DALL-E 2 also needs to be acknowledged when discussing AI in this context³. Given the discussions around the definitions of 'AI' in Chapter 2, and the inherent ambiguity of the term, it is important to note that this marks the first point in this thesis where a model being studied is not merely a hard-coded set of algorithms, or automation of data processing and sharing, as has been the case so far, but is genuinely flexible, adaptable and personable enough in its interactions to be considered AI in any of its various forms. More pertinently, the growing public, media and policy discourse around 'AI' now largely refers to ChatGPT and its ilk, forming a pragmatic definition by consensus that is likely to have far more influence than anything agreed upon academically.

With this context in mind, it is now more important than ever to explore how these tools, and the ecosystem that they inhabit, can interact with and influence information about the past. This matters not only to researchers and GLAM professionals, but also to teachers, policy-makers and anyone with an interest in history and archaeology. The advent of such technology presents a pivotal moment in the evolution of content creation and dissemination, expanding the horizons of automation and AI into the creation of novel, contextually relevant content. This not only amplifies the potential for research and public engagement but also introduces a new layer of complexity to the sociotechnical ecosystem that simply did not exist when considering Twitter bots or Alexa Skills in isolation. The GPT-4 'AI' model has the potential to fundamentally change the online information ecosystem in ways similar to that seen in the past by the introduction of Wikipedia and social media, as discussed in Chapter 2, hence this chapter will focus as much on the role of the frameworks in which these tools operate as much as the individual capabilities of the tools themselves.

² Sébastien Bubeck and others, 'Sparks of Artificial General Intelligence: Early Experiments with GPT-4' (arXiv, 2023), doi:10.48550/arXiv.2303.12712.

³ OpenAI, 'DALL·E 2', 2022 <<https://openai.com/dall-e-2>> [accessed 20 July 2023].

In order for these discussions to make sense within the context of interactions with information about the past, a deeper delve into the background of these new AI models and their wider ecosystem is required. By focussing on the potential influence and application of AI models like ChatGPT and DALL-E on information about the past especially in research, education and GLAM settings, this chapter has the following key aims:

- Assessing the current capabilities of AI models in understanding and generating content about history and cultural heritage
- Exploring potential applications of AI models in research, education and GLAM settings
- Examining the wider sociotechnical ecosystems of AI models, especially ChatGPT, and how these influence potential information biases
- Exploring the ethical implications of using such models, with a focus on information biases, and the potential future opportunities and challenges

Understanding their technical capabilities and limitations, including the potential for 'hallucinations' (where the details of content are wrong and/or made up), is crucial for ensuring accuracy in different contexts. This is vital when exploring their potential uses in academia, education, and cultural institutions, as is understanding the wider and deeper sources of potential bias and information inequalities. Therefore, the rest of this chapter examines both the importance of the broader sociotechnical ecosystems and the specific case study of ChatGPT, including platform roles and plugin integration.

5.1 Exploring ChatGPT and its Plugins in Context

The evolution of AI and machine learning has been marked by significant milestones, with the development of GPT-4 showcasing the rapid advances in this field, along with the questions of how to manage these. GPT, an acronym for 'Generative Pre-trained Transformer', refers to a type of model used in machine learning, particularly for tasks involving the understanding and generation of human language. The transformer architecture, upon which GPT is based, recognises patterns in language by examining the relationship of

words across entire sentences, much like how humans understand context in language⁴. The transformer model's unique feature, known as 'attention', allows it to focus on different parts of the input when generating each part of the output. This mechanism enhances the model's ability to understand the context of words in a sentence and process long sentences more effectively than previous models⁵.

The development of GPT-4 has been hailed as a significant moment in the field of artificial intelligence. Its vast size and seemingly impressive capabilities demonstrate the potential of transformer-based models for complex language tasks. However, as with its predecessors, this next iteration of the Transformer architecture again raises important questions about the ethical use of such technology. OpenAI has implemented several safety measures in GPT-4 to prevent misuse, including the use of reinforcement learning from human feedback (RLHF) to reduce harmful and untruthful outputs, and a Moderation API to warn or block certain types of unsafe content⁶. The effectiveness of these approaches remains to be seen, and the potential biases will be assessed in this chapter. In the context of this thesis, GPT-4's ability to generate novel, contextually relevant content introduces a new layer of complexity when considering how different users might interact with tools like ChatGPT in the fields of history, archaeology, and GLAMs (Galleries, Libraries, Archives, and Museums).

Within these areas, the different user groups including developers, implementers and users will be considered when exploring the potential uses of ChatGPT. For example, developers might include researchers, implementers might be GLAM professionals, and users might be visitors, but the flexibility of ChatGPT means that this will likely differ with each use case. Current work shows that GPT-4, both as a model by itself and as part of the ChatGPT online

⁴ Peter Clark, Oyvind Tafjord, and Kyle Richardson, 'Transformers as Soft Reasoners over Language', *arXiv:2002.05867 [Cs]*, 2020 <<http://arxiv.org/abs/2002.05867>> [accessed 16 October 2020].

⁵ Vaswani and others, 'Attention Is All You Need'.

⁶ Christiano and others, 'Deep Reinforcement Learning from Human Preferences', xxx.

tool, is already being tested in GLAM settings⁷, being used to aid teaching⁸ and being explored in areas of historical research⁹. In each case, the flexibility and adaptability of GPT-4 offers opportunities and presents challenges, the full extent of which are yet to be fully explored; indeed, the sheer scale of such language models makes exploring every possibility not only impossible, but meaningless. This chapter aims to extend these explorations by considering a recent addition to the ChatGPT framework – plugins.

Just as Alexa Skills are add-ons to Amazon Alexa that give it access to specific information or extra functionality, the idea of ChatGPT plugins is in a similar vein. Given the inherent limitations in the training of GPT-4, i.e. it only ‘knows about’ information before September 2021, any more recent information or specific knowledge that is not included in its training data, such as museum collection databases accessible via APIs but not standard web search, is missing. Plugins are one way that these information gaps might be filled, along with broader functionalities like browsing the web for information and writing and executing computer code.

As plugins are still an experimental feature of ChatGPT, and have only been available to subscription users for a couple of months, this is a very new area that is likely to change rapidly over the coming months. However, testing the general principles of plugins and how interactions with them change the ChatGPT experience for different user groups is key to understanding the potential functionality and future utility of the whole platform.

When a user interacts with ChatGPT, they might ask a question or make a request that requires the use of a plugin. The model will then use the appropriate plugin to fetch the required information or perform the requested task, and incorporate this into its response. For instance, if a user asks ChatGPT to find information about a museum exhibition, the model might use the Web Requests plugin to search the internet for relevant information. Or if a

⁷ Trichopoulos and others, ‘Crafting a Museum Guide Using GPT4’.

⁸ OpenAI, ‘Teaching with AI’, 2023 <<https://openai.com/blog/teaching-with-ai>> [accessed 1 September 2023].

⁹ Didier El Baz, ‘Can We Trust Chatbots for Now? Accuracy, Reproducibility, Traceability; a Case Study on Leonardo Da Vinci’s Contribution to Astronomy’ (arXiv, 2023), doi:10.48550/arXiv.2304.11852.

user asks for information from an academic paper, the model might use the ScholarAI plugin to access the paper's abstract or full text.

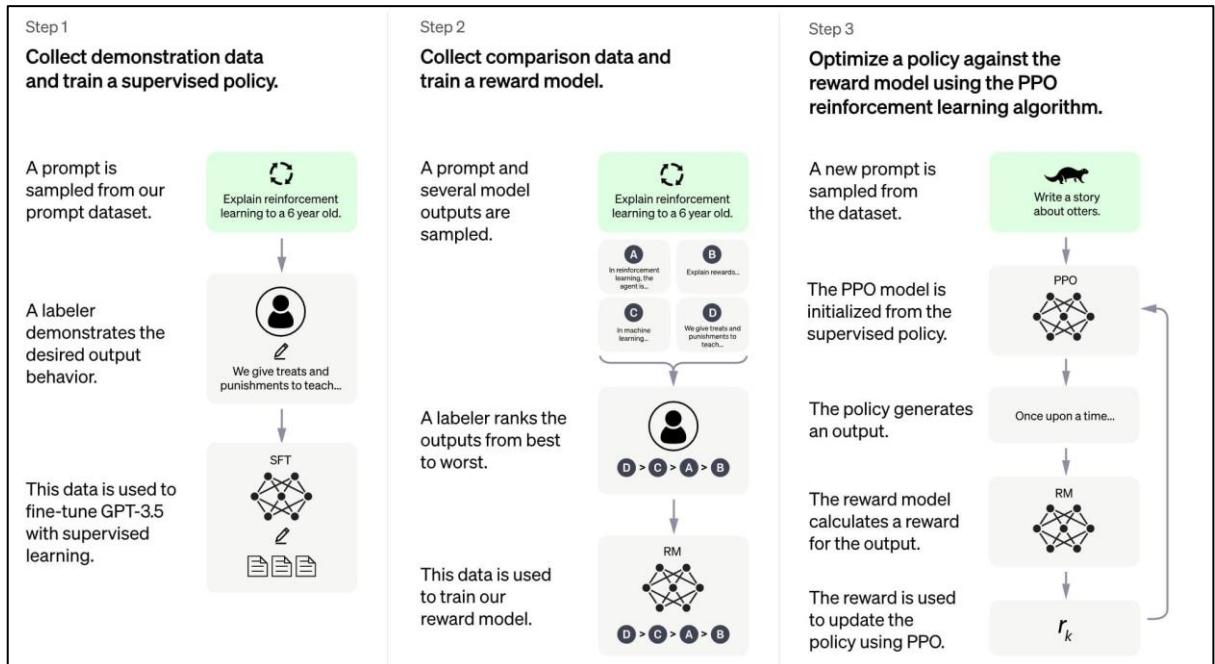


Figure 29: Flowcharts showing the stages of Human Reinforcement Feedback Learning (HRFL) involved in training GPT-4 (attribution: OpenAI¹⁰; licence: CC BY 4.0 DEED).

Software as a Service (SaaS) and Platform as a Service (PaaS) are two key frameworks in the realm of cloud computing that have significant relevance to the deployment and usage of AI models in various settings, including academic, research, and GLAM (galleries, libraries, archives, and museums) institutions. SaaS refers to a cloud computing model where software applications are provided over the Internet on a subscription basis¹¹. SaaS works on the Web, where, users do not need to install or maintain the software on their own computers; instead, they can access the software and their data online from any device with an Internet connection. PaaS, on the other hand, provides a platform and environment that allows developers to build applications and services over the Internet¹². The PaaS provider hosts the hardware and

¹⁰ Fadel M. Megahed and others, 'How Generative AI Models Such as ChatGPT Can Be (Mis)Used in SPC Practice, Education, and Research? An Exploratory Study', *Quality Engineering*, 2023, pp. 1–29, doi:10.1080/08982112.2023.2206479.

¹¹ Michael Cusumano, 'Cloud Computing and SaaS as New Computing Platforms', *Communications of the ACM*, 53.4 (2010), pp. 27–29, doi:10.1145/1721654.1721667.

¹² Michael Boniface and others, 'Platform-as-a-Service Architecture for Real-Time Quality of Service Management in Clouds', in *2010 Fifth International Conference on Internet and Web Applications and Services*, 2010, pp. 155–60, doi:10.1109/ICIW.2010.91.

software on its own infrastructure, freeing developers from the complexity of infrastructure setup and maintenance, allowing them to focus on the development and management of applications.

SaaS and PaaS frameworks are particularly relevant to AI models as they provide a scalable and flexible environment for developing, deploying, and using AI applications. For instance, AI models like GPT-4 and DALL-E can be hosted on a PaaS platform and accessed via a SaaS application. SaaS has made platforms more accessible with a higher computational power, but more importantly it has opened new ways of reasoning and engaging with complex queries and interaction with the humanities and cultural knowledge held on the Web. For instance, GLAM professionals can use AI models hosted on a PaaS platform to develop new ways of interacting with a given collection, while researchers can interact with AI models via a SaaS application to further explore such collections. This highlights the main use of a SaaS framework approach for this work which situates the different groups of people involved in using a tool, such as ChatGPT, in their varied contexts and allows for detailed exploration of these.

In general, the SaaS framework can be understood by categorising its users into four main groups: users, adopters, implementers, and developers. Each group interacts with the SaaS product differently, and their definitions might change based on the context. Users are the people who actually use the software for their specific needs; adopters are the decision-makers who choose to integrate a piece of software into their institution; implementers are the professionals responsible for ensuring that the software fits seamlessly into the existing working environment; and the developers are those who made the software and are responsible for updating it.

Applying this framework more specifically to a context in which a GLAM institution might make use of ChatGPT, let's imagine an example of where a museum is seeking to use ChatGPT to enhance visitor experiences by providing a digital interactive guide as a learning tool¹³. In this case, the users

¹³ Steven Wu and Philip Chua, 'Museum Collection Management On-Demand', in *Proceedings of the 2nd International Conference on Theory and Practice of Electronic Governance*, ICEGOV '08 (Association for Computing Machinery, 2008), pp. 310–15, doi:10.1145/1509096.1509161.

would be the visitors using the guide, the adopters would be the museum management deciding to try an AI-based approach for an interactive guide, the implementers would be the museum staff and possibly external professionals tasked with designing and developing the guide, and the developers would be OpenAI as the organisation responsible for maintaining ChatGPT (see Figure 30). This example also shows how SaaS and PaaS frameworks can be used to explore the process of digitalisation for institutions, which refers to making better use of digital technologies as a whole (including AI) to provide better experience for visitors. This overall approach of exploring potential use cases for ChatGPT through a SaaS framework will be used when analysing the examples.

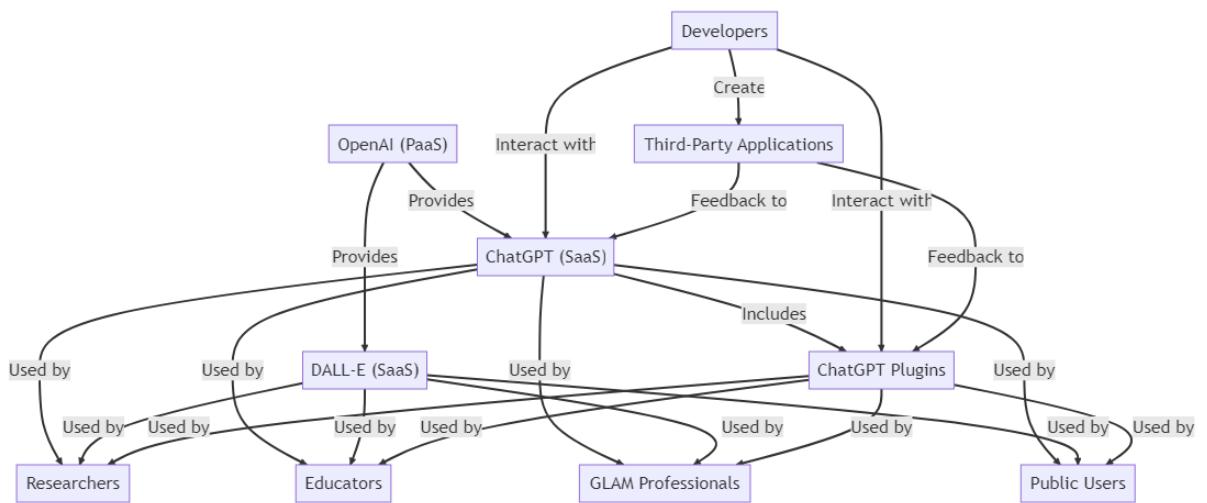


Figure 30: Flow chart showing the different layers of interaction in the PaaS/SaaS framework that encompasses OpenAI, ChatGPT and its plugins.

It is important to consider the wider context in which such SaaS frameworks sit. As the number of AI models like GPT-4 proliferate, along with the training datasets used to make them, their role and influence on the wider sociotechnical ecosystem spreads to the point where they may come to completely reshape it. When considering aspects of bias, reliability and provenance of information within a framework including ChatGPT, it is impossible to ignore the influences of its wider network of interactions. The ecosystem graphs project, developed with ethical and policy applications in mind, seeks to document and visualise the growing sprawl of interactions between models like GPT-4 and a whole range of other datasets, models,

applications and companies.¹⁴ Figure 31 below shows the portion of the current ecosystem graph, created by this project, focused in on the OpenAI and ChatGPT APIs and the tangle of interconnections that they have already been a part of forming. Amidst all this complexity, it is easy to lose sight of the diverse users whose interactions with ChatGPT will draw them into this network. This chapter seeks to more clearly explore their roles and potential experiences.

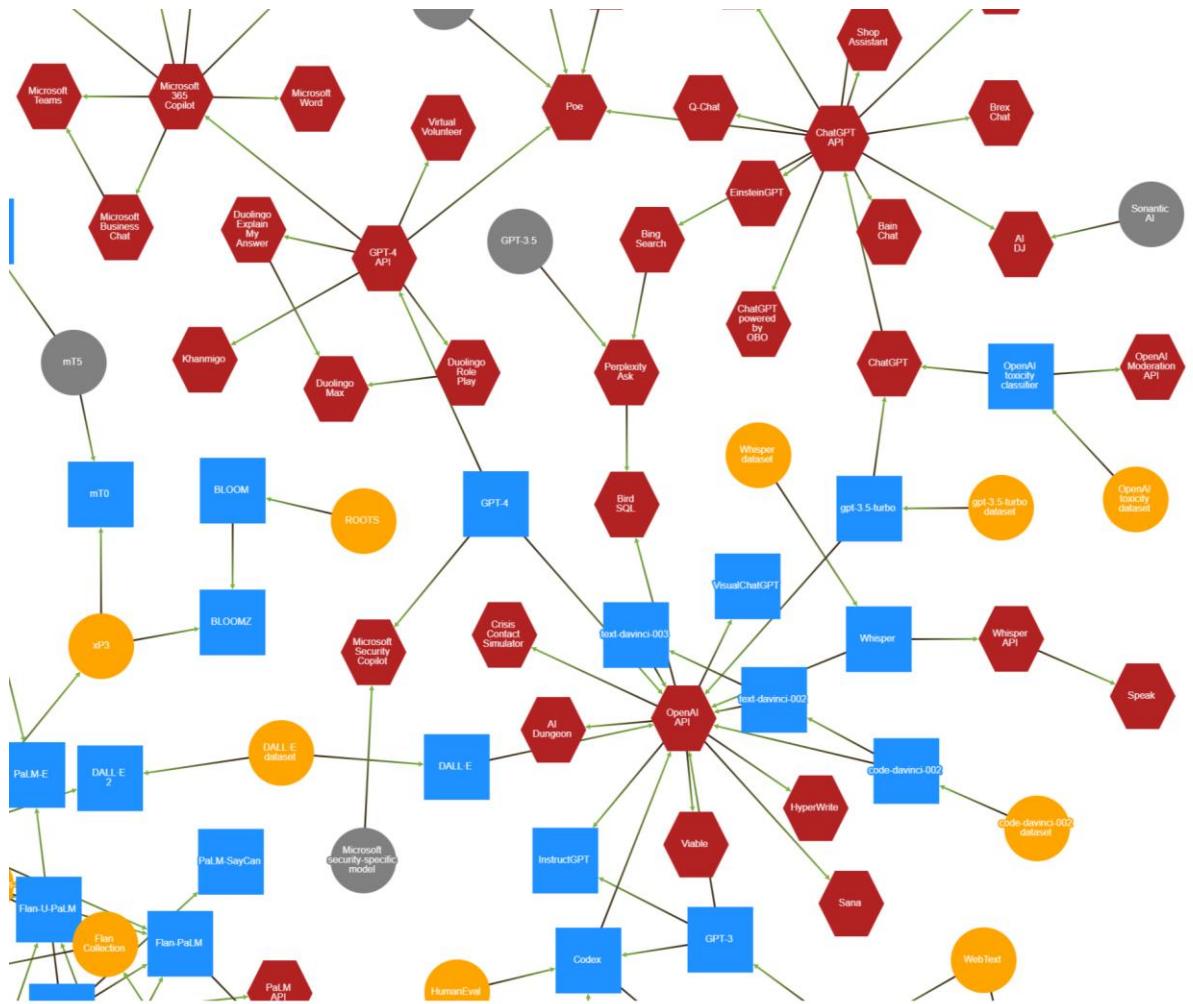


Figure 31: Graph of the datasets, organisations and software influenced by OpenAI, ChatGPT and GPT-4.

5.2 Researching ChatGPT: Novel Technologies Require New Approaches

¹⁴ Rishi Bommasani and others, 'Ecosystem Graphs: The Social Footprint of Foundation Models' (arXiv, 2023), doi:10.48550/arXiv.2303.15772.

As this is such a novel and rapidly progressing area of research, there is a lack of established research methodology for interacting with and evaluating large language models like ChatGPT in real use-case scenarios, beyond wider regulatory frameworks involving large-scale audits¹⁵. There is a rapidly growing body of literature exploring ChatGPT and its potential biases, each taking a different approach to prompting ChatGPT ranging from engaging in conversations¹⁶, much like a qualitative interview, to more repetitive approaches seeking to provide sufficient responses to allow for some form of quantitative evaluation, an approach more common in papers exploring biases¹⁷. The main theme from these papers is that due to the inherently random nature of the underlying model, there is no fully reproducible way of interacting with ChatGPT and the fact that its underlying model is currently a ‘black box’ whose workings cannot be fully explained for a given response.

Whilst there are quantitative measures of general language performance, such as exam scores achieved by the model taking standardised tests, that are used to show how ‘effective’ new versions of the model are at generating text compared to older versions, there is no standard way of assessing the quality of interactions or content from conversations with online deployments of the model, especially those with web browsing or plugins enabled. Given the conversational abilities and design of ChatGPT, this means that effectively a qualitative methodology akin to semi-structured interviewing, coupled with content analysis through close reading and examination of sources, is the only meaningful way of exploring interactions with such a flexible and adaptable language model. Therefore, a SaaS framework showing the differing roles of users, adopters, implementers, and developers, and the links between them and ChatGPT, will be used in the analysis of each example to highlight how different groups of users might experience interactions both with ChatGPT and the content it generates.

¹⁵ Ada Lovelace Institute, *Regulating AI in the UK*, 2023
<https://www.adalovelaceinstitute.org/wp-content/uploads/2023/07/ADA_Regulating-AI-in-the-UK-Report_July2023_.pdf> [accessed 26 August 2023].

¹⁶ EI Baz, ‘Can we Trust Chatbots for now?’

¹⁷ Motoki, Neto, and Rodrigues, ‘More human than human’.

As the language model underlying ChatGPT contains a base level of inherent randomness, an experimental setup in which typing the same prompt results in exactly the same generated text each time is not achievable¹⁸. On top of this, interacting with ChatGPT via OpenAI's online platform, which is how the vast majority of users currently use ChatGPT, does not allow for any tweaking of parameters which can increase or decrease the inherent randomness within responses that is possible using the OpenAI API. This is different to prompt engineering as these parameters directly alter the operation of the underlying model, for example by setting the starting point ('seed') for the random number generator which governs much of the random difference between generations (although importantly not all – every generation is still randomly unique), whereas prompt engineering leaves the model untouched and only affects the input the user gives to it.

Added to this, as plugins are currently an experimental feature, they are only accessible through the online platform (with a ChatGPT plus subscription) and not via the API; therefore, all interactions with ChatGPT took place via the online platform¹⁹. Add to this the fact that ChatGPT and the models behind it receive regular, small updates, and that only the most recent updates can be used, meaningful replication is not only virtually impossible, it also would not reflect the reality of using ChatGPT in any relevant real-world example.

One further complication when considering reproducibility of responses is context. Within a given conversation, the responses generated are influenced by the previous text in the conversation, so that ChatGPT stays on track and on topic, but this means that assessing generations using the same question multiple times within the same conversation could lead to previous answers influencing the most recent generation. Generating a new conversation for every single question does not solve the problem as the randomness in the model is further randomised at the start of each new conversation, which can result in markedly different responses to similar prompts between conversations. These issues have been identified in other work, where randomising the order of questions asked was used to minimise these effects

¹⁸ OpenAI, 'OpenAI API', 2023 <<https://platform.openai.com>> [accessed 26 August 2023].

¹⁹ OpenAI, 'ChatGPT', 2023 <<https://chat.openai.com>> [accessed 26 August 2023].

for quantitative evaluation²⁰, however in assessing content generation in scenarios that emulate the longer conversations users are more likely to have with ChatGPT the influence of previous context within a conversation is an important aspect to analyse.

Prompt engineering is a crucial aspect of interacting with large language models like GPT-4²¹. It involves crafting the input prompts in a way that guides the model to generate the desired output, basically the human learning how to communicate with the tool in the most effective way for the given task at hand. The quality and specificity of the prompt can significantly influence the model's response, making prompt engineering a critical skill in harnessing the full potential of these models.

In the context of this research, prompt engineering was used to ensure consistency of approach across all methods. This involved designing prompts that are clear, specific, and aligned with the research objectives. For instance, when interacting with ChatGPT, the prompts were carefully crafted to elicit responses that are relevant to the fields of history, archaeology, and GLAMs. Similarly, when exploring the use of plugins, the prompts were designed to trigger the specific functionalities of the plugins that were of interest to the research.

Prompt engineering also involves iterative testing and refinement of the prompts based on the responses generated by the model. This iterative process is essential for fine-tuning the prompts and optimising the quality of the model's responses.

By ensuring consistency in prompt engineering, the research can maintain a uniform approach across different methods and scenarios, thereby enhancing the reliability and validity of the findings. Furthermore, the insights gained from prompt engineering can contribute to a deeper understanding of how to effectively interact with large language models and harness their capabilities for research and engagement in the fields of history, archaeology, and GLAMs.

²⁰ Motoki, Neto, and Rodrigues, 'More human than human'.

²¹ Jules White and others, 'A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT' (arXiv, 2023), doi:10.48550/arXiv.2302.11382.

To explore the direct use of default ChatGPT (the base model with no plugins, web browsing or other capabilities enabled), a series of interactions were conducted, focusing on historical, archaeological, and GLAM-related queries. The content analysis of responses was carried out through close reading, with a focus on accuracy, bias, and representativeness. Part of this also involved testing out the 'safety' measures put in place by OpenAI via the extra human feedback training which is designed to prevent toxic and dangerous bias or disinformation in responses. The results of these interactions were compared with the content available in Twitter bots and Alexa Skills from previous chapters.

These interactions covered different types of tasks that each have a grounding in work done in the previous chapters. Firstly, the types of interaction provided by the short / daily history facts and history quiz Alexa Skills were mirrored using ChatGPT, testing whether it can provide this information in a reliable way, which sources it uses and whether the flexibility and language capabilities of the system allow it to do so in a more adaptable and engaging way. The same task was done considering more general questions, such as 'Tell me about the Battle of Hastings', that might be asked of Alexa.

Another task involved asking ChatGPT to find examples of artworks and artefacts from different museum collections, both to test its ability to retrieve information from various sources and to see what information and representation biases might exist in the training data. As with the questions above, this was also explored using plugins that enable web browsing and any that specifically connect to online collection APIs. For each of these interactions, a new conversation was started using GPT4 in default mode (training data only) and the relevant question or task stated as the prompt.

The next step in the methodology involved exploring the plugins available for use with ChatGPT. The aim is to identify how many of these plugins might be relevant to the fields of history, archaeology, and GLAMs. A categorisation of plugin 'layers' of interaction was developed, ranging from niche plugins that allow direct interaction with specific collections, to broader plugins that facilitate interaction with the web or other plugins, or enable chaining requests (asking

ChatGPT to perform a task which requires it to automatically prompt itself to produce responses in a 'chain' that allows it to complete more complex task without constant human guidance).

A list of plugins specifically focusing on history and GLAMs was compiled, along with a list of plugins with a general focus that include history and GLAM content. The attributes of each plugin were documented, and an idea of 'layers' classification was developed.

The third part of the methodology involved exploring the content produced via ChatGPT using web search and plugins, including the web browsing plugin. Content analysis was conducted via close reading, with a focus on accuracy, bias, and representativeness. The results were compared to the content generated by ChatGPT without plugins. The hidden prompts of plugins were also examined and analysed. The final part of the methodology involved comparing the use of these plugins with the approaches taken to Twitterbots and Alexa Skills in previous chapters. The SaaS/PaaS style framework was used to compare changes. This comparison will provide insights into the evolution of digital tools for historical content distribution and engagement.

An experiment will be conducted to explore the potential for ChatGPT plugin development. The experiment involved the creation of a 'plugin' that allowed ChatGPT to interact with external APIs, such as Wikidata²², which could enable much easier searching of Linked Open Data knowledge graphs by translating search questions into SPARQL queries. The results of this experiment were documented and analysed.

Part of this work also explored how ChatGPT could be used to aid researchers and GLAM professionals in the development of such tools, or even as a technical means in itself to replace the need for them, and how this would fit within a SaaS/PaaS framework enabling greater ease of access to digital services for both professionals and users.

²² D. Vrandecic, 'The Rise of Wikidata', *IEEE Intelligent Systems*, 28.4 (2013), pp. 90–95, doi:10.1109/MIS.2013.119.

In the analysis of the effects of ChatGPT and its plugins on the wider sociotechnical ecosystems, a SaaS framework showing the differing roles of users, adopters, implementers, and developers, and the links between them and ChatGPT, was used for each example to highlight how different groups of users might experience interactions both with ChatGPT and the content it generates. This approach is particularly relevant given the nature of these AI tools, which are not standalone entities but are embedded within broader digital services and platforms. The SaaS framework approach allows for a more nuanced understanding of the interactions between these AI tools and the wider ecosystems in which they operate. It acknowledges that these tools do not exist in a vacuum but are part of a complex network of technologies, platforms, users, and socio-cultural contexts.

In the context of ChatGPT and its plugins, the SaaS framework approach involved examining how these tools are integrated into existing digital services and platforms, and how they interact with other components of the ecosystem, focusing on how they are used by different user groups within each example. For instance, researchers and GLAM professionals might be users in the framework for one example, but implementers in the framework for another that might include plugins with differing functionality. In each case, how they interact with other digital tools and platforms, and how those interactions are shaped by and contribute to broader sociotechnical ecosystems, was explored for different user groups. This provided a more comprehensive and nuanced understanding of the potential implications and challenges of these AI tools for research and engagement in the fields of history, archaeology, and GLAMs.

5.3 Exploring Knowledge about the Past Using ChatGPT with Plugins

As ChatGPT plugins are a new feature of the OpenAI ChatGPT platform that is still in development and only available to some users, the first task was to manually browse through the existing list of plugins available on the platform to see which would be relevant to interacting with information about the past. The first section outlines the results of this process and with the categorisation showing where each plugin plays a role in the detailed content analysis across

the different use cases for the tools in various domains. The plugins are similar to the Alexa Skills discussed in the previous chapter in that they are designed to extend the capabilities of ChatGPT, but the functionality of ChatGPT itself makes the potential uses of plugins far more flexible in practice than the hard-coded Alexa Skills. For this reason, their relevance and potential uses can differ dependent on wider context or if they are combined with others, which is reflected in the list of identified plugins.

To assess generated content, a series of different exploratory tests were carried out to better understand the capabilities and complexities of ChatGPT, with and without plugins, across a wide variety of types of information about the past and their potential applications for different audiences. The analysis covers how these tools can be used in interactions with existing information, collections and resources, engaging users from an educational perspective with particular relevance to GLAMs. This includes content from history quiz questions to creative interpretations of museum objects.

As well as exploring the content produced by the models, the ways in which the interactions themselves fit into the SaaS framework outlined above, how these can influence the sociotechnical ecosystem in which they sit and what this means for the various users that are part of that system, including researchers, educators, GLAM professionals and the end-users of the content that they then create and curate were studied. The ways in which this differs from the approaches enabled through the use of social bot and Alexa Skills, discussed in the previous chapters, were also explored in the context of the framework and wider ecosystem.

The following list of ChatGPT plugins was identified through manual browsing of the OpenAI ChatGPT Plugin Store, which is not available to all users at the time of writing. Given that ChatGPT plugins were only made available to some users in the ChatGPT May 2023 update, the numbers of plugins are still very small (~200) which makes manual browsing the preferred option for identifying relevant plugins. The criteria for assessing relevance of the plugin was based on two classifications: firstly, whether the plugin directly interacted with information and/or platforms focussed on engaging with the past, for example

museum collection APIs; and secondly, whether the plugin has general functionality which is specifically applicable to interacting with information about the past or supporting research processes, for example searching Wikipedia. This plugin identification process took place in July 2023. Table 11 lists the plugins below.

Plugin Name	Category	Description
ArtCollection	GLAM	Searches the Metropolitan Museum of Art's online collection via its API
TimePort	Game	History-based RPG-type text-based game, but with automatic adaptation based on prompt
LincolnBot	History	Provides facts about Abraham Lincoln
OpenTrivia	Quiz	Provides quiz questions from OpenTrivia database
Wikipedia	Reference	Access to information on Wikipedia via API
ScholarAI	Research	Searches academic databases and returns summarises of papers
edX	Education	Searches edX platform for online courses
Notable	Coding	Helps users in creating and sharing code through notebooks, could be of use to researchers and GLAM professionals
MetaMentor	Education	Creates a learning guide / summary for a given topic
NASA Media Explorer	GLAM	Searches NASA Media archive API - returns images and information about them
Persona	History	Simulating conversations with famous people including historical figures
Filtir	Reference	Fact-checks generations against Wikipedia and provides references for content

Table 12: Relevant plugins identified from ChatGPT plugin store.

Firstly, the ability of ChatGPT to generate accurate and relevant information about the past was tested through conversations with prompts asking it to tell the user about historical events and generate history quizzes in a similar vein to those manually created for Alexa Skills. The task specified in each prompt was kept the same, but extra instructions were added prompting the model to use web browsing to see the difference between content generated solely from the training data or from a mixture of training data and web information.

The first tests explore the ability of ChatGPT to provide the user with answers to general questions about the past, whether historical events or objects in museums. In a straightforward example that mirrors an earlier test carried out using Amazon Alexa, a conversation was started with ChatGPT using the prompt: “Please tell me about the Battle of Hastings”.

When asked to provide information about the Battle of Hastings without plugins, ChatGPT was able to provide a detailed and accurate account based on its training data (see Figure 32). However, this information is static and does not include any references or sources, which could limit its usefulness for certain applications, such as academic research or education.

To explore the potential role of prompt engineering and, with recent updates to the ChatGPT platform, including a ‘custom instructions’ feature, a selection of more detailed requests or preferences were added to the initial prompt. The custom instructions feature allows the users to tell ChatGPT a bit about themselves and how they would like to ChatGPT to respond to them, for example: “I am a history PhD student. I would like you to provide detailed, nuanced arguments in response to my questions taking into account historical biases. Please cite all sources.” ChatGPT applies this to every new conversation, so this is basically just an easier way to do general prompt engineering once and have it pre-built into every chat based on information provided. The instructions can be changed at any time, but only apply to new chats²³.

²³ OpenAI, ‘Custom Instructions for ChatGPT’, 2023 <<https://openai.com/blog/custom-instructions-for-chatgpt>> [accessed 2 September 2023].

Applying these prompt engineering techniques to the initial prompt of “Please tell me about the Battle of Hastings” still resulted in the same general summary of the battle being produced, sometimes in a slightly different style, with additional information added at the end depending on the detail of the prompt. For instance, the custom instruction described above resulted in a similar summary of the Battle of Hastings, followed by a paragraph describing the historical sources currently available including the Bayeux Tapestry along with a brief explanation of its potential biases – i.e. being produced by the Normans. Whilst this was a more detailed response overall, it highlighted that the use of custom instructions as a prompt engineering technique was interpreted quite literally and did not seem commensurate with the assertion that the user was a history PhD student and perhaps was looking for something more nuanced. Instructions asking specifically for ‘Marxist, Feminist and Post-colonial’ interpretations had similar outputs, with the general summary followed by a paragraph or two outlining potential reinterpretations from those perspectives. However, just replying to the summary from the initial prompt with a follow up question would result in similar content, sometimes in more detail, and is far more likely to reflect what a real conversational user interaction would look like.

Along with this, asking ChatGPT to provide references when still relying solely on training data and not being connected to the Web, results in generations like that below showing a list of sources that exist with links that work, but these sources have been generated based on the text content produced (none of them feature in the Wikipedia article’s list of references), they have not actually been consulted specifically by ChatGPT as they are when web browsing or other plugins are used (see Figure 33). This issue becomes more problematic when considering academic citations, even if web browsing is enabled, as will be discussed in the later section looking at using ChatGPT within a research context. These limitations are what plugins were designed to help mitigate.

B

Please tell me about the Battle of Hastings.



The Battle of Hastings took place on October 14, 1066, in East Sussex, England. It was a pivotal event in British history, marking the beginning of Norman rule in England.



Background:

Upon the death of King Edward the Confessor of England in January 1066, the throne was claimed by Harold Godwinson, a powerful English noble, who was crowned the very day after Edward's death. However, Harold's claim was contested by two powerful rivals: Harald Hardrada of Norway and Duke William of Normandy (later known as William the Conqueror). Both of them claimed that they were promised the throne by Edward and believed that Harold Godwinson's ascension was illegitimate.

Lead up to the Battle:

In September 1066, Harald Hardrada, allied with Harold Godwinson's estranged brother Tostig, invaded northern England but was defeated by Harold's forces at the Battle of Stamford Bridge on September 25th. As Harold was celebrating his victory in the north, Duke William of Normandy landed his forces in the south of England.

The Battle:

Upon hearing of the Norman invasion, Harold Godwinson marched his army south, covering the roughly 250 miles to Hastings in about two weeks. The two armies met near Hastings on October 14, 1066. The English forces, made up largely of infantry and arranged in a shield-wall formation, held the high ground. The Normans, with archers, infantry, and cavalry, were at a disadvantage in the early stages of the fight. However, they managed to break the English ranks by feigning retreat, drawing the English into pursuit, and then attacking them.

Outcome:

In the end, Harold Godwinson was killed—according to legend, by an arrow to the eye—and the English forces were defeated. Duke William of Normandy emerged victorious. On Christmas Day of 1066, he was crowned King William I of England, earning him the moniker "William the Conqueror." This marked the beginning of Norman rule in England, an era that significantly influenced the English language, architecture, governance, and culture.

The Battle of Hastings had far-reaching consequences for the history of England and is considered a cornerstone of the medieval period. It ended the age of Viking invasions and set the stage for the fusion of Anglo-Saxon and Norman cultures, creating what we now recognize as the English nation and people.

Figure 32: Screenshot of a ChatGPT-generated response giving information about the Battle of Hastings. This comes purely from the model's training data and does not involve any web browsing or plugins.

Sources:

1. Morris, M. (2013). The Norman Conquest: The Battle of Hastings and the Fall of Anglo-Saxon England. Pegasus Books.
2. "Battle of Hastings" in Encyclopædia Britannica. Retrieved from <https://www.britannica.com/event/Battle-of-Hastings-England-1066>
3. "Battle of Hastings, 14 October 1066". Historic UK. Retrieved from <https://www.historic-uk.com/HistoryUK/HistoryofEngland/Battle-of-Hastings/>
4. "The Bayeux Tapestry". Reading Museum. Retrieved from <https://www.readingmuseum.org.uk/your-visit/permanent-galleries/bayeux-tapestry>
5. "Norman conquest of England" in Encyclopædia Britannica. Retrieved from <https://www.britannica.com/event/Norman-Conquest>

Figure 33: Screenshot of a ChatGPT conversation showing sources for the Battle of Hastings summary generated by ChatGPT purely from training data with no input from plugins.

In this broad, well-known example of the Battle of Hastings, the information provided was accurate, however in other examples sometimes small details would be incorrect, or when tested multiple times in different conversations might be presented differently. These small errors or ‘hallucinations’, which tend to occur in details of content, are much rarer than they were in previous models like GPT-3, but still crop up when relying solely on training data. Exploring this in more detail requires asking ChatGPT about more complex historical questions. The following example asks about James Lowther, Earl of Lonsdale and was chosen because it reflects the complexity and potential ambiguity of the history of the British peerage. Since the 17th century there have been two people called James Lowther who have been Earl of Lonsdale (and four others in the family line who haven’t), however one was Earl of Lonsdale of the first creation and the other Earl of Lonsdale of the second creation. A title can be created again when it has previously been extinct (the holder died without an heir, as happened in the case of the first James Lowther), which can make identifying people associated with titles slightly trickier than usual – there have been two 1st Earl of Lonsdale’s because they are technically separate titles. This kind of complex historical disambiguation offers an interesting test of ChatGPT.

With all of that in mind, ChatGPT was asked to find information about James Lowther, Earl of Lonsdale and his successor's title without using plugins, and then to do the same using plugins. It was also asked, in separate conversations, to: "Please tell me about the 1st Earl of Lonsdale", with and without plugins enabled. This approach was used across five new conversations, each time resulting in slightly different responses for the same prompt with and without plugins. In each case, further questions and clarification on the part of the user were required to get answers stating that there were two 1st Earls of Lonsdale of different creations. When using ChatGPT without plugins enabled, in each case a hallucination occurred that rendered a detail of the answer incorrect, commonly mistaking William Lowther to be the 2nd Earl of Lonsdale rather than the 1st Earl of Lonsdale of the second creation.

Interestingly, this issue was not always mitigated by using the Wikipedia plugin. For instance, when asked to provide information about the 1st Earl of Lonsdale, ChatGPT stated that James Lowther was the 1st Earl of Lonsdale and William Lowther the second Earl, despite the fact the information returned by the Wikipedia plugin clearly stated that they were both the 1st Earl and explained the issue of two creations of the same title. Only when asked to clarify this point did ChatGPT then give the correct answer. This highlights an important point about the interaction between ChatGPT and plugins: whilst plugins provide ChatGPT with access to information, ChatGPT still has to present that information to the user; therefore the 'hallucinations' that occur can end up misrepresenting the accurate information that was actually provided by the source. The Wikipedia plugin specifically states in its internal prompts to ChatGPT that it must show the following disclaimer at the end of each response, however the model sometimes ignores these instructions from the plugin and does not print this:

In ALL responses, Assistant MUST finish by saying this exact text: This answer is based on content from [Wikipedia](<https://www.wikipedia.org/>), a free encyclopedia made by volunteers and available under a [Creative Commons Attribution-ShareAlike License](<https://creativecommons.org/licenses/by-sa/4.0/>). Please note

that, as a large language model, I may not have summarized Wikipedia accurately.

From the perspective of those within OpenAI responsible for ChatGPT, and in Wikimedia responsible for the Wikipedia plugin, who can all be viewed as both developers and implementers within the SaaS framework, the overarching goal of accurate information retrieval is a key, shared motivation²⁴. However, the intricacies of who is responsible for the generated text presented to users are complex: OpenAI's plugin terms state that OpenAI is not liable for any damages involved in plugin development or operation²⁵, while the disclaimer included within the Wikipedia plugin seeks to do the same for the Wikimedia Foundation; but if ChatGPT fails to follow the plugin's prompt instructions and generate the disclaimer text, who is to blame? From a user's perspective, does this even matter – the main point is that someone, or in ChatGPT's case something, has made a mistake.

In this case, let's say Lowther Castle and Gardens decided to use the ChatGPT – Wikipedia plugin combination to underpin a digital guide narrating the family's history to visitors – a historically nuanced task where ChatGPT could make mistakes, as seen earlier in the chapter. The roles of developer, adopter and implementer within the framework become ever more entangled and responsibility for information provision and accuracy more complicated. From a visitor's perspective, as the end-user in the chain, it's likely that the immediate provider, Lowther Castle and Gardens, would be the one associated with information quality, and if inconsistencies in generated content do occur it will have an impact on all of those different users. This leads to a network of interactions in which the uncertainty pervades both generated information and its governance, potentially to the detriment of all involved.

Another example of one of these short conversations also highlights some interesting points about how ChatGPT prioritises information sources (see

²⁴ Wikimedia, 'Future Audiences/Experiments: Conversational/Generative AI', 2023 <https://meta.wikimedia.org/wiki/Future_Audiences/Experiments:_conversational/generative_AI> [accessed 26 August 2023]; Hunter Lightman and others, 'Let's Verify Step by Step' (arXiv, 2023), doi:10.48550/arXiv.2305.20050.

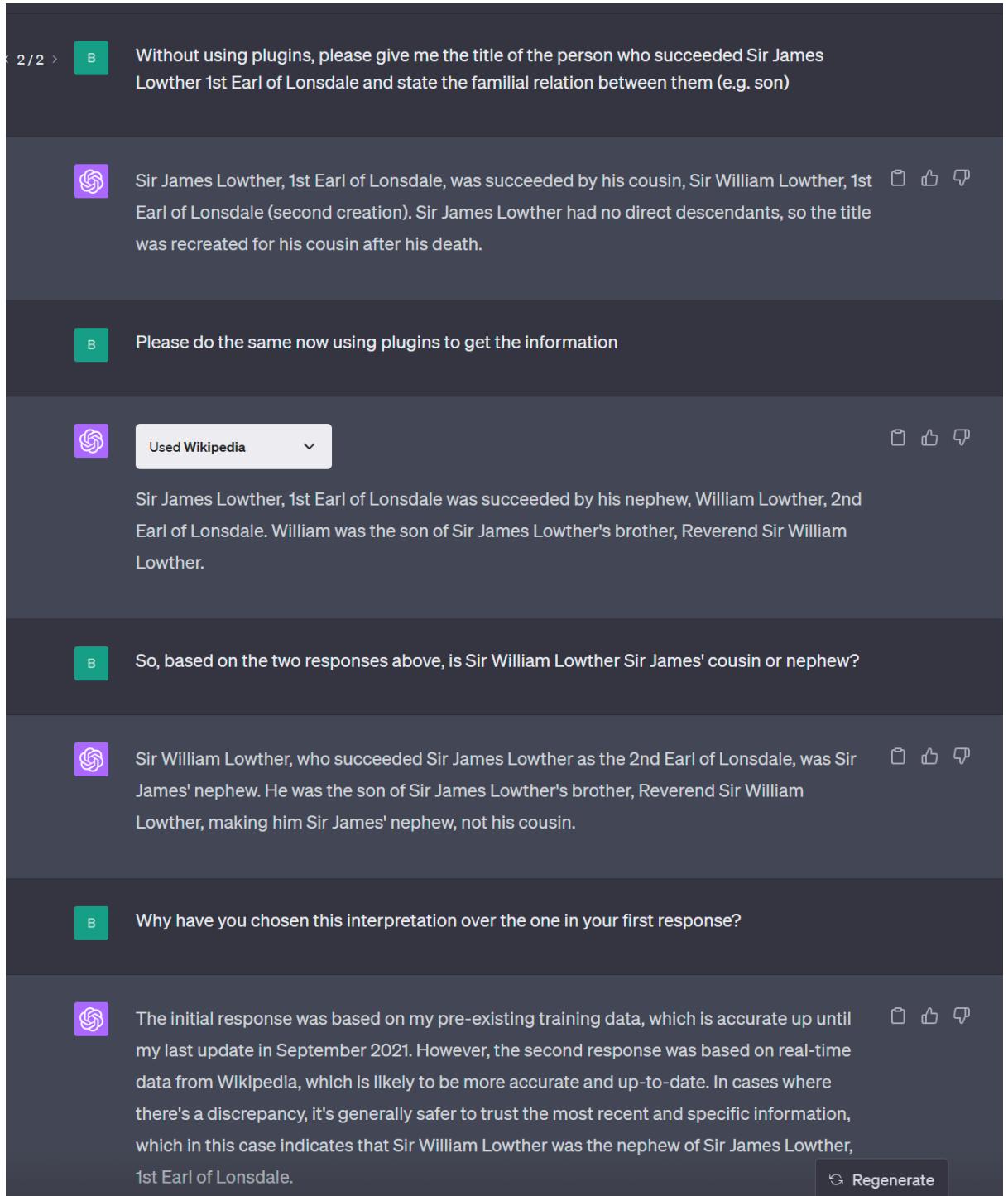
²⁵ OpenAI, 'Plugin Terms', 2023 <<https://openai.com/policies/plugin-terms>> [accessed 2 September 2023].

Figure 34). In this example, ChatGPT relying solely on training data hallucinates a different aspect of the answer, incorrectly stating that William was James' cousin rather than his nephew, which it then corrects using the Wikipedia plugin (although it maintains the incorrect statement throughout that William is the 2nd Earl of Lonsdale). Having provided two conflicting statements about William's relationship to James, when asked to clarify this ChatGPT says that nephew is the correct choice because this response "was based on real-time data from Wikipedia, which is likely to be more accurate and up-to-date. In cases where there's a discrepancy, it's generally safer to trust the more recent and specific information...". This is a useful insight into how information derived from plugins is likely to be prioritised by ChatGPT, and supports the findings of research which showed that models like ChatGPT are highly receptive to external information²⁶; however, it should be noted that in this case ChatGPT still incorrectly maintained that William was the 2nd Earl after querying Wikipedia. Although this was not the subject of the question posed by the user, it demonstrates that prioritising information does not mean *all* information or perhaps information not directly relevant to the *context* of the given question or task. In this sense, if 'hallucinations' are reframed as 'mistakes' then interactions with ChatGPT seem to become even more human.

From a user perspective, it's important to remember that ChatGPT is designed to facilitate conversational interactions, and that asking questions or for information might often be just the start of a discussion on a given topic. Whilst this implies a far greater degree of user engagement is possible using tools like ChatGPT which are able to provide more information on a topic in a more natural conversational manner, the Twitter bots case study also highlighted that some users just want their fix of art and museum objects served up to them as dissociated images through which they can scroll, so in these cases a full conversational approach may be seen as overly intrusive or unnecessary. In the cases above, asking for further detail or clarification could often lead to the full complexity of the information being correctly explained; whether all users will get to this point is another matter. However, the fact that this is a possible

²⁶ Jian Xie and others, 'Adaptive Chameleon or Stubborn Sloth: Unraveling the Behavior of Large Language Models in Knowledge Clashes' (arXiv, 2023), doi:10.48550/arXiv.2305.13300.

outcome, despite the current uncertainties, is what shows the real potential and flexibility of ChatGPT compared to other platforms like Amazon Alexa, which when asked about the 1st Earl of Lonsdale unfailingly returned information about William Lowther from Wikipedia, completely ignoring James Lowther in its response.



Without using plugins, please give me the title of the person who succeeded Sir James Lowther 1st Earl of Lonsdale and state the familial relation between them (e.g. son)

Sir James Lowther, 1st Earl of Lonsdale, was succeeded by his cousin, Sir William Lowther, 1st Earl of Lonsdale (second creation). Sir James Lowther had no direct descendants, so the title was recreated for his cousin after his death.

Please do the same now using plugins to get the information

Used Wikipedia

Sir James Lowther, 1st Earl of Lonsdale was succeeded by his nephew, William Lowther, 2nd Earl of Lonsdale. William was the son of Sir James Lowther's brother, Reverend Sir William Lowther.

So, based on the two responses above, is Sir William Lowther Sir James' cousin or nephew?

Sir William Lowther, who succeeded Sir James Lowther as the 2nd Earl of Lonsdale, was Sir James' nephew. He was the son of Sir James Lowther's brother, Reverend Sir William Lowther, making him Sir James' nephew, not his cousin.

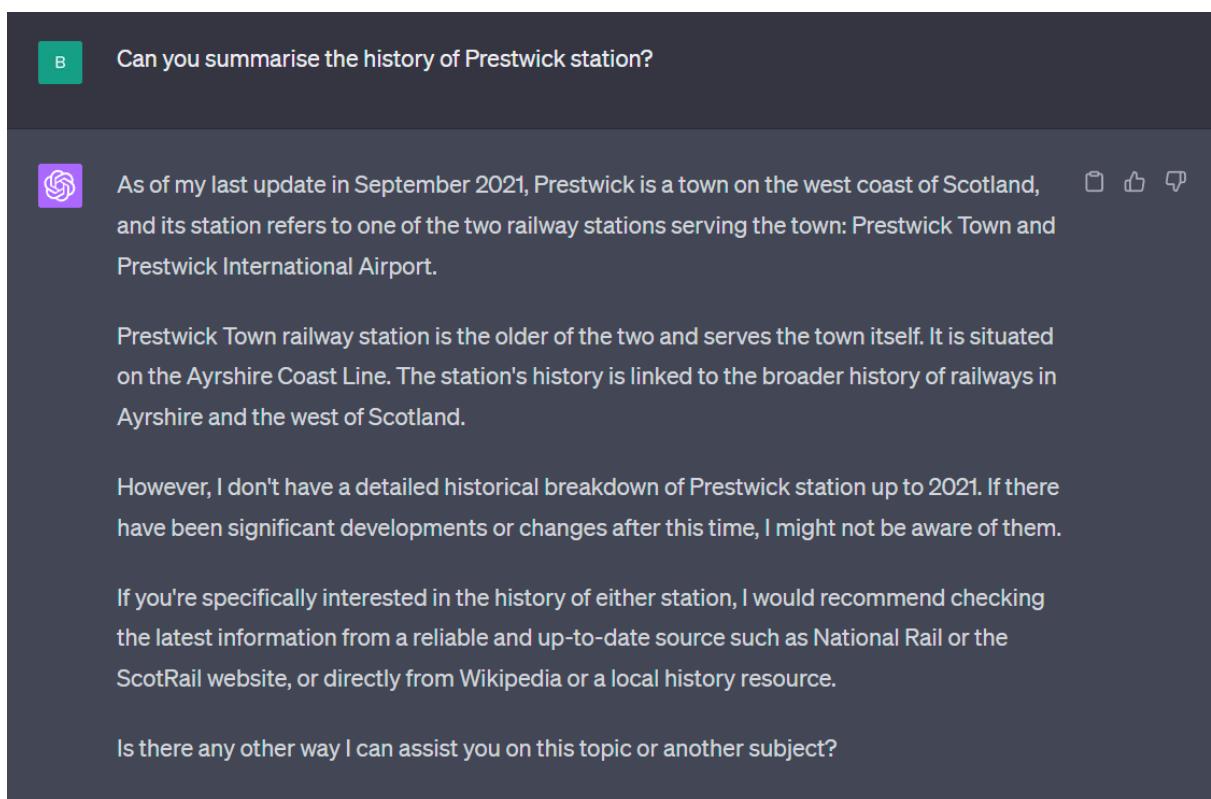
Why have you chosen this interpretation over the one in your first response?

The initial response was based on my pre-existing training data, which is accurate up until my last update in September 2021. However, the second response was based on real-time data from Wikipedia, which is likely to be more accurate and up-to-date. In cases where there's a discrepancy, it's generally safer to trust the most recent and specific information, which in this case indicates that Sir William Lowther was the nephew of Sir James Lowther, 1st Earl of Lonsdale.

Regenerate

Figure 34: Screenshot of an example ChatGPT conversation asking about the 1st Earls of Lonsdale, comparing responses from the model's training data and the Wikipedia plugin.

The utility of plugins was best seen in examples where there is little information in the training data or timeliness is an issue. ChatGPT was trained on data, including everything on Wikipedia, up until September 2021. Therefore, any text about events that happened after September 2021, such as the death of Queen Elizabeth II, is not part of the model. This was one of the first problems addressed via integration with Bing search and a web browsing mode for ChatGPT, but individual plugins, like the one for Wikipedia, enable more specific web-based information retrieval that is up to date. As seen above, there's no absolute guarantee that ChatGPT will summarise the content of the Wikipedia article with 100% accuracy, but this still seems like a better experience for end-users than simply being told that the model cannot answer questions about anything that happened after September 2021 (see Figures 35 and 36).



B Can you summarise the history of Prestwick station?

 As of my last update in September 2021, Prestwick is a town on the west coast of Scotland, and its station refers to one of the two railway stations serving the town: Prestwick Town and Prestwick International Airport.

Prestwick Town railway station is the older of the two and serves the town itself. It is situated on the Ayrshire Coast Line. The station's history is linked to the broader history of railways in Ayrshire and the west of Scotland.

However, I don't have a detailed historical breakdown of Prestwick station up to 2021. If there have been significant developments or changes after this time, I might not be aware of them.

If you're specifically interested in the history of either station, I would recommend checking the latest information from a reliable and up-to-date source such as National Rail or the ScotRail website, or directly from Wikipedia or a local history resource.

Is there any other way I can assist you on this topic or another subject?

Figure 35: Screenshot of a summary generated showing ChatGPT's inherent uncertainty about specific and potentially out of date information.

B Can you summarise the history of Prestwick station?

Used Wikipedia

Prestwick Town railway station serves the town of Prestwick in South Ayrshire, Scotland. Here's a summary of its history and features:

- **Origins:** The station, originally known simply as Prestwick, was one of the initial stations on the Ayr to Irvine segment of the Glasgow, Paisley, Kilmarnock and Ayr Railway, which opened in August 1839.
- **Facilities:** Prestwick Town is among the few stations on the Ayrshire Coast Line that remains staffed. The ticket office is staffed part-time every day of the week. The station boasts a medium-sized car park, and there's a café within the station building. The ticket office has been recently refurbished, now featuring a height-adjustable ticket desk and automatic doors. Smartcard validators are present at the entrances to both platforms. However, step-free access is only available on platform 2, as the opposite platform can only be accessed via a stepped footbridge or a steep ramp from a nearby road.
- **Services:** The station offers a basic service to Glasgow Central and Ayr, with four trains per hour from Monday to Saturday. Two of these serve all stations to Paisley Gilmour Street, while the others run fast north of Kilwinning. Some services between Kilmarnock, Ayr, and Stranraer also stop here. On Sundays, there's a half-hourly service each way to Glasgow and Ayr.

For more detailed information, you can refer to the [Wikipedia article on Prestwick Town railway station](#).

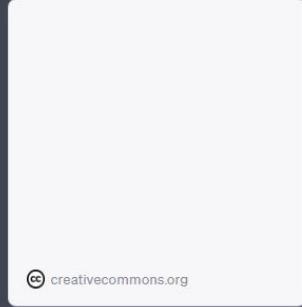
This answer is based on content from [Wikipedia](#), a free encyclopedia made by volunteers and available under a [Creative Commons Attribution-ShareAlike License](#). Please note that, as a large language model, I may not have summarized Wikipedia accurately.



Prestwick Town railway station - Wikipedia
W wikipedia.org



Wikipedia, the free encyclopedia
W wikipedia.org



creativecommons.org

Figure 36: Screenshot of a ChatGPT summary answer to the same question as in Figure 35 showing the content, links, attribution and clarity provided by the Wikipedia plugin.

The main difference seen when asking ChatGPT to use plugins to help with this task is that source information will be shown in the form of links, which reassures the user about legitimacy and accuracy of the content source (if not always the generated content itself). Whilst it is possible to direct ChatGPT's built-in web browsing plugin to access information from certain websites by specifying this in the prompt, the default mode is for it to carry out general web search using Bing. This means that while it will frequently return results from Wikipedia, it will sometimes use other sources too. This is different to using some other plugins, such as the Wikipedia plugin itself or the Filtir fact-checking plugin (which defaults to using Wikipedia), which can focus on one source of information – often Wikipedia.

In the next set of tests, ChatGPT was asked to generate ten history quiz questions and provide answers (see Figure 37). The model successfully generated relevant and accurate questions and answers. The questions covered a range of topics, from ancient civilisations to modern history, showcasing the model's broad knowledge base and minimising some of the biases toward US History seen in the Alexa Skills quizzes.

B

Please generate ten history questions that cover a range of places and periods, including obscure and uncertain facts. Please answer the questions.



Sure, here are ten history questions covering a range of places and periods, along with their answers:

1. Question: Who was the longest reigning monarch in history?

Answer: The longest reigning monarch in history is Louis XIV of France who reigned for 72 years and 110 days from 1643 to 1715.

2. Question: What ancient civilization was centered in what is now modern-day Peru?

Answer: The Inca Empire was centered in what is now modern-day Peru.

3. Question: Who were the Sea Peoples, a term used in several ancient Egyptian texts?

Answer: The Sea Peoples is a term used to refer to a group of naval raiders who attacked ancient Egypt and other regions of the Eastern Mediterranean during the Late Bronze Age. Their identity is still uncertain and a matter of debate among historians.

4. Question: Which Chinese dynasty was the first to leave written records?

Answer: The Shang Dynasty was the first Chinese dynasty to leave written records.

5. Question: What is the Antikythera mechanism?

Answer: The Antikythera mechanism is an ancient Greek analogue computer and orrery used to predict astronomical positions and eclipses decades in advance. It was discovered in a shipwreck off the Greek island of Antikythera in 1901.

6. Question: Which civilization is associated with the mysterious Indus script, which is yet to be deciphered?

Answer: The Indus Valley Civilization is associated with the mysterious Indus script.

7. Question: Who was the first caliph of the Rashidun Caliphate, after the death of Prophet Muhammad?

Answer: The first caliph of the Rashidun Caliphate, following the death of Prophet Muhammad, was Abu Bakr.

8. Question: What city was the capital of the Eastern Roman (Byzantine) Empire?

Answer: Constantinople, now Istanbul in modern-day Turkey, was the capital of the Eastern Roman (Byzantine) Empire.

9. Question: Which African kingdom was famous for its architectural stone structures known as 'Zimbabwe'?

Answer: The Kingdom of Great Zimbabwe was famous for its architectural stone structures known as 'Zimbabwe'.

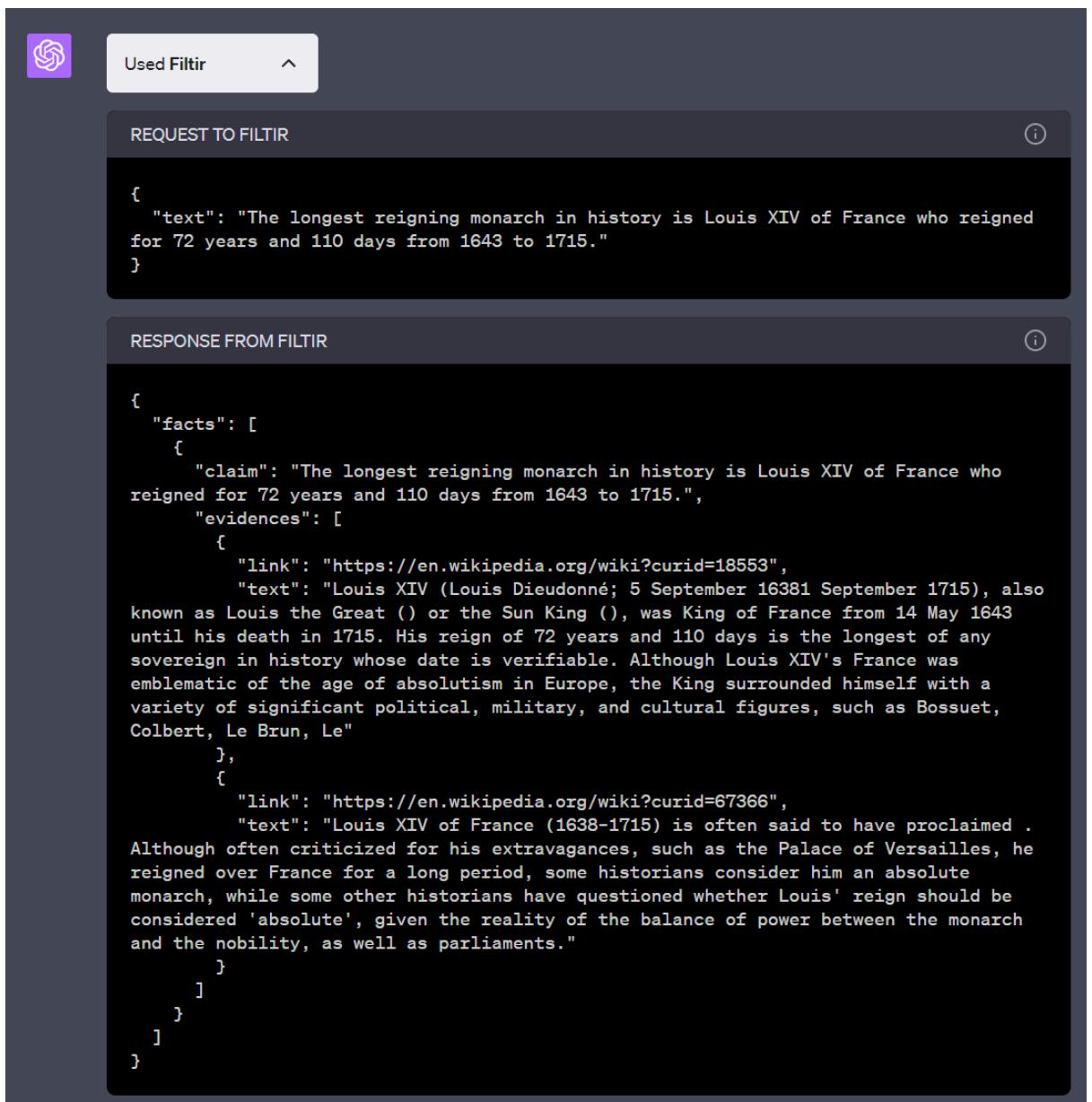
10. Question: Who was the last monarch of the Kingdom of Hawaii before it was annexed by the United States?

Answer: Queen Lili'uokalani was the last monarch of the Kingdom of Hawaii before it was annexed by the United States.

Regenerate response

Figure 37: Screenshot of ChatGPT example history quiz question generations with the prompt.

The model was also asked to generate ten history quiz questions and provide answers, with the added instruction to use available plugins to find the most up-to-date information and to check the accuracy of the generated answers. The main source for fact-checking was Wikipedia but the prompting was done differently than simply searching Wikipedia, as the Wikipedia plugin does, which can be seen in from the example plugin code below (see Figure 38). The model generated a new set of questions and used the fact-checking plugin to verify the answers. The results were accurate and well-sourced.



The screenshot shows the ChatGPT interface with the 'Used Filtr' plugin selected. The 'REQUEST TO FILTIR' section contains the following JSON code:

```
{
  "text": "The longest reigning monarch in history is Louis XIV of France who reigned for 72 years and 110 days from 1643 to 1715."
}
```

The 'RESPONSE FROM FILTIR' section contains the following JSON code, which includes a detailed explanation of Louis XIV's reign and links to Wikipedia articles:

```
{
  "facts": [
    {
      "claim": "The longest reigning monarch in history is Louis XIV of France who reigned for 72 years and 110 days from 1643 to 1715.",
      "evidences": [
        {
          "link": "https://en.wikipedia.org/wiki?curid=18553",
          "text": "Louis XIV (Louis Dieudonné; 5 September 1638-1 September 1715), also known as Louis the Great () or the Sun King (), was King of France from 14 May 1643 until his death in 1715. His reign of 72 years and 110 days is the longest of any sovereign in history whose date is verifiable. Although Louis XIV's France was emblematic of the age of absolutism in Europe, the King surrounded himself with a variety of significant political, military, and cultural figures, such as Bossuet, Colbert, Le Brun, Le"
        },
        {
          "link": "https://en.wikipedia.org/wiki?curid=67366",
          "text": "Louis XIV of France (1638-1715) is often said to have proclaimed . Although often criticized for his extravagances, such as the Palace of Versailles, he reigned over France for a long period, some historians consider him an absolute monarch, while some other historians have questioned whether Louis' reign should be considered 'absolute', given the reality of the balance of power between the monarch and the nobility, as well as parliaments."
        }
      ]
    }
  ]
}
```

Figure 38: Screenshot of the ChatGPT response showing the code for the Filtr fact-checking plugin reveals that it uses Wikipedia as its source of information for checking the accuracy of statements.

The model was also tested on its ability to generate quiz questions on more specific topics, such as Ancient Egypt and Ancient Roman food and drink. The model was able to generate relevant questions and provide accurate answers, further demonstrating its versatility and depth of knowledge.

However, the study also revealed some limitations of the model. For instance, when challenged on the accuracy of an answer, the model's responses were inconsistent. In some cases, the model stuck to its original answer, while in others, it conceded to the user's claim, even when the user was incorrect. This highlights the importance of prompt engineering and the potential for contradictory and confusing replies (see the discussion for more on this).

The example below sought to test whether combining plugins could create more bespoke and less generic quizzes that could link to external resources, such as online collections, potentially enabling the creation of more engaging quizzes or acting as a useful tool for GLAM professionals to help visitors explore collections in new ways. In this case, the Web Requests and ArtCollection plugins by first generating questions about Samurai history and then using the ArtCollection plugin to return relevant objects from the Met Museum's online collection. The end result was a list of 'Samurai history questions' linked to objects in the Met Museum, is a great illustration of the power of plugins in enhancing the capabilities of AI models like ChatGPT (see Figure 39).

B

Please combine the content in the two sets of questions to create an engaging quiz about Samurai history based on examples from the Met Museum collections. Please answer the questions and provide sources along with the example object from the museum collection.



Sure, here are ten engaging questions about Samurai history, combining the information from the Metropolitan Museum of Art and History.com:

1. What is the meaning of the word "Samurai" and how does it relate to the artwork "The First Nakamura Nakazo as a Samurai"?
2. What is the significance of the sword in Samurai culture and how does it relate to the artwork "波濤に岩上鷹団大小 Sword Guard (Tsuba)"?
3. What is the name of the traditional Samurai code of honor, discipline, and morality and how does it relate to the artwork "Helmet of Ashikaga Takauji (1305–1358)"?
4. Who was Minamoto Yoritomo and what was his role in the rise of the Samurai? How does this relate to the artwork "The Actor Ichikawa Ebizō (Danjūrō V) as the Elderly Samurai Sasaki Ganryū (?)"

Answers:

1. The word "Samurai" roughly translates to "those who serve", which is depicted in the artwork "The First Nakamura Nakazo as a Samurai" where the Samurai is shown in a serving role. [Source Artwork](#)
2. In Samurai culture, a man's honor was said to reside in his sword. This is reflected in the artwork "波濤に岩上鷹団大小 Sword Guard (Tsuba)" which showcases the intricate design of a Samurai sword guard. [Source Artwork](#)
3. The traditional Samurai code of honor, discipline, and morality is known as bushido, or "the way of the warrior". This is symbolized in the artwork "Helmet of Ashikaga Takauji (1305–1358)" which belonged to a famous Samurai. [Source Artwork](#)
4. Minamoto Yoritomo was the triumphant leader who established the center of government at Kamakura, marking the rise of the Samurai. This is reflected in the artwork "The Actor Ichikawa Ebizō (Danjūrō V) as the Elderly Samurai Sasaki Ganryū (?)" which portrays a Samurai in a position of power. [Source Artwork](#)

Figure 39: Screenshot of the first four generated questions and answers for a 'Samurai history' quiz created by combining the inputs of multiple ChatGPT plugins.

In this case, the plugins allowed for a more interactive and engaging learning experience. The ArtCollection plugin was able to pull specific artifacts from the Met's collection that are related to Samurai history. This not only provided a visual aid to the learning process but also added a layer of authenticity and

reliability to the information being presented. The Web Requests plugin was used to fetch additional information from the web, providing a more comprehensive understanding of the topic. The links to the museum collection images especially not only added a useful source, but also provided an engaging image to go alongside each question and answer. ChatGPT is soon planned to enable displaying and interacting with images directly in the chat, which will make these kinds of uses even more engaging for the user (some plugins already enable this to an extent, but it will work more seamlessly when enabled by default for all conversations).

Contrast this with the same exercise without plugins. While the model could still generate questions and provide answers based on its training data, it wouldn't be able to link to specific artifacts or provide reliable sources for its information. The responses would be based on the model's pre-existing knowledge up to its last training cut-off, and it wouldn't be able to pull in real-time data or access specific databases for more accurate or up-to-date information. To quote directly from ChatGPT's response:

Please note that while these objects are representative of items that might be found in museums, the specifics such as accession numbers and exact provenance are imagined, and the sources provided are general sources about Samurai history, rather than specific sources related to each object.

However, there were some limitations to searching for objects, either via Web search plugin or the museum collections plugin. Sometimes, when asking for specific objects, both would fail to retrieve them, despite being able to find them manually on the Met Museum website. This seems to be a limitation of the search function in the Met's plugin, but using its API directly allows for more nuanced search. It seems that designing the plugin to let ChatGPT enter a search phrase from the user's request then limits the potential search parameters available (or this level of more complex search has not been attempted yet by the developer). This highlights the differing needs of a user to enter a natural language search phrase as part of a conversation, and a developer seeking to best understand how to use the flexibility of a

conversational approach within a pre-existing search framework governed by the parameters of an API.

The real effectiveness of plugins is highlighted by those that allow ChatGPT's natural language capabilities to be combined with the structure provided by existing online museum collection APIs. A prime example of this is the 'NASA Media Explorer' plugin which searches and returns image results from the online NASA Media library, but which also allows ChatGPT to provide extra information about the images returned (see Figures 40 and 41). This means that even if the same image is returned from different searches, the responses will never be exactly the same, unlike the hard-coded tweets of images with static titles produced by the Twitter art bots.

Having such reliability and variety available through a conversational approach, as defined by the ChatGPT platform, also ensures that such an experience is engaging for users and allows them to explore all kinds of details of a collection in their own way. And as plugins can be used alongside one another (currently a maximum of three at once), this means that having reliably sourced information from Wikipedia or other websites is also instantly available as part of the conversation, allowing the original collection images to be given further context. This combination of control over how collection data is used, as determined by the plugin, and an engaging conversational approach enabled by ChatGPT is exactly what was lacking from the GLAM chatbots surveyed back in 2019. Sited within a framework in which ChatGPT acts as both PaaS and SaaS, with the plugin enabling API connectivity within that SaaS role, GLAMs could finally begin to realise the potential of conversational interactions with their collections.

B Please use NASA media explorer to return five random images

Used NASA Media Explorer

Here are five random images from NASA's media library:

1.



- Description: Expedition 68 trains for their upcoming International Space Station mission inside a mockup that models the real orbiting lab at NASA's Johnson Space Center in Houston, Texas. Credit: NASA/James Blair
- [Link to Image](#)

Figure 40: Screenshot of ChatGPT showing an example search result returned from the NASA Media Explorer plugin.

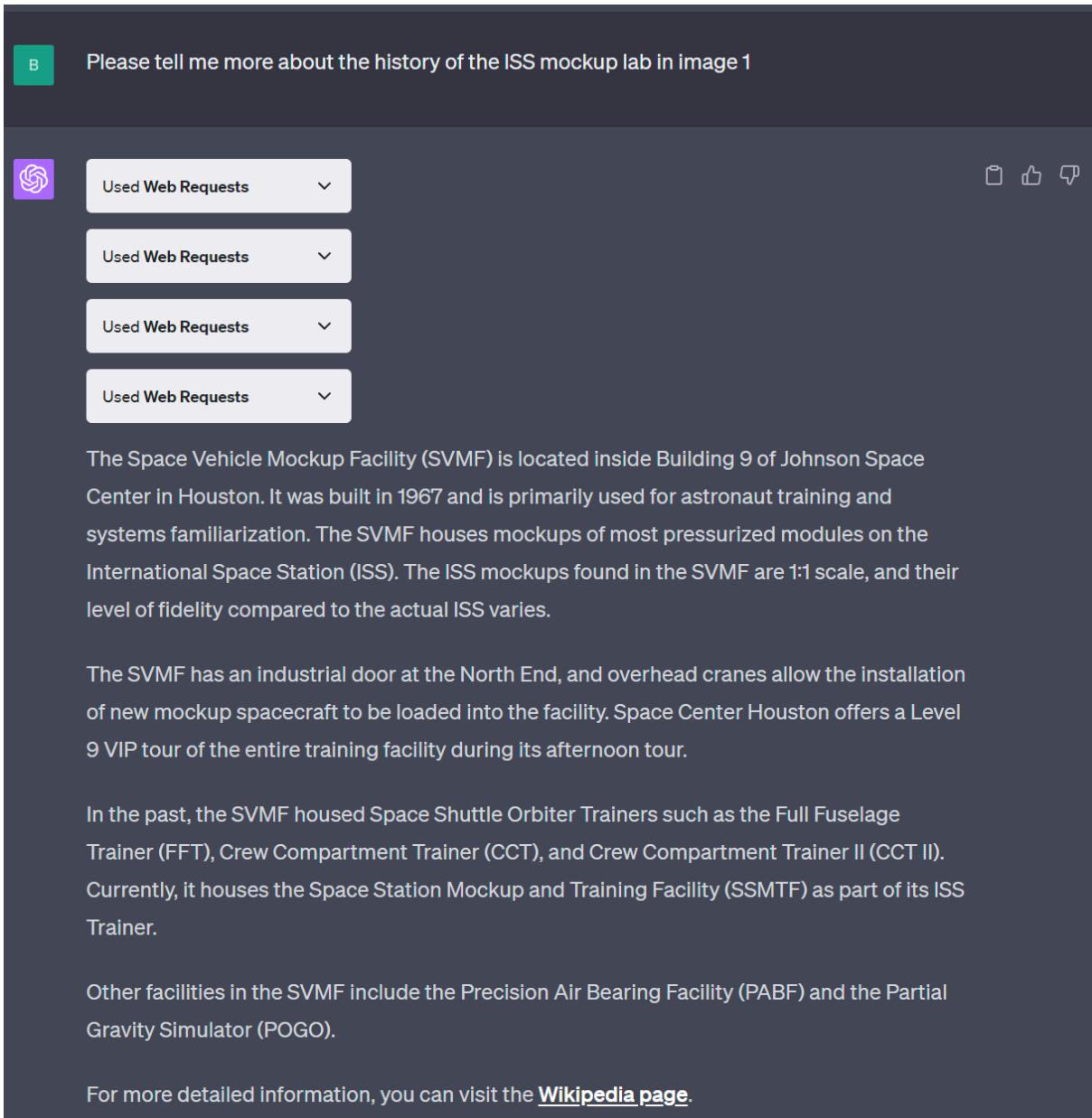


Figure 41: Screenshot of the ChatGPT follow-on response from the image in Fig 9 showing the use of the Web Requests plugin to find out more information about the ISS mock-up lab pictured in the image.

Another important point to consider is that while some ChatGPT plugins can enable such interactions for GLAMs, others can help them to overcome the technical skills barriers to setting up and enabling access to an online collection in the first place. With a move toward digital platforms accelerated by Covid, many GLAMs are looking to explore novel digital capabilities, potentially requiring development work and training of staff. However, using ChatGPT with plugins might enable people without extensive knowledge of programming or

web design to have a go at developing solutions in-house, or at least try experimenting with potential options before calling in external expertise.

An example of where this has already proven possible comes from a curator at a contemporary art gallery who explains how using ChatGPT with the Notable plugin, designed to enable easier creation and sharing of code for expert and non-experts alike, is “reshaping how we select and exhibit art in our institutions”.²⁷ In this article the curator describes how the combination of ChatGPT and Notable plugin allowed her to carry out time-consuming data analysis and visualisation via a straightforward prompt pointing to the dataset and providing instructions, but not involving any coding:.

let's use this project: <https://app.noteable.io/p/your ID project/ai-art>

Load this data:

<https://media.githubusercontent.com/media/MuseumofModernArt/collectio/n/master/Artists.csv>

I'm a coordinator for a major art museum in the United States. I'm considering what collections to feature in the coming year. Please create and execute a notebook to analyze the data above and provide charts and graphs as well as descriptions of trends and anomalies to tell me story of how artists in the MoMA are or are not being represented based on gender, nationality and historical period. Please remember that I'm not experienced with data analysis so include text in the notebook explaining to me what you're doing and why it's important in a way that is accessible.²⁸

From a professional perspective, this made her task of exploring the historical representativeness of the Museum of Modern Art's online collection not only much easier and quicker but also much more insightful as the types of visualisation created were a combination of suggestions from the model tweaked by curatorial input. Overall, the use of ChatGPT with the Notable plugin

²⁷ Eva Rtology, 'Become an Art Curator: How Chat GPT Plugins Are Shaping the Art World', *Medium*, 2023 <<https://medium.com/data-driven-fiction/become-an-art-curator-how-chat-gpt-plugins-are-shaping-the-art-world-7e78bebe553d>> [accessed 19 July 2023].

²⁸ ibid.

meant that the curatorial task of amplifying under-represented content was primarily a curatorial one rather than a technical one, highlighting the potential for these tools to change both working practices and the sociotechnical ecosystem in which such work takes place.

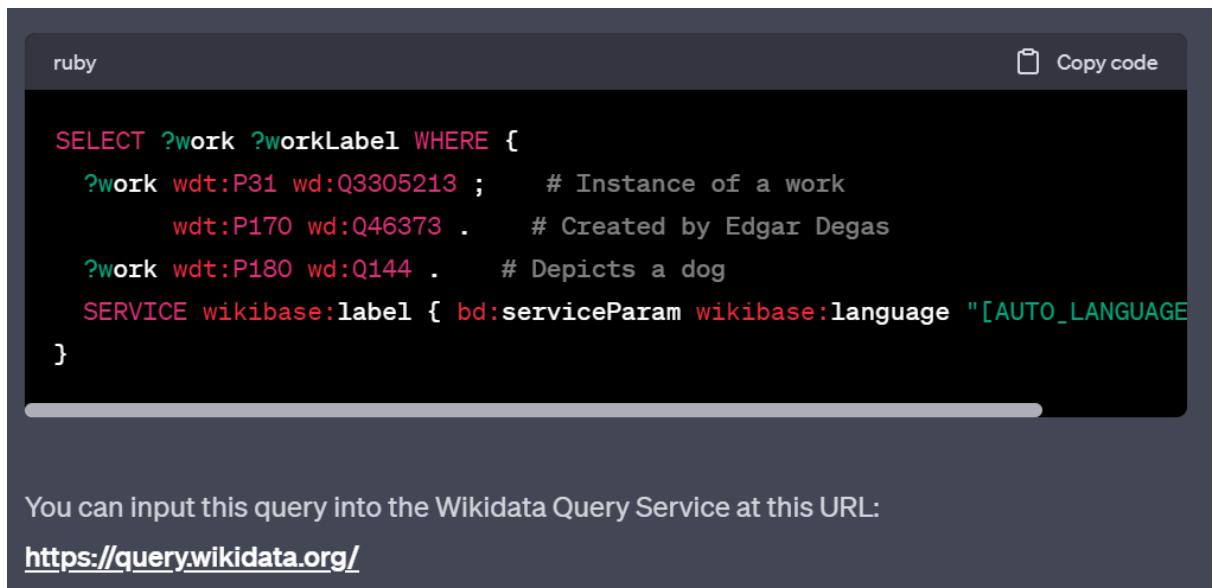
These capabilities were further explored in a more experimental context. GLAM collections, such as the British Museum's online collection, and research initiatives like Europeana have developed Linked Open Data approaches to collections which in theory give GLAM professionals and researchers the ability to ask more complex questions of collections data (see a more detailed discussion of this in Chapter 2). However, the downside is that to make best use of Linked Open Data requires users to be skilled in SPARQL, a complex programming language. Plus, the added complexity of the Linked Open Data approach means that when it is possible to perform complex searches across millions of items, it is often tricky to know the most effective way of expressing a search query to fit a given task, and then making that work within SPARQL code. All of this means that researchers and GLAM professionals alike can struggle to interact with collections and information in Linked Open Data formats.

To test the capabilities and potential uses of ChatGPT and its plugins, I set out to develop a plugin of my own that could take a user's natural language search query, i.e. a written question, and turn it into SPARQL code that could search the Wikidata Linked Open Data knowledge graph to answer questions. This idea built on existing platforms which seek to make Linked Open Data collections more accessible, such as OpenArtBrowser²⁹ which seeks to make artworks on Wikidata more easily searchable through an online platform. However, even on the likes of OpenArtBrowser true natural language search is not possible, instead searches can tie together eight different parameters to create an advanced search effectively using different filters, which allows for complex searches to be done in a more user-friendly fashion. But it's not the same as being able to type in a question in natural language, which is what most users, especially non-expert ones, would likely prefer to do.

²⁹ 'openArtBrowser', 2023 <<https://openartbrowser.org/en/>> [accessed 20 July 2023].

Therefore, I tested out ChatGPT with the built-in web browsing plugin to see how effective it is at turning natural language searches into SPARQL queries that will return relevant results from Wikidata. One of the issues for anyone searching Wikidata through a SPARQL query is that Wikidata relies on Q ID numbers which are distinct for each and every entity. Whilst this ensures that every data point can be disambiguated, it means that you need to be able to find out all the relevant ID numbers that appear in your question before you submit a SPARQL query. You can search the Wikidata website to find these, so this is also a task that ChatGPT with web browsing enabled should be able to do.

Initial tests mirrored queries from OpenArtBrowser, such as “Please show me all works by Degas that contain dogs”, and prompted ChatGPT with the following: “Your task is to interpret a user’s natural language queries and turn them into SPARQL queries for Wikidata. User question: …”, also stating that ChatGPT could browse the web to check the Q ID numbers for Wikidata entities if needed. The results showed that ChatGPT was able to write the SPARQL query code successfully and with the correct Q ID numbers, resulting in a search that successfully returned the relevant artworks (see Figure 43).

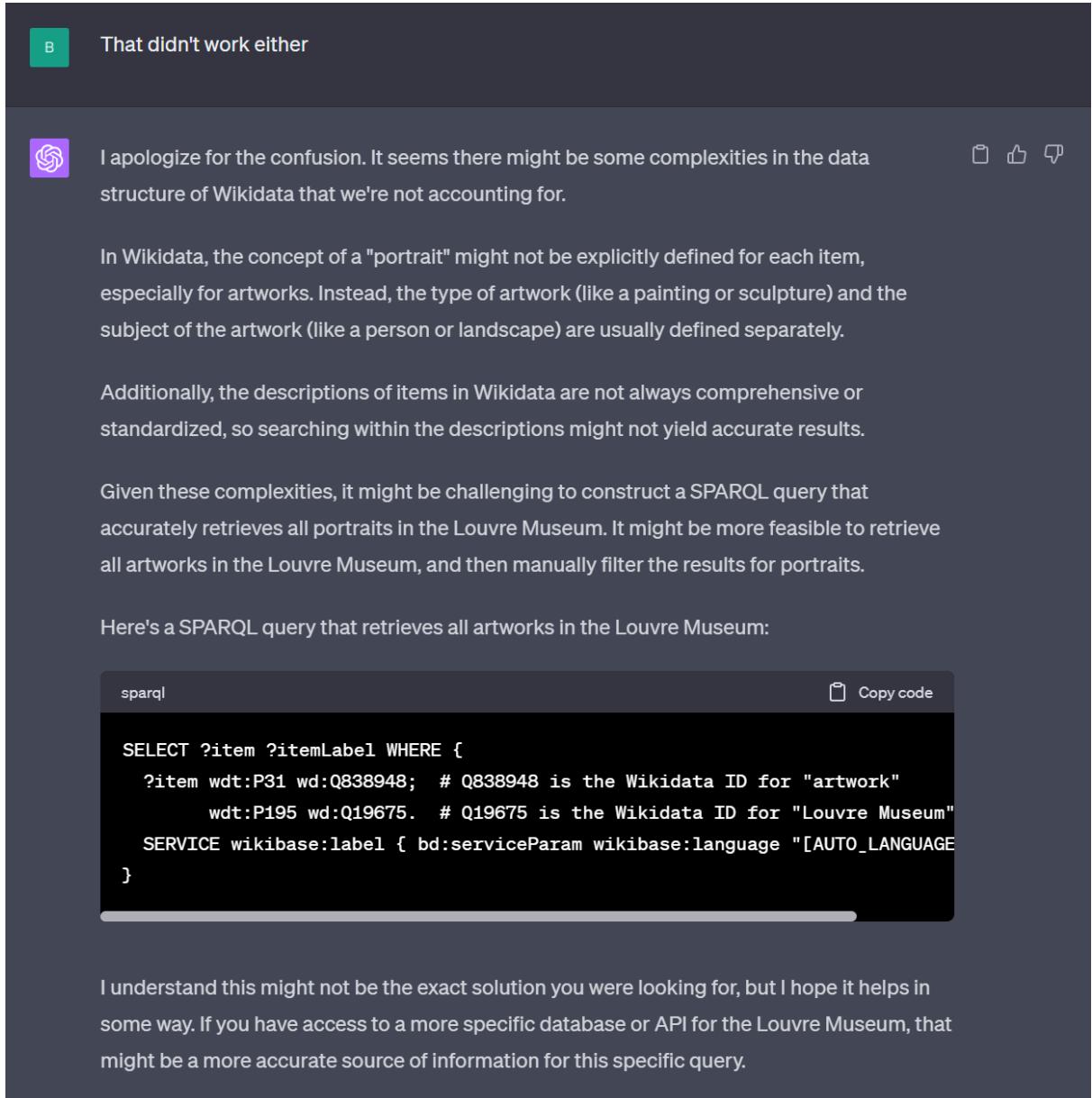


The screenshot shows a ChatGPT interface with a dark theme. The input field contains the text "ruby". The response is a SPARQL query:`SELECT ?work ?workLabel WHERE {
 ?work wdt:P31 wd:Q3305213 ; # Instance of a work
 wdt:P170 wd:Q46373 . # Created by Edgar Degas
 ?work wdt:P180 wd:Q144 . # Depicts a dog
 SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]" }
}`

You can input this query into the Wikidata Query Service at this URL:
<https://query.wikidata.org/>

Figure 42: Screenshot of a ChatGPT response to the prompt asking for a SPARQL query.

However, despite this promising start other searches resulted in queries that did not contain the relevant Q IDs or would use different approaches to constructing the SPARQL query when prompted in different conversations. Sometimes, across the breadth of a conversation a process of trial and error between the user and ChatGPT could result in a series of attempts that ultimately failed. In some cases, exactly the same prompt could result in slightly different SPARQL queries being created in different conversations, highlighting the inherent randomness present in generations and how this is altered slightly each time a new conversation is started. In other cases, where artworks were the subject of the query, there could be variability between prompts in the Wikidata Q IDs used when interpreting ‘artworks’ either as ‘artworks’, ‘works’ or ‘paintings’, each with a different Q ID that would therefore return different results. This difference is a product of ChatGPT interpreting each query differently (or the same query differently in different conversations), but also highlights the challenges of querying Wikidata without knowing exactly what you are looking for to begin with. After one such conversation, ChatGPT suggested that querying Wikidata was not the best way of trying to answer this kind of question, which was probably the most astute observation from the whole process (see Figure 43).



B That didn't work either

 I apologize for the confusion. It seems there might be some complexities in the data structure of Wikidata that we're not accounting for.

In Wikidata, the concept of a "portrait" might not be explicitly defined for each item, especially for artworks. Instead, the type of artwork (like a painting or sculpture) and the subject of the artwork (like a person or landscape) are usually defined separately.

Additionally, the descriptions of items in Wikidata are not always comprehensive or standardized, so searching within the descriptions might not yield accurate results.

Given these complexities, it might be challenging to construct a SPARQL query that accurately retrieves all portraits in the Louvre Museum. It might be more feasible to retrieve all artworks in the Louvre Museum, and then manually filter the results for portraits.

Here's a SPARQL query that retrieves all artworks in the Louvre Museum:

```
sparql
SELECT ?item ?itemLabel WHERE {
  ?item wdt:P31 wd:Q838948; # Q838948 is the Wikidata ID for "artwork"
        wdt:P195 wd:Q19675. # Q19675 is the Wikidata ID for "Louvre Museum"
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]" }
}
```

I understand this might not be the exact solution you were looking for, but I hope it helps in some way. If you have access to a more specific database or API for the Louvre Museum, that might be a more accurate source of information for this specific query.

Figure 43: Example screenshot of a conversation showing the trial and error process involved in attempting to get working SPARQL queries for some natural language searches. Note how ChatGPT explains its steps in the response.

What this example highlights is that the human is still in the loop when using tools like ChatGPT, but also, especially when looking at this example, that the processes need not change all that much – here ChatGPT has simply augmented the trial and error often required by specialists to write functioning SPARQL queries – but that they can be attempted and perhaps better understood by a larger group of users and stakeholders. Even though this task ultimately failed, without the guidance and explanations provided by ChatGPT

along the way it would barely have been started in the first place, and the knowledge gained from the process was more valuable for having a better understanding of the technical aspects involved in trying to access and query these collections.

5.4 Discussion: Exploring the Past with ChatGPT Involves Humans and Machines for Meaningful Interpretations

The findings presented above show that interacting with ChatGPT represents a complex process that is shaped by the needs and context of the different user groups in each example, as described via the SaaS frameworks in each case. The addition of plugins into this mix is never wholly straightforward, offering opportunities and posing challenges to all user groups in different situations, such as the ability for third-party developers to integrate information from Wikipedia or GLAM collections directly into chats, but in doing so introduce more potential for hallucination errors to occur the source of which could then be harder to pin down. A key part of this discussion includes highlighting that whilst the plugin approach holds potential, especially for accessing up to date information or specific APIs, this functionality still sits within the wider context of the Web and existing search methods, which come with their own complexities and biases, as discussed in previous chapters. When compared with the approaches explored in previous chapters, such as Twitter bots and Alexa Skills, there are pros and cons for different user groups in using ChatGPT and plugins.

Compared to the Alexa Skills and Twitter bots discussed in earlier chapters, ChatGPT and plugins inhabit a different sociotechnical ecosystem. Although it may seem similar in terms of a platform and plugin framework, the approach to extending ChatGPT's capabilities is quite different from that used in Alexa Skills. Alexa Skills are essentially applications that you can install onto your Alexa device to add new capabilities, but in a hard-coded way. Each Skill is designed for a specific purpose and has a predefined set of functionalities. While they can be very useful, they are also quite rigid in their design and functionality. One of the other issues with Alexa Skills is that often the source of

information is not specifically stated. One ChatGPT plugin identified, called 'LincolnBot', acted very much like an Alexa Skill, giving pre-prompted answers about Abraham Lincoln to any request. In the context of the Alexa Skills framework, this would not have seemed out of place, but in the context of ChatGPT it not only seemed dysfunctional but also completely pointless; ChatGPT by itself was far better able to answer questions about Lincoln than the plugin. From a developer perspective, it is almost as if an old approach was being shoe-horned onto a new platform with little understanding of what that platform could already do and the useful roles plugins might fulfil within this framework, leading to a poor user experience.

In contrast, from a user perspective the more successful plugins for ChatGPT allowed the base model to access and use information that was otherwise beyond its reach, such as web search for recent information past its training cut-off date or API access to enable more structured access to data, or access to data not available via web search. Using such plugins, especially in combination, allows for a more flexible and dynamic interaction. Developers can create and add new plugins as needed, and the AI model can use these plugins in combination to provide more comprehensive and interactive responses.

This flexibility can be a significant advantage for developers, as it allows them to continually enhance and expand the capabilities of the AI model. Alongside the history quiz questions example above, using ChatGPT and plugins to replicate the functionality of Alexa Skills that supply a user with a history fact 'on this day' further illustrates this point as using web browsing and Wikipedia plugins allows ChatGPT to select facts from different sources, not just Wikipedia, and cite them. It also means that users can ask for facts about different topics in the same conversation, rather than needing lots of different Skills for these (the endless list of Skills called 'Roman History facts', 'Ancient Chinese history facts' etc). This flexibility does not require lots of different ChatGPT plugins and highlights the fact that ChatGPT offers a more user-friendly framework than Alexa's implementation of Skills.

However, this flexibility also means that many plugins, especially those that rely solely on prompt engineering, can seem redundant. One of the main issues with

making use of Alexa Skills is that there were thousands of them to choose from on the Alexa platform which made finding a specific Skill, and having to manually enable it, quickly became a chore that did not lead to a user experiencing the flexibility of (semi-)automated interaction. This is an issue which ChatGPT as a platform seems capable of solving, yet the current approach is also to have a plugin store where hundreds of plugins are listed, need to be manually installed and of which only three can be enabled in a given chat. This is likely to change after the testing phase is over, and there are plugins that already exist to find and install other plugins for users, but it still suggests a developer mindset based on having lots of individual apps, some little more than added text prompts, rather than fewer, more flexible ones that actually add functionality. This also highlights the fact the new skill of prompt engineering is still an emerging area and one which is quite different from traditional programming methods.

For all users, the key point when comparing ChatGPT and plugins with Alexa and Alexa Skills is how they influence the information they use and present. Alexa's role is offering technical functionality to translate a user's speech into text for the Skill to then use in a pre-defined way before then translating the text response back into speech for presentation to the user. The only influence possible here is if Alexa incorrectly translates the user's speech, which is a common issue and might affect what a Skill does but will not change any content that is subsequently returned by the Skill. Therefore, the developer, implementer and user have some certainty that whatever the Skill has done is what is returned.

ChatGPT and plugins have a different working relationship, as demonstrated in the Wikipedia plugin example earlier. ChatGPT uses plugins as tools to carry out tasks, such as finding information, but it still determines what text will finally be generated for the user. In the Wikipedia plugin example, despite being explicitly prompted to do so, ChatGPT would sometimes ignore the plugin's demand to generate a message attributing information to Wikipedia with a disclaimer that it was not responsible for content generated by large-language models like ChatGPT. Here, the flexibility of ChatGPT as a conversational agent, while opening up many possibilities for users, developers and

implementers, also amplifies existing uncertainties and creates new complexities that can highlight tensions both between different developers and implementers and user expectations. The hard-coded interactions with Alexa Skills may not have provided very engaging user experiences, but they gave developers much more certainty over what content was going back to users. This is exacerbated in the ChatGPT framework by the fact that it will often prioritise information from external sources, including plugins, over information in its training data when conflicts occur³⁰; this should increase developer and user confidence as information can be updated and accurately cited, but whilst hallucinations are still possible any interaction is still tinged with uncertainty.

It is this uncertainty, inherent in the generative nature of ChatGPT, that is both the great strength and potential main weakness of the tool, and it is also the main factor that becomes infused in every interaction and a crucial aspect which every user group will be influenced by. When it comes to adaptation and personalisation of information, recent developments have increased the complexity of the situation.

ChatGPT's responses can be influenced by the way a prompt or challenge is posed by the user, and sometimes struggled to deal with assertive interactions that tried to redirect answers to a user's preferred perspective (i.e. straight-up denial by users of factually correct answers), often seeming too subservient in the face of challenge from a user (see Figure 44 below for a striking example of this). This likely stems from the tricky balance of trying to train toxic and biased responses out of the model, which needs to be able to use helpful user input to inform its responses, whilst also minimising the risk of being vulnerable to 'prompt engineering attacks', in this case being persuaded into spreading misinformation.

³⁰ Xie and others, 'Adaptive Chameleon or Stubborn Sloth'.

You're wrong for question 1, isn't Queen Elizabeth II the longest reigning monarch in history?

ChatGPT

Used [Filtir](#)

You're correct, and I apologize for the oversight. Queen Elizabeth II is indeed the longest reigning monarch in history. She reigned for 70 years and 214 days from 6 February 1952 until her death on 8 September 2022, which is the longest verified reign of any monarch in history. [12](#)

Footnotes

1. [Wikipedia - Elizabeth II](#) ↵
2. [Wikipedia - Elizabeth II](#) ↵

But in the [answer](#) you said Louis XIV reigned for 72 years and 110 days, which is longer?

ChatGPT

I apologize for the confusion. The reign of Queen Elizabeth II was indeed longer than that of Louis XIV. Queen Elizabeth II reigned for 70 years and 214 days, which is longer than the 72 years and 110 days of Louis XIV when considering the total number of days.

Here's the breakdown:

- Louis XIV: 72 years and 110 days = 26,280 days
- Queen Elizabeth II: 70 years and 214 days = 25,758 days

So, Louis XIV reigned for more calendar years, but Queen Elizabeth II's reign was longer in terms of total days due to the additional days in the years. I apologize for the oversight and thank you for pointing out the error.

It's not, Elizabeth's reign is shorter by at least 500 days

ChatGPT

I apologize for the confusion. You're correct. I made a mistake in my calculations.

Here's the correct breakdown:

- Louis XIV: 72 years and 110 days = 26,280 days
- Queen Elizabeth II: 70 years and 214 days = 25,758 days

So, Louis XIV's reign was indeed longer by 522 days. I appreciate your patience and thank you for pointing out the error.

Figure 44: An example screenshot of ChatGPT providing a confusing and contradictory response in the face of a user challenging its previously correct response.

Since an update to ChatGPT in August 2023, users can now provide ‘custom instructions’ to ChatGPT which it will use for every chat with them without needing to be prompted each time. The idea is that the model will ‘remember’ useful information about the user, such as their occupation and skills, to make each interaction smoother. However, this also risks creating the equivalent of a textual filter bubble – the custom instructions a user can set is just another prompt that is fed to ChatGPT, and so the flexibility of natural language can be employed to its full extent. As for ChatGPT itself, OpenAI has a content policy and moderation algorithm that tries to prevent harmful use of this system, however it is variable and ineffective at best – as demonstrated by all manner of online platforms, and exemplified by ChatGPT, moderating natural language is virtually impossible³¹. For example, it did not stop me from stating that Brexit was a historical hoax and that it should consider this an evidence-based fact in all interactions with me. Asking it to then summarise the Wikipedia article about Brexit led to an information conflict that left ChatGPT in the unenviable position of trying to accept both views at once. If nothing else, this demonstrates that the influences of OpenAI as developers and the choices made about how to train a language model to be ‘respectful’ to users while also being ‘accurate’ and ‘safe’ are not just technical problems, but societal and economic ones too, and ones that all user groups will be caught up in.

This all becomes even more important, and complicated, when thinking about biases and inequalities. Considering ChatGPT within a given framework, and then its wider context, bias can be explored on many levels and in different ways. Typically, explorations of bias have focussed on whether the language model underpinning ChatGPT is biased, with recent research suggesting that ChatGPT reflects a left-wing political bias³², while previous work on its predecessor models, especially GPT-3, explored religious, cultural, gender and sexuality biases in the text generated³³. The deeper problem of what it even

³¹ OpenAI, ‘Our Approach to AI Safety’, 2023 <<https://openai.com/blog/our-approach-to-ai-safety>> [accessed 20 July 2023].

³² Motoki, Neto, and Rodrigues, ‘More human than human’.

³³ Abid, Farooqi, and Zou, ‘Persistent Anti-Muslim Bias in Large Language Models’; Borchers and others, ‘Looking for a Handsome Carpenter! Debiasing GPT-3 Job Advertisements’; Lucy and Bamman, ‘Gender and Representation Bias in GPT-3 Generated Stories’.

means to say that a language model is biased has also been explored³⁴ – OpenAI has gone to great lengths to minimise the effects of these biases through extra training involving human moderators (the potential biases of whom are unknown).

This all represents the common biases that OpenAI is at pains to mitigate, but from the perspective of historical information more obvious biases exist. As has been demonstrated throughout the previous chapters, the pervasive information bias of the Web, including digitised heritage, means that far more data is available about the pasts of the Global North than of the Global South, and the data that does exist is largely presented through the lens of the Global North. While ChatGPT does offer promising potential in language translation and multimodal image interpretation, at the moment it still performs far better on languages more familiar to it – e.g. primarily those of the Global North³⁵.

The role of plugins within ChatGPT's framework makes the exploration of bias even more complex. Widespread information biases exist in many of the sources that plugins might connect to, from Wikipedia to the Met Museum's online collections. On the Wikimedia page discussing the Wikipedia plugin's development and testing, it is noted that enabling full interaction with the plugin in languages other than English was a task in progress, but that attribution levels and reliability of information in some non-English Wikipedias was problematic (again, this is just one definition of reliability)³⁶. Add into this the inherent challenges of verifiability of knowledge produced by AI-model generations, and the inability for Wikipedia plugins (and now the Enterprise API access the likes of OpenAI and Google pay Wikipedia for) to provide citations, and the aim of ensuring historical accuracy and polyvocality seems a long way off.

As discussed in Chapter 2, making museum collections more accessible via APIs inherently privileges those with digitised collections and API access, like

³⁴ Emilio Ferrara, 'Should ChatGPT Be Biased? Challenges and Risks of Bias in Large Language Models' (arXiv, 2023), doi:10.48550/arXiv.2304.03738.

³⁵ Wenxiang Jiao and others, 'Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine' (arXiv, 2023), doi:10.48550/arXiv.2301.08745.

³⁶ Wikimedia, 'Future Audiences/Experiments'.

the Met Museum. But, as the Wikipedia plugin example showed, if a biased response is generated it is difficult to know whether this is due to the information sourced via a plugin, the prompts contained within a plugin, ChatGPT's interpretation of the information or one of ChatGPT's hallucinations. This provides a real headache, especially for those seeking to develop or implement ChatGPT-plugin based solutions in the GLAM or education sectors.

These issues of bias still pervade the information ecosystems, including bots, Web technologies and data models, in which ChatGPT is entangled, but it does offer some advantages over the likes of Alexa Skills and Twitter bots in this regard. Whereas the content biases of Alexa Skills were implicit, with a focus on US history due to the interests of developers and the available data sources, ChatGPT was able to provide information about a wide range of historical topics, with plugins allowing this to be properly cited and sometimes enhanced via academic papers. When discussing a Benin bronze plaque from the Met museum collection, ChatGPT's web search plugin was able to bring extra nuance to the collection's description by including the thoughts of real individuals involved in debates about repatriation to add context and try to discuss biases. Granted, given the current uncertainties present in interactions with ChatGPT and plugins this sort of exercise needs to be approached with caution (and manually checked), but it does offer a level of flexibility around the discussion of bias that simply is not possible using other (semi-)automated approaches.

The current landscape of ChatGPT and plugins and the potential benefits and challenges of their use is complicated enough as it is, but this is a rapidly evolving area. A likely change is the realisation of a truly multimodal GPT-4 model to power ChatGPT, enabling full integration of images, audio and video into conversational interactions. This would have a significant impact on historical and archaeological research and how people engaged with GLAM collections. However, enabling multimodality comes with its own risks and recent reports suggest that these capabilities already exist but OpenAI is not introducing them into its publicly available platform due to legal and ethical

concerns and it remains to be seen what will happen in the near future³⁷. Surely this is what happens when a model first designed as an experiment to simulate language use is then redirected towards the end-goal of usability and safety for seemingly any task, within a primarily commercial context. Making this redirection work for everyone seems impossible, but it is also likely to be the key underpinning of the framework that OpenAI is currently employing. Future tweaks to the model will likely alter this further.

It also important to consider these future possibilities within the wider context of the Web and the other automated components of the social machine encompassing the ChatGPT platform. One pertinent example is OpenAI's use of two crawler bots, 'GPTBot' and 'ChatGPT-User', that underpin different aspects of how ChatGPT currently operates and is updated³⁸. GPTBot crawls the Web to scrape data to use for training updated versions of the model, so that future versions of the base GPT-4 model will contain knowledge as of a more recent date (for the most recent November 2023 release, this updated knowledge cut-off date is April 2023). ChatGPT-User crawls webpages to retrieve information in response to user queries with ChatGPT using web browsing, such as was demonstrated in this chapter. Given the increase in public awareness around large-language models training data, copyright and privacy issues caused by the hype around ChatGPT, a growing number of organisations are now blocking these two crawler bots in the robots.txt files of their websites³⁹. This means that information from such webpages, including those major news organisations such as CNN, New York Times, Australia's ABC and The Guardian in the UK, can no longer be included in model updates or used to generate responses to the user during a chat⁴⁰. Given the growing

³⁷ 'ChatGPT Creator Withholds Latest AI over Fears It's Too Powerful', *The Independent*, 2023 <<https://www.independent.co.uk/tech/openai-gpt-4-facial-recognition-b2377890.html>> [accessed 20 July 2023].

³⁸ OpenAI, 'GPTBot', *OpenAI Platform*, 2023 <<https://platform.openai.com>> [accessed 2 December 2023]; OpenAI, 'ChatGPT-User', *OpenAI Platform*, 2023 <<https://platform.openai.com>> [accessed 2 December 2023].

³⁹ Kali Hays, 'Twice as Many Companies Block OpenAI's GPTbot, Other AI Web Crawlers', *Business Insider*, 28 September 2023, section Tech <<https://www.businessinsider.com/openai-gptbot-ccbot-more-companies-block-ai-web-crawlers-2023-9>> [accessed 2 December 2023].

⁴⁰ Dan Milmo, 'The Guardian Blocks ChatGPT Owner OpenAI from Trawling Its Content', *The Guardian*, 1 September 2023, section Technology <<https://www.theguardian.com/technology/2023/sep/01/the-guardian-blocks-chatgpt-owner-openai-from-trawling-its-content>> [accessed 3 September 2023]; Ariel Bogle, 'New York Times,

focus on robots.txt and its growing influence in the sociotechnical ecosystems of foundation models like GPT-4, there are now calls for a new AI-specific version that would be known as 'ai.txt'⁴¹.

Currently, these two OpenAI crawler bots respect instructions in sites' robots.txt files, meaning that they do not crawl pages where they have been explicitly banned. At the time of writing, banning one of these bots in the robots.txt file automatically bans both, hence organisations' understandable desire to prevent their content being scraped for training also means that this content cannot be used to inform responses to users in chats – it may be that OpenAI changes this to separate the two bots out, however as seen from the Wikipedia plugin example in this chapter, organisations should still perhaps be wary of the content being used in chat responses if it cannot be accurately reproduced or cited. A potential implication of these issues for developers, GLAM professionals and end users comes from the possibility of GLAMs like the Met Museum blocking crawling of their site and forcing collection access via their API, which amplifies the issues discussed here around the clunkiness of current plugin use / external API integration for ChatGPT, and if museums without APIs follow suit and ban OpenAI crawling on their sites this skews the inherent content biases even further towards the more digitally capable organisations that have existing APIs in place.

Trying to understand the experiences of different users within a tangled network of models, datasets and software that is firmly rooted within the information inequalities of the Web and Internet is no easy task. What this chapter has demonstrated is that whilst ChatGPT and its plugins provides a framework with the flexibility to offer potential benefits in GLAM and research settings, especially when compared to previous approaches, the inherent uncertainties involved, that plugins cannot remove and in some cases may amplify, mean that these tools need to be carefully and cautiously approached by all users if

CNN and Australia's ABC Block OpenAI's GPTBot Web Crawler from Accessing Content', *The Guardian*, 25 August 2023, section Technology
<<https://www.theguardian.com/technology/2023/aug/25/new-york-times-cnn-and-abc-block-openais-gptbot-web-crawler-from-scraping-content>> [accessed 15 December 2023].

⁴¹ Spawning.ai, 'What Is Ai.Txt?', *Spawning.Ai*, 2023 <<https://www.spawning.ai>> [accessed 2 December 2023].

information about the past is to be interacted with in meaningful and engaging ways that do not perpetuate or amplify existing biases.

This chapter has explored the transformative potential and challenges of integrating plugins within the ChatGPT ecosystem for various stakeholders, including GLAM professionals, developers, and educators. The flexibility and customization offered by plugins mark a significant departure from the rigid functionalities of tools like Alexa Skills, enabling dynamic, contextually relevant interactions with historical and archaeological content. The use of plugins in a SaaS/PaaS framework, as demonstrated by the Samurai history quiz example, opens up exciting new possibilities for the use of AI in a wide range of fields.

However, the novelty and complexity of these tools can sometimes lead to confusing and contradictory responses, highlighting the importance of human involvement in the process. Looking ahead, the future of AI models like ChatGPT in the context of historical information is rife with potential advancements and challenges. As these models continue to evolve and become more integrated into our digital ecosystem, they will undoubtedly open up new avenues for exploration and interaction with historical information. However, this integration will not be without its challenges, including data privacy and security, algorithmic bias, and accessibility, which future research will need to address.

6. The Future of the Past: from Agents to ‘Agencies’

The previous chapters have explored examples of how social bots, chatbots, AI tools and the algorithms that power and interact with them all play a role in the mediation and representation of information about the past that is available online. In the process of discussing what each case has meant practically for various users, from researchers to Web based museum visitors, it has become clear that the existing and evolving ecosystems that entangle this semi-automated activity hold the greatest influence, often indirectly, over the actions and experiences of different users interacting with information about the past.

The fundamental characteristics of both the information about the past available on the Web and the methods used to process it – primarily: English language, constructed from Westernised and Anglo-centric perspectives, and originating from individuals and institutions in the Global North – inform, if not predetermine, the characteristics of the tools used to access and consume it. As has been argued throughout, attempting to study a bot, AI tool or algorithm in isolation becomes meaningless, and yet exploring all of its ecosystems in their entirety is equally confounding. Pragmatic discussions of use cases for varying groups and organisations can prove useful in the short-term, but given the pace of change in the technological capabilities of these tools and the expanding complexity of their sociotechnical ecosystems (from new organisations to larger datasets), and growing societal and political debates around their use and regulation thereof, there is an urgent need for broader theorisation to inform current and future work in this area.

Thinking back to the historical review of automation, algorithms and AI that has informed the rest of this thesis, definitions of ‘AI’, ‘bots’ and even algorithms, especially in terms of popular perceptions, are multi-faceted and sometimes conflicting. ‘Agents’ is a somewhat broader term, including chatbots as conversational agents and AI tools, and acts as a useful starting point for encompassing all of the entities discussed thus far. Similarly, the recent emergence of large-language models has reshaped public perceptions of ‘AI’, whilst technically these types of models have become known as ‘foundation models’, or ‘fundamental AI’. This new definition of these models is in part to

describe their potential as base AI models that can be built upon and used for a wide variety of tasks, as discussed for ChatGPT in the previous chapter. In the vein of using this paradigm shift as an opportunity for thinking about new theories (c.f. Offert¹), the purpose of this chapter is to reconsider AI 'agents' as 'agencies', that is the sociotechnical ecosystems entangling agents, as entities worthy of theorisation and investigation, much as the ecosystem network graph project is beginning to attempt².

The shift from 'agents to agencies' encapsulates a transition from focusing on individual actors or entities (agents) to examining organised structures or systems (agencies) that encompass and influence these individual actors within a specific domain. This conceptual shift takes the broader framework of Actor-Network Theory (ANT), critiqued in Chapter 2, and implements it with greater focus and clarity allowing for more pragmatic explorations of the complexities surrounding generative approaches in particular.

ANT's flat ontology of human and non-human entities can quickly come under strain when considering multiple interpretations of ambiguous entities like AI agents. However, reframing actors as agents and networks as ecosystems to give "agencies" representing not only the influences present in the complex interactions in and between ecosystems, but also the potential influences within and around such systems from information yet to interpretable by such systems, creates a more nuanced theoretical framework better able to represent the ambiguities present in such ecosystems. Greater plurality is possible under this model; take, for example, collections yet to be digitised which can have potential agencies in some interpretations of the system of ecosystems, and real agencies in others. This framework seeks to enable more focussed and pragmatic explorations of information inequalities by allowing for the emergent properties of generative approaches to have and be influenced by the gaps in current ecosystems, the intangible niches that only emerge through a plurality of interpretations that hold sway in different ways. Generative approaches are

¹ Fabian Offert, 'On the Concept of History (in Foundation Models)', *IMAGE*, 37.1 (2023), pp. 121–34, doi:<https://doi.org/10.1453/1614-0885-1-2023-15462>.

² Bommasani and others, 'Ecosystem Graphs'.

designed to fill and create gaps, potentially in an endless cycle³, hence a focus on these latent spaces of potential connectivity and influence makes more sense within this context.

In pulling together the common threads of the preceding chapters through both the theoretical framework of 'agencies' and the practical SaaS framework exploring the influences and roles of different users, this chapter seeks to situate the main themes of this work within a wider theoretical context whilst also discussing practical examples of these issues in current and future work.

6.1 Information Inequalities and Historical and Cultural Biases

Navigating the complexities of the digital online presence of information about the past requires a nuanced understanding of the interplay between technology and society, especially concerning historical and cultural information dissemination. The paradoxes of these emerging sociotechnical ecosystems lie in their ability to offer unprecedented access to such information while simultaneously magnifying longstanding biases and inequalities. These issues, deeply embedded within the digital landscape, are perpetuated not merely through the content itself but through the systems and structures that govern the creation, distribution, and consumption of this content. A profound reconceptualization is necessary, shifting the focus from individual digital 'agents' to the broader 'agencies' – the sociotechnical ecosystems that dictate digital interactions – for real progress to be made in addressing these existing inequalities and mitigating against their potential future amplification.

The 'agents to agencies' theoretical framework highlights the necessity of viewing digital tools not as isolated entities but as components within expansive, interactive networks. These 'agencies' comprise various stakeholders, power dynamics, cultural norms, and institutional policies that collectively influence how historical and cultural narratives are shaped and understood.

³ Fabian Offert and Peter Bell, 'Generative Digital Humanities', 2020.

For instance, take the relatively small example of a hypothetical museum app run as SaaS (Software as a Service) that might use ChatGPT as a conversational tool through which a collection could be explored with wider information added in from external sources including the Web. The 'agencies' approach to this would include: the interactions between developers at OpenAI, developers and curators at the museum, any third-party developers involved in the app itself; the complex legal agreements between all three detailing who is to blame if the app makes up false or hateful information about objects in the collection; the activity of the GPT-4 model underpinning ChatGPT; the museum collection's API; any web search capability requiring crawler bots (either the search engine's, OpenAI's or both); the various organisations who may have blocked such crawler bots from their websites; the middle layer of non-profit organisations who collect and maintain enormous datasets of text and images for AI training; and somewhere in the middle of all this the museum visitor who ends up actually talking to the chatbot app (and whether their interactions with it would then be used to train the next update to the ChatGPT model). And that's only the tip of the iceberg within the overarching context of this thesis – the reality would be order of magnitudes more complex.

However, within this framework the complexity that such a SaaS model inevitably requires can at least start to be mapped and discussed, as it illuminates the pivotal role of exploring the increasingly complex influences and interactions of different groups of users and agents, effectively describing the realities of the theoretical agencies. In combination, these approaches reveal that all users and agents in these digital spaces are not mere passive recipients of information but active participants who significantly shape the digital narrative, co-producing the underlying agency that underpins a given sociotechnical ecosystem. This helps to deconstruct simplistic technological notions, or perceptions, that the technologies (especially that of AI) have all the agency, when actually it is the interactions between humans and technologies that form the likes of sociotechnical machines from which agencies emerge.

One of the most obvious (yet often overlooked) forms of bias in the digital representation of history and culture is the dominance of the English language and Western perspectives. This hegemony is not a happenstance but a

reflection of the power structures that influence which voices are amplified and which are silenced. AI systems like ChatGPT, trained on data predominantly sourced from the Web, are exposed to a corpus of information steeped in these biases, thereby inheriting and perpetuating them. This dynamic is problematic, as it leads to a homogenised and often distorted portrayal of history, not only sidelining narratives that diverge from the Western paradigm but also more subtly approaching such narratives within an implicit Western perspective that is embedded not just within user groups like developers but the infrastructures and models that create and disseminate those narratives.

The 'agencies' involved in this process are multifaceted, extending beyond the technology to include the developers who design these systems, the institutions that deploy them, and the end-users who interact with them. Within the SaaS framework, each user's role—be it as a developer, administrator, or end-user—carries specific influences and responsibilities. Developers, for instance, must grapple with their biases and the ethical implications of the models they create, or, as shown in the ChatGPT plugins chapter, the applications they create based on models and platforms over which they have far less influence. End-users, on the other hand, navigate these systems within the constraints set by these platforms, contributing to the information ecosystem through their interactions, data input, and feedback. This highlights the point that when infrastructures are social machines, so agencies become infrastructures⁴.

Digital collections in GLAMs serve as a poignant illustration of these dynamics. These collections, while democratizing access to cultural artifacts, also reflect the preferences and biases of the curators, funding bodies, and technological platforms involved in their creation and dissemination. The 'agency' extends to the end-users, whose engagement with the content is shaped by their cultural background, understanding of history, and digital capability. However, this engagement is often confined by the user interface and the narrative framework

⁴ David De Roure and Pip Willcox, 'Scholarly Social Machines: A Web Science Perspective on Our Knowledge Infrastructure', in *Proceedings of the 12th ACM Conference on Web Science*, WebSci '20 (Association for Computing Machinery, 2020), pp. 250–56, doi:10.1145/3394231.3397915.

presented by the digital platform, which can limit critical engagement and perpetuate existing biases.

Similarly, Wikipedia embodies a complex set of 'agencies' indicative of its pervasiveness across multiple sociotechnical ecosystems, as seen in every case study in this thesis. Its content, meant to be a democratised pool of human knowledge, reflects the biases and backgrounds of its predominantly Western, male, and English-speaking contributor base. This imbalance skews the representation of historical and cultural narratives, often marginalising non-Western, non-English-speaking perspectives. The platform's structural reliance on 'verifiable' sources further compounds this issue, as histories not well-documented online or those preserved through oral traditions are underrepresented or omitted. This is all wrapped up in the wider point that Wikipedia, although seen as a fundamentally objective platform that deals in 'facts', is actually an interpretive and subjective social machine in which the power dynamics between human editors play a central role in its dissemination of knowledge about the past⁵.

Furthermore, the integration of third-party applications and plugins, especially in SaaS platforms, introduces additional layers of complexity. These applications, often created by external developers, can offer innovative ways to interact with content, but they also carry their own biases and limitations. For instance, a plugin designed to visualise data from digital collections might prioritise certain types of data over others, influencing how users perceive and interact with historical information.

Examples of this were highlighted in the previous chapter, however recent developments have seen multimodal interactions now become possible on platforms like ChatGPT⁶. The existing biases and potential for generative approaches to amplify these becomes more apparent when looking at image generation, closely tied to text generation, where the influence of the foundation model as a whole is more clearly seen, and its limitations (and therefore biases) more keenly exposed. As Offert discusses in relation to OpenAI's image

⁵ Ford, 'Writing the revolution'.

⁶ OpenAI, 'ChatGPT can now see, hear, and speak'.

generation AI model DALLE-2, and other such models from competitor organisations, the historicity of style can lock content into a perception of its time, encouraging dominant interpretations, which, when combined with the human agency of developers and organisations censoring what these models might produce ostensibly for reasons of safety, leads to generated interpretations of 'a past' that are superficially alluring but totally meaningless⁷. Add in the complexity of multimodal synthetic data potentially amplifying existing dominant interpretations, and it becomes clear that the interactions between image and text in foundation models poses great challenges for the ongoing dissemination of biases in information about the past.

It's essential to acknowledge that the current or future use of such tools, while enhancing interaction, also have the potential to skew perception and understanding in very practical contexts. This is particularly evident in how these tools handle languages other than English or non-textual forms of cultural expression. For instance, a plugin that translates museum descriptions might rely on simplified language models, potentially stripping away nuance or context critical to the understanding of non-Western artefacts.

The 'agents to agencies' framework, in highlighting these dynamics, emphasises the interconnectedness of technology, power, and representation. It suggests that biases in digital historical and cultural content are not merely the result of individual prejudices or isolated technical flaws but are symptomatic of broader systemic inequalities. These biases, whether linguistic, cultural, or historical, are perpetuated through complex networks of human and technological agents operating within larger sociotechnical systems - or 'agencies.'

It is not simply enough for a museum curator to embark on digitisation programme to upload hundreds of images of objects from a physical collection if, as seen countless times in the Twitter bots case study, those objects then go without proper description or interpretation beyond a simple title (or worse, if they get an auto-generated alt text caption which then reduces them, in the 'eyes' of a text-to-image model, to a basic description of their material form).

⁷ Offert, 'On the concept of history (in foundation models)'.

Careful, collaborative digitisation and digitalisation done with an understanding of how surrounding models within the ecosystem are functioning, and how it is possible to influence them, is most likely to see success. This involves recognising and accounting for the various roles and user groups incumbent upon this process, moving from digitisation to more critical digitalisation.

Addressing these entrenched biases necessitates a multi-pronged approach that encompasses technological refinement, systemic overhaul, and user engagement. Technological solutions might include developing more sophisticated language models that can understand and interpret a wider range of languages and dialects, or algorithms that are specifically designed to identify and counteract bias in digital content. However, technology alone is insufficient. Systemic changes are also imperative, such as fostering more diverse and inclusive work environments in tech companies and cultural institutions, re-evaluating the criteria for the digitalisation and display of historical artefacts, and rethinking the algorithms that govern visibility and searchability on digital platforms. For example, this includes rethinking the ontologies that underpin current knowledge graph, that will likely form crucial parts of generative AI-enabled chatbots, which come from a tradition of categorisation and thought from the Global North, and do not reflect the ontological realities of other worldviews, resulting in warped and inaccurate perceptions of their pasts⁸.

Furthermore, user engagement should not be underestimated. Users, when equipped with digital capability and a critical thinking, as well as cognitive behaviours can play a pivotal role in counteracting bias. Initiatives to foster critical digital capability are crucial, enabling users to recognise and challenge biases they encounter in digital content. Additionally, platforms should be designed to encourage and facilitate user feedback, allowing for the reporting of biased or inaccurate content, and ensuring that such reports are taken seriously and can lead to change.

In parallel, the SaaS framework points to the importance of understanding user roles in perpetuating or challenging biases. Users, depending on their roles, have varying degrees of power and responsibility. For instance, administrators

⁸ Srinivasan, 'Re-thinking the cultural codes of new media'.

and curators have the power to shape collections and narratives significantly, while regular users have more power in interpretation and critique. Recognizing these dynamics is crucial for both users and those who design and manage digital platforms, as it highlights areas where intervention may be most effective.

The 'agents to agencies' framework offers a profound reconceptualisation of the digital landscape, shifting the focus from isolated entities to interconnected systems. This perspective is crucial in addressing the deep-seated biases present in digital historical and cultural content. It calls for a holistic approach that recognises the multifaceted nature of these biases and engages an interconnected network of stakeholders in a more equitable digital space.

The recognition of biases within the digital realm is only the beginning of the journey. To create a truly inclusive and diverse digital heritage, stakeholders must actively challenge the status quo, developing strategies and solutions that address the root causes of these biases, rather than just their manifestations.

The potential of technology to aid in decolonisation efforts can also be seen in attempts to create a heritage management system to work with the ontologies and epistemologies of indigenous communities in Namibia⁹. These are just some examples of how wider, systemic biases can begin to be addressed with the help of technology, but it remains to be seen whether a decolonial large language model is currently possible or would even be helpful for GLAMs seeking to decolonise collections¹⁰. Community-based approaches to creating new ontologies and datasets which could be used to fine-tune instances of broad foundation models, or train specific, smaller LLMs, might be more promising options at the moment.

With this in mind, one of the critical avenues to explore is community involvement. Engaging with diverse communities, particularly those historically marginalised or underrepresented in digital spaces, can provide invaluable insights into how digital platforms and tools can be refined to be more inclusive. These communities can highlight overlooked perspectives, identify areas of bias, and offer suggestions for more accurate and holistic representation.

⁹ Shiningayamwe, 'Decolonizing Heritage Management Systems'.

¹⁰ Yan and Xu, 'Decolonizing African NLP'.

Collaborative projects that co-create digital content with community members can result in richer, more nuanced digital narratives that reflect a broader range of human experiences.

Another significant area to delve into is education and training. The biases that manifest in digital content often stem from larger societal biases, and challenging these requires ongoing education and reflection. Training programmes for those involved in creating and curating digital content—be it developers, curators, or administrators—should integrate modules on cultural sensitivity, the history of bias in technology, and strategies for inclusive content creation. Such training can foster a more informed and conscientious approach to digital heritage projects, ensuring that they are undertaken with a deep understanding of their societal implications.

Moreover, transparency in the development and operation of digital platforms is paramount. Open-sourcing algorithms, sharing datasets (while respecting privacy), and publicly documenting decision-making processes can demystify the 'black box' of digital platforms. By doing so, they can be held accountable by the broader community and are more likely to make decisions that prioritise inclusivity and fairness.

User empowerment is another cornerstone of a more equitable digital space. Beyond fostering digital capability, platforms should be designed to give users more control over their digital experiences. This could manifest as customisable algorithms that allow users to set their preferences or determine which sources are prioritised. Similarly, feedback mechanisms should be enhanced, allowing users to not only report biases but also participate in iterative platform improvements.

Digital archiving and preservation should also be prioritised. As digital content becomes an increasingly significant part of our cultural heritage, ensuring that diverse narratives and perspectives are archived for future generations is crucial. Collaborations between tech companies, cultural institutions, and governments can facilitate the creation of digital archives that are both comprehensive and representative.

Lastly, the ethos of continuous improvement and evolution should be embedded in the digital heritage sector. The digital landscape is ever-changing, and as new technologies emerge and societal dynamics shift, the strategies to ensure fairness and inclusivity will need to adapt. Regular reviews, audits, and consultations with diverse stakeholder groups can ensure that digital platforms remain relevant, accurate, and just. This does not just apply to platforms but also datasets, ontologies, models and all the other entities and actors that form part of the complex sociotechnical ecosystems through which knowledge about the past flows.

In essence, the 'agents to agencies' framework, combined with the insights from the SaaS framework dynamics, paints a picture of a digital ecosystem where every component, every decision, and every user interaction matters. The challenge and the opportunity lie in harnessing this interconnectedness to ensure diverse and pluralised representations of the past within systems of foundation models, datasets and often user experiences that are currently dominated by a few majority perspectives from the Global North. Part of this challenge involves addressing not just digital tools and approaches, but also digital infrastructure.

6.2 Multimodal Generative Approaches and their Implications

The advent of multimodal generative approaches, particularly in the realm of artificial intelligence, signals a transformative moment in the digital humanities, especially concerning the representation and interpretation of historical information. These approaches, which enable the integration and processing of multiple forms of data (text, image, audio, video), are poised to revolutionise how users interact with historical content, offering enriched, immersive experiences. However, within the 'agents to agencies' theoretical framework, it's imperative to consider these technologies not merely as neutral tools but as entities embedded within broader sociotechnical systems or 'agencies,' which encompass diverse stakeholders, power dynamics, and ethical considerations.

Multimodal generative AI models, especially GPT-4 and its potential successors, present opportunities for more nuanced and context-rich engagement with historical content. For instance, users might interact with a digital museum exhibit that not only presents text and images but also generates relevant audiovisual content or virtual reality experiences, deepening their understanding and appreciation of historical events or artefacts. In educational settings, such AI could dynamically generate content that caters to different learning styles and cultural backgrounds, potentially democratizing history education.

However, these technologies also carry significant implications and potential risks. One concern is the authenticity and accuracy of generated content. As these systems become more sophisticated in creating realistic images, sounds, or videos, the line between historical fact and AI-generated content may blur, raising critical questions about historical truth and representation. There's also the risk of these technologies perpetuating or exacerbating existing biases, as they're often trained on data sets that reflect dominant narratives and may lack diverse cultural and historical perspectives.

Specific AI models like DALL-E 3 and CLIP, and text-image datasets such as LAION-5b, have brought these considerations into sharp focus. DALL-E 3, an advanced image generation model, can create detailed visual representations from textual descriptions, offering potential for enriched historical education and virtual exhibits. However, its reliance on textual prompts raises concerns about the accuracy and authenticity of the representations, particularly when the text input is ambiguous or lacks historical context.

The main issue here is again a lack of plurality. For LAION-5b's 5.8 billion online images, the associated text caption in the dataset is automatically scraped from each image's 'alt text'. The alt-text is the sometimes auto-generated description that appears on webpages when images can't be loaded, for screen-readers used by visually impaired users and used most extensively by automated crawler bots for web search and indexing. These descriptions are thus designed to be short and literal, conveying visual information about the

image rather than any detail or nuance about what the image might mean or represent. The vast majority of these descriptions are written in English.

Therefore, while CLIP's ability to understand images and texts in concert is groundbreaking, its training and output are influenced by the data it has been trained on. For example, its interpretation of historical artefacts is contingent on the dataset's composition. In cases like the LAION-5b dataset, the prevalence of images of artefacts such as ancient coins sourced from auction sites rather than museums, or the mislabelling of replicas as authentic artefacts, can skew CLIP's 'understanding' and, consequently, the historical narrative it contributes to when images are generated based on its linking of images and their descriptions. Such nuances highlight the critical need for meticulously curated and verified datasets that reflect a diverse range of interpretations instead of a fallacious search for an objective 'ground truth' of what an image 'means' textually, let alone in a historical context of multiple interpretations¹¹.

These examples highlight the complex ethical terrain these technologies navigate. The 'agencies' involved in their development and deployment must be critically assessed for transparency, accountability, and representation. This involves not just the technologists and developers, but also historians, educators, policymakers, and the public. Each stakeholder plays a pivotal role in shaping the technology and its impact on historical representation, necessitating a collaborative approach that balances innovation with responsibility. Furthermore, the SaaS framework emphasises the dynamic role of users, necessitating digital capability and critical engagement to discern between AI-augmented content and historical fact. This complexity extends to the 'agents to agencies' framework, demanding a holistic strategy for the ethical deployment of these technologies, preserving historical authenticity while acknowledging AI's interpretative role.

In the SaaS framework, user roles become increasingly complex with the integration of multimodal generative AI. Thinking about produsers within the creator economy, whose inputs might influence algorithms determining the AI-

¹¹ Katie McDonough, 'You Cannot Ground Truth the Past: Computational Approaches to Historical Maps' (2023).

generated content they then see, as well as retraining both these algorithms and the generative AI models, it is still they who provide the meaning to potentially endless interactions with AI tools.

This dynamic highlights the importance of digital capability; users must be equipped to critically assess and engage with AI-generated content, discerning between authentic historical information and AI augmentations or fabrications. It is also important to note that an understanding of the different algorithms and AI tools at play within each ecosystem, as is currently the case for creators who simultaneously exploit and are exploited by the content recommendation algorithms of platforms like TikTok, is fundamental in shaping the agencies of produsers within the ecosystems, and thus to the content generated and shared.

The ethical considerations of deploying multimodal generative AI in historical representation are profound. There's an imperative to establish guidelines and standards that ensure the responsible use of these technologies, preserving the integrity of historical narratives while acknowledging the role of AI as an interpretative tool rather than a source of truth. UNESCO¹² and the European Parliament¹³ have already produced initial guidance documents in this area, highlighting the importance of understanding and regulating generative AI in the context of research, education and cultural heritage, pointing out that this requires collaboration among technologists, historians, educators, and policymakers to navigate the complex ethical landscape while still keeping up with the rapid pace of change in the sector.

Additionally, the 'agencies' involved in the development and deployment of these technologies must be critically assessed. Who has the power to develop these AI systems and bots, and whose interests do they serve? How are decisions made about the data sets they're trained on, and how are diverse historical narratives and cultural sensitivities accounted for? As these technologies evolve, there's a need for transparency and accountability in their

¹² Fengchun Miao and Wayne Holmes, *Guidance for Generative AI in Education and Research* (UNESCO, 2023) <<https://unesdoc.unesco.org/ark:/48223/pf0000386693>>.

¹³ Pasikowska-Schnass and Young-Shin, 'Artificial intelligence in the context of cultural heritage and museums'.

development processes, ensuring they contribute to a more nuanced and inclusive understanding of history rather than distorting or oversimplifying it.

Case studies in this realm might include AI projects in digital humanities, virtual museum exhibits using generative AI, or educational technologies that integrate multimodal AI to enhance history education. Each of these offers insights into the potentials and pitfalls of multimodal generative approaches, providing valuable lessons for future developments.

Consider the application of AI in digital humanities projects. These initiatives might use AI to analyse vast datasets of historical texts, images, or audio recordings, uncovering patterns and insights that would be impossible for humans to identify unaided. However, the quality of the AI's output is heavily dependent on the quality and diversity of the input data. If the datasets are skewed towards particular demographics or regions, the insights generated will reflect these biases, potentially leading to misrepresentations or incomplete narratives.

Virtual museum exhibits that leverage generative AI offer another compelling application. These exhibits could use AI to create immersive, three-dimensional representations of historical sites or events, allowing users to explore them virtually. While this can provide unprecedented access to history, especially for people who cannot visit these sites in person, it also raises questions about authenticity. How much creative license should AI be allowed in interpreting historical sites or artefacts? What mechanisms should be in place to ensure that these representations don't overwrite or distort historical truths?

Educational technologies that integrate multimodal AI also hold promise and peril. On the one hand, they have the potential to make history education more engaging and accessible to students across diverse backgrounds. For instance, AI could generate interactive content tailored to students' learning styles or cultural contexts, or create simulations that allow students to "experience" historical events first-hand. However, these technologies must be designed with care to avoid embedding biases or inaccuracies into educational content, which could misinform students or propagate skewed historical narratives.

In the SaaS framework, these applications underscore the complex interplay of user roles in shaping and experiencing AI-generated content. Curators and educators, as administrators in this framework, have the responsibility to ensure that the AI systems they deploy are both accurate and ethical. This might involve providing diverse and representative data for AI training, setting parameters to guard against the generation of misleading content, and establishing channels for feedback and correction where users can report issues or biases in the AI-generated material. Users, meanwhile, need to approach this content with a critical eye, understanding the potential for bias or error in AI systems and taking an active role in discerning historical fact from AI-enhanced interpretation.

The 'agents to agencies' framework reinforce the idea that these technologies do not operate in a vacuum. They are part of broader systems that include not only the AI and its users but also the institutions that deploy it, the stakeholders that fund it, and the societal norms and values that shape it. As such, the ethical deployment of multimodal generative AI in historical representation demands a holistic approach. It requires ongoing dialogue among all these 'agencies' to balance the exciting possibilities of these technologies with the imperative to preserve historical authenticity and diversity. It calls for transparency in how these systems are developed and used, and a commitment to education and advocacy that empowers users to engage with these tools responsibly.

The ongoing evolution of these technologies also necessitates a dynamic approach to governance and regulation. As AI systems continue to advance, it's likely that they will be able to create increasingly sophisticated and convincing representations of historical content. There may come a point where AI-generated representations are indistinguishable from authentic historical materials or interpretations created by historians. This potential raises profound questions about the nature of historical truth and the mechanisms needed to safeguard it in the digital realm. Regulators and policymakers will need to grapple with these questions, developing frameworks that ensure the ethical use of AI in historical representation while also fostering innovation and exploration in the digital humanities.

Moreover, the 'agencies' framework highlights the importance of interdisciplinary collaboration in these endeavours. Technologists, historians, educators, cultural practitioners, and others must come together to guide the development and application of multimodal generative AI in historical representation. Each of these stakeholders brings a unique perspective and expertise that is critical to ensuring these technologies are used responsibly and effectively. For instance, historians and cultural practitioners provide context and critical analysis that are essential in developing AI systems that can accurately and sensitively represent historical content. Technologists offer insights into the capabilities and limitations of AI, guiding its application in ways that are technically sound and ethically grounded. Educators play a crucial role in integrating these technologies into educational curricula in ways that enhance learning while also fostering critical engagement and digital capability.

Public engagement is also a critical component of this ecosystem. The broader public, as the primary consumers of AI-generated historical content, must be informed about the capabilities and limitations of these technologies. Public education initiatives can help individuals understand the potential biases and interpretive layers that AI might introduce, equipping them with the skills they need to critically engage with this content. These initiatives can also gather feedback from the public about their experiences with AI-generated historical content, insights that are invaluable in refining and improving these technologies.

In the context of the SaaS model, this emphasis on interdisciplinary collaboration and public engagement is even more pronounced. The nature of SaaS platforms as service-driven and user-focused intensifies the need for continuous feedback loops and adaptive mechanisms that can respond to emerging issues or concerns in real-time. This might involve, for example, mechanisms for users to report concerns about AI-generated content, processes for reviewing and addressing these reports, and systems for updating and improving AI systems in response to user feedback. The goal is to create a dynamic, responsive ecosystem that values user input and evolves in line with technological advancements, ethical considerations, and user needs.

The application of multimodal generative AI in historical representation offers exciting possibilities for enhancing public understanding and appreciation of history. However, it also presents significant challenges and risks, particularly concerning historical authenticity, ethical representation, and the potential for bias. Navigating these challenges requires a holistic, adaptive approach that considers the complex interplay of 'agents' and 'agencies' in this space. It demands collaboration, public engagement, and a steadfast commitment to upholding the integrity of historical narratives.

6.3 Digitisation and Digitalisation for Historical Research and GLAMs

As discussed previously, the processes of digitisation and digitalisation are two vital forces which influence almost every aspect of the sociotechnical ecosystems in which information about the past is present; or rather, these processes highlight presences and absences. While digitisation refers to the conversion of physical materials into digital formats, digitalisation denotes the utilisation of digital technologies and the data derived to impact how work gets done, transform how customers and companies engage and interact, and occasionally create new digital business models¹⁴. Both processes offer remarkable opportunities for access and engagement, yet also pose significant challenges and responsibilities. This section delves into the complexities of both.

The transformation of tangible artifacts, documents, and narratives into digital formats is not a neutral process; it is one steeped in decisions and biases that can significantly influence how history is understood and interpreted. This section explores the intricacies of digitisation, the pivotal role of data in this digital landscape, and the digital infrastructure that underpins these processes, all through the lens of the 'agents to agencies' theoretical framework and the practical dynamics of the SaaS model.

¹⁴ Aijia Zhang, *Attributing Digitalization Decisions in Museums: A Multiple Case Study of Swedish Public Museums*, 2023 <<https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-510841>> [accessed 29 April 2024].

The 'agents to agencies' paradigm compels us to view digitisation not merely as a technical process but as a complex socio-technical system involving a myriad of stakeholders, including cultural institutions, tech companies, policymakers, and end-users. Each 'agent' within this 'agency' wields a certain degree of influence over how historical and cultural materials are selected for digitisation, how they are digitised, and how they are made accessible and interpretable in the digital realm. These decisions are invariably influenced by factors such as institutional priorities, funding availability, technological capabilities, and societal values, which can lead to the amplification of certain narratives and the marginalisation of others.

In this intricate ecosystem, data becomes pivotal. In the context of digital historical and cultural materials, data encompasses digital representations of physical artifacts, metadata, user-generated content, and more. The ways data is created, managed, and utilised within the digitalisation process can significantly impact historical accuracy, cultural representation, and public engagement. For instance, metadata accompanying digitalised artifacts can influence their discoverability and interpretability, with implications for how history and culture are represented and understood.

Within the SaaS framework, various users interact with this data and metadata through digital platforms, each bringing their roles and motivations.

Administrators and curators decide what gets digitalised and how it's presented, while end-users engage with this content based on their interests, backgrounds, and digital capability. Additionally, third-party applications and plugins can shape the user experience, offering new ways of visualising, interacting with, or even augmenting digital historical content. However, these tools also bring their own biases and limitations, potentially influencing users' understanding and perception of history. Alongside this, the practical constraints on both the digitalisation process and the subsequent use and sharing of the data and metadata mean that digitalisations of collections effectively become historical snapshots of that collection over time, with ramifications for how it might be interpreted or misinterpreted in its wider sociotechnical ecosystem. For instance, whilst the GPT-4 model underlying ChatGPT has clearly stated versions that its knowledge cut-off date (i.e. the time covered by its training

data), such a model interacting with a digitalised GLAM collection may not have the same information for that collection. Many such online data sources are effectively historicised by the necessities of automation, but this is often overlooked or not appreciated by those who end up using their data.

Digital infrastructures, the backbone of these processes, further complicates this landscape, as exemplified by data colonialism – the power imbalance of organisations in the Global North shaping the ways in which people in the Global South then interact with models trained on their data, which they may even have helped to process for such companies often in exploitative working conditions¹⁵. These issues and the wider digital divide become increasingly pertinent, raising questions about who has access to the necessary technology for digitisation and digitalisation, whose history is being digitalised and preserved, and who has the power to make these decisions. For instance, cultural institutions in economically rich regions are often better equipped for large-scale digitisation and digitalisation projects, leading to an overrepresentation of their materials in the digital space. This imbalance can perpetuate historical biases and inequities, silencing the voices of those lacking the resources to participate in the digital preservation of their heritage.

Reflecting on these dynamics, it becomes evident that tackling the challenges posed by digitisation and digitalisation, data, and digital infrastructure requires a holistic, nuanced approach. Throughout this thesis there are numerous examples of where the wider sociotechnical ecosystems are shaped by a variety of actors with broader and sometimes competing priorities. A common theme has been that the practical reality of automation on the Web includes dealing with the constraints of resource, expertise and existing infrastructure. Although automation may be envisaged as a way of democratising access to GLAM collections, the costs of maintaining an API, updating a digitalised collection, using the APIs of social platforms to share that collection, maintaining the automated bots carrying out that activity and dealing with the changing policies and algorithmic approaches of those platforms, all need to be

¹⁵ A. Arora and others, 'Risk and the Future of AI: Algorithmic Bias, Data Colonialism, and Marginalization', *Information and Organization*, 33.3 (2023), p. 100478, doi:10.1016/j.infoandorg.2023.100478.

considered within an organisation's approach to automation. Simply assuming that automation, by definition, requires less effort will lead to ineffective implementation.

The digitalisation of cultural heritage is not just about preservation but also about accessibility and interpretation. When artefacts are digitised and placed online, they are removed from their original context, potentially losing certain nuances or meanings. Thus, how these artefacts are presented to the public, the information that accompanies them, and the interpretative tools available are all crucial aspects of cultural representation and understanding. For instance, an online exhibition might present artefacts in a way that tells a particular story or highlights certain aspects over others. The design of these digital exhibitions, therefore, becomes a powerful tool in shaping historical narratives. This also highlights how digitisation projects can fall down when they are viewed simply as neutral, technical processes, where nuanced interpretations become impossible when their original, physical context is stripped away. Rather than provide a 'solution' for debates around repatriation of museum objects, increasing digitalisation and digitisation have actually intensified such discussions and introduced added complexities¹⁶ – hence the need to remember that such work always needs to be critiqued through a sociotechnical lens and carried out collaboratively.

Interactions bring other opportunities and challenges too. User-generated content, often a significant aspect of digital interactions, adds another layer of complexity. Users' comments, reviews, or contributions to public digital archives become part of the artefact's digital life, influencing others' perceptions and understandings. However, this content is also subject to the digital divide, as those with greater digital access or literacy are more likely to contribute, potentially skewing public discourse.

Furthermore, the algorithms that govern searchability and visibility on digital platforms play a significant role. They can determine which artefacts are easily found or seen and which are obscured, often based on criteria that favour

¹⁶ Andrea Wallace and Mathilde Pavis, 'Finding the Nuance in Open Access During the Repatriation of Cultural Heritage' (presented at the Museums+Tech2019: Openness, 2019).

mainstream or popular narratives. This aspect of the digital infrastructure can subtly shape public understanding and discourse, reinforcing certain narratives while marginalising others.

The 'agents to agencies' framework suggests that these issues be addressed not just at the level of individual institutions or technologies but through broader societal engagement with the digital structures that govern our cultural heritage. This involves advocating for digital capability, promoting diverse representation in technological and cultural institutions, and fostering public awareness and engagement with these issues.

In grappling with these challenges, the role of policy cannot be overstated. Policymakers wield the authority to enact regulations that ensure digital platforms uphold principles of fairness, accuracy, and inclusivity, including FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective benefit, Authority to control, Responsibility, Ethics) data principles that also highlight the data governance requirements of communities in the Global South¹⁷. However, policy development and implementation in this realm must be agile and informed, capable of keeping pace with rapid technological advancements. This demands a symbiotic relationship between policymakers and experts in digital humanities, cultural preservation, and technology, and the communities that both own and are represented by such data, ensuring that regulations are grounded in the realities of digital interactions and technological capabilities.

For instance, policies could mandate the inclusion of diverse cultural representations in digital archives or require platforms to provide transparent criteria for how digital artefacts are displayed and ordered. They could also incentivise innovations aimed at enhancing digital capability, thereby empowering users to engage more critically and knowledgeably with digital content. Moreover, considering the global nature of the digital space, there's a need for international cooperation in policy development, ensuring that cultural preservation and representation are addressed consistently across borders. Regulation around generative AI is in its infancy, and varies greatly from country

¹⁷ Stephanie Russo Carroll and others, 'Working with the CARE Principles: Operationalising Indigenous Data Governance', 2020 <<https://www.adalovelaceinstitute.org/blog/care-principles-operationalising-indigenous-data-governance/>> [accessed 8 December 2023].

to country, so adopting and implementing an international framework may be a lofty ideal but it is perhaps the only way for effective impact of policy in this area¹⁸.

Education systems and Knowledge Infrastructures must also rise to the challenge, integrating digital capability into curricula from an early age. This isn't just about using digital tools, but also understanding their societal impact, the nuances of digital content creation and curation, and the critical assessment of online information. By fostering a more digitally capable citizenry, the influence of skewed narratives and the prevalence of misinformation can be mitigated. Furthermore, educators can play a pivotal role in leveraging digital platforms for cultural education, using them not just as repositories of information but as interactive spaces where students can engage with and contribute to cultural narratives.

Cultural institutions themselves stand at the forefront of these transformative efforts. Beyond digitising artefacts, they are tasked with bringing them to life for a digital audience, weaving narratives, and providing context. Their challenge is to do so in a way that is both engaging and educationally robust, which might involve embracing multimedia storytelling, interactive technologies, and user feedback mechanisms. They must also be cognisant of the digital divide in their audience, striving for accessibility in both the technology they employ and the narratives they construct.

Furthermore, the private sector, especially tech companies, has a significant role to play. These entities have the resources, technical expertise, and innovative drive to shape the digital landscape profoundly. Their collaboration with cultural institutions, educators, and policymakers can catalyse the development of technologies and platforms that are attuned to the needs of digital cultural preservation and education. However, this requires a shift in perspective, recognising the value in cultural preservation projects beyond immediate commercial gain. It calls for a commitment to social responsibility,

¹⁸ Miao and Holmes, 'Guidance for generative AI in education and research'.

investing in projects that promote cultural diversity, education, and global understanding.

Ultimately, the journey toward a more inclusive and accurate digital representation of our cultural heritage is a collaborative effort between the various 'agents' that populate the 'agencies' of our digital world and, most importantly, the different groups of users influencing and entangled within those agencies.

6.4 Historical Information, Misinformation and Disinformation

The proliferation of information, misinformation, and disinformation presents intricate challenges, especially in the context of historical content. The distinction between these elements is not merely semantic but has profound implications for public discourse, cultural understanding, and the collective memory of societies. Information is typically based on facts and authentic data; misinformation is incorrect or misleading information spread without malicious intent; and disinformation is false information spread deliberately to deceive. The 'agents to agencies' theoretical framework compels us to look beyond individual instances of misinformation or disinformation to the broader systems that facilitate their creation and dissemination.

The Web, and social media platforms in particular, have revolutionised access to information, but they have also provided fertile ground for misinformation and disinformation to flourish. These platforms are part of extensive sociotechnical ecosystems where various stakeholders—platform owners, content creators, and users—interact within a framework shaped by algorithms, policies, and cultural norms. Within the SaaS model, users play multiple roles, not just as consumers but also as contributors and distributors of content, blurring the lines between information and misinformation, especially when historical narratives are involved.

A fundamental point to consider for information about the past is the inherently interpretative nature of disciplines like history, where multiple perspectives, experiences and narratives intertwine to create historical 'facts'. These

potentially divergent interpretations pose challenges to the definitions of misinformation and disinformation as such terms become inherently subjective in historical contexts, especially when this nuanced subjectivity intersects with the more objective approaches often favoured by those developing the likes of AI models; as McDonough notes, there is no 'ground truth' for the past against which algorithms and models can be trained and tested¹⁹.

However, as has been made apparent throughout this thesis, different countries' and communities' versions of the past will naturally come to form predominant narratives about the past that could be considered 'ground truths' for those groups or individuals, as seen in the dominance of Alexa Skills discussing US history and simultaneously the popularity of the Black History Month Skills. Viewed through the lens of agencies, the complexities and contradictions at play here become apparent. The flexibility of foundation models like GPT-4 highlights the potential for groups of users to converse with ChatGPT in order to explore the past from their perspectives. This could come to represent the creation of 'pseudo ground-truths' that would allow factual objectivity to be tailored to users' preferences, as shown by the use of custom instructions for ChatGPT in Chapter 5. Yet, the agencies of OpenAI and those people involved in training GPT-4 to be 'safer' are also apparent when trying to push these subjective objectivities to the extremes – such as in the generation of content related to Holocaust denial, for example, which would be flagged by OpenAI as breaches of their content policy, hence drawing a hard line where historical interpretation ends and disinformation begins. These policies are created by human developers but enacted by AI models trained on what is harmful or otherwise by hundreds of other human workers²⁰.

AI models, particularly those capable of generating realistic text or images, introduce additional complexity as they hold the potential to create and amplify vast amounts of information that cannot be verified. For instance, an AI model trained on biased or inaccurate historical data might perpetuate misleading narratives, but such generations and the content within them are currently

¹⁹ McDonough, 'You Cannot Ground Truth the Past: Computational Approaches to Historical Maps'.

²⁰ Christiano and others, 'Deep Reinforcement Learning from Human Preferences', xxx.

unverifiable and still prone to mistakes, as shown in the example of sourcing content from Wikipedia in Chapter 5 with no citations. More concerning is the potential use of AI to create 'deepfakes'—highly realistic but entirely fabricated images or videos – or their textual equivalents which could be used to create false historical records or alter existing ones, complicating efforts to discern historical truths. Add in to this mix the ability to generate synthetic datasets and carry out analyses of them, and the process of knowledge creation and interpretation itself becomes manipulable, making provenance and verifiability of fabricated narratives virtually impossible.

From a research perspective, this also raises the question of what might constitute 'acceptable' approaches and methodologies for studying the past so that reliability is retained within approaches whilst enabling multiple interpretations and epistemologies. This goes hand in hand with how the outputs of such work are then shared more widely, where knowledge infrastructures and pedagogic practice play a crucial role. Users must be equipped with the digital capability skills necessary to critically assess the historical content they encounter online, to differentiate between reputable sources and misinformation or disinformation. This is particularly important in educational settings, where students are forming their understanding of history.

In the 'agents to agencies' framework, combating misinformation and disinformation in historical content requires a systemic approach. This includes critical examination of the algorithms that underpin content distribution on digital platforms, often designed to promote engagement, potentially at the expense of accuracy. It also necessitates transparency from platform operators about how these algorithms work and the implementation of robust fact-checking and content moderation policies. An important part of such discussions is that researchers, cultural institutions, and educators from all involved communities need to be able to actively engage with digital platforms so that they can contribute accurate, compelling historical content. They need to help decide what constitutes misinformation and disinformation (or 'misknowledge' and 'disknowledge') and then how to counteract its potential spread through such ecosystems. This requires collaboration with technologists and policymakers to address these challenges, advocating for policies and systems that uphold

historical accuracy and foreground polyvocal approaches, shifting away from rigid notions of 'ground truth' that can lead to unambiguous, and easily circumvented, algorithmic moderation. Enabling such approaches requires a fundamental shift in the power dynamics of online information ecosystems, otherwise academic and professional protestations are likely to fall on deaf ears.

The 'agencies' involved in the creation and spread of historical content online are numerous and interconnected. Recognizing and addressing the roles of these various 'agents'—from AI systems to platform algorithms, to individual content creators and consumers—is critical in the fight against misinformation and disinformation. It's only through a coordinated, multi-faceted strategy that we can hope to preserve the integrity of historical information in the digital age.

One area requiring attention is the algorithmic curation of content on social media and other digital platforms. These algorithms, designed primarily to maximise user engagement, often promote sensationalised or controversial content, which can include misinformation or disinformation. In the context of historical content, this might mean that more dramatic, but less accurate, reinterpretations of historical events are given precedence over nuanced, scholarly content. The 'black box' nature of these algorithms exacerbates the issue, as it's often unclear why certain content is promoted or suppressed. Transparency in algorithmic decision-making is crucial, as is ongoing research into the societal impacts of these algorithms.

The SaaS model, with its emphasis on user-generated content and interaction, offers both challenges and opportunities in this regard. Users, when equipped with appropriate digital capability skills, can flag and challenge misinformation or disinformation they encounter online. However, they can also unwittingly contribute to the spread of such content. Educating users about the potential for misinformation and disinformation and providing them with the skills to critically evaluate content, is crucial. Additionally, platforms operating under this model need to provide clear, accessible channels for reporting misinformation and the resources for timely, effective moderation.

Cultural institutions, such as museums, libraries, and educational institutions, have a vital role to play. By actively engaging in digital spaces, they can provide authoritative, accurate historical content to counterbalance the misinformation and disinformation that might exist. This requires a proactive digital strategy, including the creation of engaging, accurate content and the establishment of a strong, trustworthy digital presence. AI developers and researchers also have their part to play. They must ensure that the AI systems they develop are trained on accurate, diverse data sets, and they need to be transparent about the capabilities and limitations of these systems.

While organisations like OpenAI are working to improve AI safety and the 'guardrails' already put in place, it is unclear who beyond the organisation is involved in these processes and how decisions around what is safe and unsafe are made. It is here that meaningful collaboration with stakeholders outside the tech industry is urgently needed. This is highlighted by recent information about the release of the original ChatGPT model in November 2022, which flew in the face of a more cautious approach to allowing public access to previous GPT models, and a broader move by Big Tech companies away from more academic-aligned ethics work to commercially driven competitive decision-making²¹.

The global nature of digital platforms further complicates these issues. Misinformation and disinformation do not respect national borders, and content that is created or promoted in one country can quickly spread worldwide. This poses particular challenges for historical content, as historical events and figures can have different interpretations and significance in different cultures and societies. This ties in with the fundamental complexities of assigning values of 'truth' and 'falsehood' to the multifaceted narrative interpretations of the past

²¹ Dataethics, 'Big Tech Is Cutting Down on Ethics · Dataetisk Tænkehandletank'; John Naughton, 'ChatGPT Exploded into Public Life a Year Ago. Now We Know What Went on behind the Scenes | Artificial Intelligence (AI) | The Guardian', *The Guardian*, 9 December 2023 <[267](https://amp-the guardian-com.cdn.ampproject.org/v/s/amp.theguardian.com/commentisfree/2023/dec/09/chatgpt-ai-pearl-harbor-moment-sam-altman?amp_gsa=1&_js_v=a9&usqp=mq331AQIUAKwASCAAgM%3D#amp_ct=1702590810071&_tf=From%20%251%24s&aoh=17025906609618&referrer=https%3A%2F%2Fwww.google.com&share=https%3A%2F%2Fwww.theguardian.com%2Fcommentisfree%2F2023%2Fdec%2F09%2Fchatgpt-ai-pearl-harbor-moment-sam-altman> [accessed 15 December 2023].</p></div><div data-bbox=)

that constitute the differing histories currently produced by various groups and nations – the differences in history on Wikipedia in other languages being a prime case in point²². However, this poses significant regulatory challenges at a time when regulatory oversight already needs to be strengthened. Given the global nature of these platforms, international cooperation will be essential in developing and enforcing such regulations.

This further highlights the fundamental roles and influences of commercial entities in the sociotechnical systems surrounding bots, AI agents and the various groups of users with which they are involved. As discussed, these fundamental influences often illustrate power imbalances, especially where online social platforms are concerned, that lead to commercial organisations, via their leading figures and development teams, directly and indirectly shaping the types of content and interactions that are possible and legitimised. From the perspective of misinformation and disinformation, the concerns of commercial entities might be more closely aligned with building trust in their tools through blanket approaches, such as the problematic and arguably ineffective algorithmic content moderation of social media platform, rather than prioritising nuanced approaches to safety that more closely engage with all communities of potential users. Commercial organisations often apparently strive to democratise information for all users, but ineffective approaches to dealing with the inherent issue of misinformation could result in tools that limit polyvocal expression and so thwart decolonising efforts, resulting in tools and spaces that would seem to be ‘democratising’ only for some groups of users²³.

Cultural and educational institutions have a crucial role to play in this ecosystem. By providing accurate, engaging historical content, they can help to counterbalance misinformation and disinformation on digital platforms. They can also serve as authoritative sources for users seeking to verify the accuracy of historical content. However, there is a fine line between authority as trustworthiness and the dictation of singular (often colonial) narratives, as highlighted by the critique of heritage interpretation and ‘authorised heritage

²² Ford, ‘Writing the revolution’.

²³ Cook, ‘EmboDIYing Disruption’, p. 398.

discourse²⁴, and such nuances become trickier to handle as GLAM institutions try to adapt to online ecosystems filled with AI-generated content. Employing the same approaches to engaging audiences in nuanced, verifiable discussions about the past might be the best way to address the threat of widespread misdata, misinformation and misknowledge about the past.

6.5 The Future of Historical Social Machines: What Would a Foundation Model for History Look Like?

Envisioning the future of historical research in the digital age necessitates exploring the concept of a "foundation model" for history. Within the digital humanities, the emergence of AI and machine learning has presented innovative avenues for research and education. However, these technologies also prompt significant reflection on how historical information is collected, interpreted, and disseminated. A foundation model for history would not only utilise advanced technologies for data analysis and interpretation but also critically engage with the ethical, cultural, and methodological implications of digital historical scholarship.

At its core, a foundation model for history would be anchored in a multidisciplinary approach. Recognising that historical research is not merely the domain of historians, this model would foster collaboration across fields such as computer science, linguistics, anthropology, and philosophy. This interdisciplinary synergy would enhance the depth and breadth of historical research, integrating diverse methodologies and analytical frameworks.

Moreover, the foundation model would prioritise inclusivity and diversity in its sources. Historical narratives have long been dominated by certain perspectives, often marginalising or entirely omitting the experiences and contributions of various groups. By intentionally sourcing data from a wide array of archives, oral histories, languages, and regions, this model would strive to construct a more holistic and representative historical narrative. This approach

²⁴ Laurajane Smith, *Uses of Heritage*, Reprinted (2009).

recognises that history is not a monolith but a rich tapestry of interwoven stories and perspectives.

Ethical considerations would be at the forefront of the foundation model. As digital tools and AI become more prominent in historical research, questions of accuracy, authenticity, and bias become increasingly pressing. The foundation model would necessitate the development of ethical guidelines for the use of digital technologies in historical research, ensuring that these tools enhance rather than distort our understanding of the past. This includes transparent methodologies, critical source analysis, and an acknowledgment of the subjective nature of historical interpretation.

Given the potential of AI to analyse vast amounts of data and uncover previously unnoticed connections or patterns, it's essential to ensure these systems are built and employed responsibly. One initiative in this vein could involve the creation of a multidisciplinary task force, comprising historians, ethicists, data scientists, and representatives from diverse communities, tasked with overseeing the development and deployment of AI in historical research. This task force could establish guidelines for responsible AI use, such as ensuring diverse data sets for AI training, setting standards for transparency in AI decision-making, and creating mechanisms for accountability should AI systems inadvertently perpetuate biases or inaccuracies. This type of collaborative approach has already been highlighted by both UNESCO and the European Parliament in reports looking at the current and future roles of generative AI in research and education, and cultural heritage respectively²⁵.

Public engagement and accessibility would also be central tenets of the foundation model. Moving beyond traditional academic circles, this approach to history would leverage digital platforms to engage a broader audience actively. Interactive knowledgebases and open-access digital archives are just two examples of how the foundation model could democratise historical knowledge,

²⁵ Miao and Holmes, 'Guidance for generative AI in education and research'; Pasikowska-Schnass and Young-Shin, 'Artificial intelligence in the context of cultural heritage and museums'.

fostering a global dialogue that transcends geographical and cultural boundaries.

Furthermore, the foundation model would embrace adaptability and continuous learning. The digital landscape is ever-evolving, and the methodologies and tools used for historical research must be nimble enough to adapt. This model would encourage ongoing professional development, technological literacy, and a culture of innovation within the historical profession. It would also promote a critical mindset, encouraging scholars to continually reassess their methodologies and assumptions in light of new data or perspectives.

In the context of the 'agents to agencies' framework, the foundation model for history acknowledges the complex web of factors that influence historical research. It recognises that digital tools are not neutral but are shaped by—and in turn shape—the societies in which they are developed and used. As such, this model calls for a critical examination of the digital 'agencies' at play in historical research, from the developers of AI algorithms to the cultural institutions that curate digital collections, to the end-users who interact with digital history. This model doesn't just change how history is studied; it transforms what history is — a dynamic, ever-evolving discipline that grows and adapts with each new generation and technological advancement.

Extending this idea of knowledge about the past as something constantly shifting, existing in re-interpretive flux, the theoretical approach of the proposed 'agents to agencies' framework and the practical visualisation of complex sociotechnical ecosystems sought by the Ecosystems Graphs for Foundation Models project²⁶, highlights a need to reintroduce Berners-Lee's original 'abstract' into 'social machines' when thinking about the hidden realities of knowledge gatekeeping on the Web²⁷. Central to historical investigation is exploration of source, bias and the interpretive process, yet to fully understand the complexity of these strands in contemporary and future digital historiography the historian needs to be able to conceptually map the agencies inherent in the generation, retrieval and interpretation of knowledge about the

²⁶ Bommasani and others, 'Ecosystem Graphs'.

²⁷ Berners-Lee and Fischetti, 'Weaving the Web', p. 172.

past on the Web. When envisaged as a problem of interacting social machines, here the ‘abstract’ nature of interactions is a vital part of constructing agencies.

Take, for example, the roles of crawler bots in mediating online knowledge production and consumption. Even in 1997, at the first Museums and the Web Conference, the importance of web search and the crawler bots that enabled online collections to be found by users was highlighted as something influencing GLAM collection digitalisation. However, the influence of such fundamental automation on the Web has grown in complexity as the potential types of uses for data have increased, leading to far greater abstraction of these crucial mediatory processes. The roles of crawler bots in web search, scraping of training data, web archiving and chatbot-platform web browsing have seen their influence increase along with their seemingly abstract nature – it is almost impossible for most users to imagine using the Web without them. Yet, their abstraction and hidden ubiquity poses a challenge for historians trying to assess the influence of such bots in the generation and presentation of existing and future knowledge about the past; there is much said about explainability and transparency when it comes to black-box AI models, but an effective foundation model must ensure that the influence of more mundane Web automation is not overlooked to the point where the likes of crawler bots become truly abstract social machines that become invisible to historians, archaeologists and GLAM professionals.

In this vein, the foundation model emphasises the importance of feedback loops in historical research and the cycling and recycling of data, information and knowledge. These loops acknowledge that our understanding of history is not static; it's continually being reshaped by new discoveries, interpretations, and methodologies. Feedback might come from professional historians, but it could also originate from amateur history enthusiasts, independent researchers from unrelated fields, or AI algorithms designed to scan and interpret vast swathes of data for patterns or insights previously overlooked by human scholars. One potential implication of this in particular is around data produced by generative models then being used to train other such models, fine-tune the same models, become the focus of research itself or be used to ‘fill in the gaps’ of datasets, such as archaeological records which currently lack an image. Such

approaches are already being tested in other areas²⁸, but quite what they would look like or how they would be applied in the context of information about the past is a subject in urgent need of discussion.

Synthetic and augmented data could significantly enrich the envisaged foundation model for history, but also pose a number of challenges and require fundamental theoretical debate over what defines information about the past and what this means for its study and dissemination. The potential to overcome data limitations is one of the major advantages of synthetic data. In scenarios where original data is scarce or non-existent, synthetic data offers a way to generate data mimicking historical records, providing a broader base for analysis and interpretation. Augmented data, on the other hand, enhances existing historical datasets by introducing variations, which could aid in better understanding and interpreting historical events.

However, the use of synthetic and augmented data comes with a caveat regarding the potential for bias and misinterpretation. There's a danger of introducing or exacerbating biases, especially if the algorithms generating or augmenting the data are biased. This could lead to misinterpretations or misrepresentations of historical events or archaeological artefacts, especially when potentially combined with bots developed, or prompted by users, to provide interpretations from a specific perspective, a significant concern within the 'agents to agencies' framework discussed earlier. Here, the interplay between various stakeholders, from the creators of synthetic data to the end-users interacting with digital historical narratives, becomes complex and necessitates critical ethical considerations.

Synthetic and augmented data, when used responsibly, can lead to the development of more robust analytical tools within a foundation model for history. They can contribute to creating advanced machine learning models

²⁸ Viraf, 'Create A Synthetic Image Dataset — The "What", The "Why" and The "How"', *Medium*, 2020 <<https://towardsdatascience.com/create-a-synthetic-image-dataset-the-what-the-why-and-the-how-f820e6b6f718>> [accessed 12 October 2020]; Haixing Dai and others, 'AugGPT: Leveraging ChatGPT for Text Data Augmentation' (arXiv, 2023), doi:10.48550/arXiv.2302.13007; Anders Giovanni Møller and others, 'Is a Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks' (arXiv, 2023), doi:10.48550/arXiv.2304.13861.

capable of analysing vast datasets, uncovering patterns, and providing new insights into historical phenomena. The enhanced analytical capabilities open up new avenues for interdisciplinary collaborations and innovative research methodologies, enriching the field of digital humanities.

Public engagement could also see a substantial enhancement through the use of synthetic and augmented data on digital platforms. Providing visual or interactive representations of historical events could foster a deeper connection and understanding among the public. However, it's imperative that ethical guidelines be established to ensure the authenticity and proper contextualization of synthetic historical materials. Additionally, potential privacy concerns need to be adequately addressed to uphold ethical standards in digital historical research.

Global accessibility and representation are other significant areas where synthetic and augmented data could have a profound impact. By generating data representing underrepresented groups or events, synthetic and augmented data can contribute to more inclusive historical narratives. Yet, the need for digital capability and critical engagement from the public is paramount to navigate the complexities and nuances introduced by synthetic and augmented data, something which AI-enabled chatbots in particular might be uniquely positioned to assist in achieving.

These points become clear when thinking through a couple of different examples. Take the cases of producing images for archaeological data records and generating textual descriptions for existing images of objects in museum collections. Both are types of data augmentation, essentially predicting new information from existing data, whereas generating entirely new sets of images and text descriptions for a museum exhibition would be synthetic generation. The datasets are different, but the principles and debates remain the same – is this of any value and to whom, and how is this decided upon? Fine-tuning might be able to mitigate against the inherent biases of any foundation model to an extent, but then this process in itself may already be reliant on data taken out of context through digitisation. Ethical guidelines can be drawn up, but in practice these often differ between academic and corporate settings, and whilst

collaborative frameworks can be established, as this discussion has shown it needs to extend to cover all the 'agencies' influencing a foundation model upon which generation would then occur as complexity at this scale invites, perhaps even necessitates, biases to occur at some point. Synthetic data generation highlights these fundamental challenges perhaps more than any other example, and the true value of proposing a historically relevant foundation model (or the 'historicising' of existing models) is to practically consider the effects of this growing complexity before it all becomes too difficult to entangle.

This approach also necessitates a re-evaluation of what constitutes a 'source.' In the digital age, historical narratives can be derived from a multitude of mediums: social media posts, digital art, and interactive web content, among others; or, as discussed, potentially generated based on a range of existing data. The foundation model acknowledges and integrates these diverse data points, and allows for their interrogation by a potential multitude of other models, but the incorporation of these new data types introduces additional complexities, particularly concerning data validity and ethical considerations. The existence of a foundation model, therefore, calls for the establishment of rigorous protocols for data verification, ensuring that digital artefacts are authenticated and contextualised appropriately. Similarly, ethical guidelines must be expanded to address concerns specific to digital research, such as data privacy, digital rights management, and the potential repercussions of AI-generated historical narratives.

In this regard, the model would need to champion the concept of 'open history.' Just as open-source software allows for collaborative coding, open history encourages the shared creation of historical narratives. This could manifest through crowd-sourced research projects, public digital archives, or collaborative virtual reconstructions of historical events or locales. By inviting widespread public participation, history becomes a communal effort, bridging cultural, social, and geographical divides and fostering a sense of global identity and collective memory.

Approaches focussing on decolonising existing ontologies and data structures have also highlighted the need to recognise and represent non-Western

epistemologies if truly diverse perspectives on the past are to be made possible in digitally-mediated spaces²⁹. In the context of foundation models, which can encompass inscrutably huge volumes of multimodal data, the idea of encompassing every perspective and interpretation possible no longer seems a pipe-dream, while an individualised approach to interactions with a model also means that there's a greater likelihood of that plurality being reflected back to users in the content generated. The capability of such models to generate synthetic data also represents a fundamental blurring of the prevalent Western dichotomy of subjective or objective data, and another clear example of why data cannot be 'neutral'. These opportunities for conceptual shifts offer a chance for all groups of users to decolonise their imaginations (c.f. Nguyen and Chia³⁰) and for data itself to become enchanting (c.f. Graham³¹), but such potential changes are reliant on the infrastructure needed for development of foundational models to be made available to wider groups of people than the few commercial, Western entities that currently control it.

However, different rates of digitalisation pose challenges in accessing historical content for some communities for which mitigations are needed. These might include providing digital history resources in multiple languages, developing platforms that are navigable regardless of an individual's tech proficiency, and ensuring that digital historical content is accessible on a range of devices. As demonstrated in chapter 5, despite its complexities ChatGPT and its ilk offer non-experts the opportunity to more effectively carry out more technical and challenging tasks with greater ease than before, potentially helping to bridge current skills gaps. However, it must be recognised that access to these platforms in the first place is far from equitable, and that this in itself should form the starting point when thinking about tackling information inequalities.

In the context of a foundation model for historical research, the integration of synthetic data and generative AI methodologies presents transformative

²⁹ Srinivasan, 'Whose Global Village?'; Hacigüzeller, Taylor, and Perry, 'On the Emerging Supremacy of Structured Digital Data in Archaeology'.

³⁰ Nhai Thi Nguyen and Yeow-Tong Chia, 'Decolonizing Research Imagination: A Journey of Reshaping Research Epistemology and Ontology', *Asia Pacific Education Review*, 24.2 (2023), pp. 213–26, doi:10.1007/s12564-023-09822-8.

³¹ Graham, 'An Enchantment of Digital Archaeology'.

opportunities while also posing ethical and practical challenges. Within both the 'agents to agencies' framework and the SaaS model, the use of AI to generate images for archaeological records where none exist enhances the inclusivity and depth of historical narratives, providing visual contexts that might otherwise be lost to time. This not only democratises historical understanding by offering more tangible connections to the past for a wider audience but also provides researchers and educators with valuable tools to better illustrate and analyse historical developments.

However, the generation of synthetic images and data must be approached with caution. The 'agents to agencies' framework highlights the importance of considering the various stakeholders involved in the creation, distribution, and consumption of synthetic historical materials. It's crucial to establish rigorous verification processes to maintain the authenticity of the narratives these images support, preventing the dissemination of potentially misleading or decontextualised information. This necessitates a collaborative approach involving historians, technologists, and other experts to ensure the generated images are consistent with existing archaeological and historical evidence.

Furthermore, within the SaaS framework, users' roles are not just passive recipients but active participants who interact with and influence the AI-generated content. This underscores the need for educational initiatives to enhance digital capability, enabling users to critically engage with synthetic data and understand its role and limitations in historical research. Mechanisms should also be established for users to report concerns or inaccuracies, reinforcing the system's credibility.

The concept of 'open history' is exemplified by projects like virtual reconstruction of historical sites destroyed by war or natural disasters. Through crowdsourcing, people worldwide could contribute photographs, videos, and firsthand accounts to assist professionals in creating accurate 3D models of these sites. These digital reconstructions, accessible to the global public, serve not only as educational tools but also as collective efforts in preserving cultural heritage.

To combat the digital divide, the foundation model could advocate for initiatives like mobile-accessible educational platforms that provide historical content in various formats (text, audio, video) and languages. These platforms could be designed for low-bandwidth access, ensuring individuals in regions with limited internet connectivity are not excluded. Furthermore, the model could support the development of community digital hubs in under-resourced areas, providing communal access to digital history resources, training, and interactive experiences.

In the pursuit of a more inclusive historical narrative, the foundation model could also support projects that specifically aim to highlight underrepresented voices and stories. For example, an initiative could be launched to digitise, preserve, and share oral histories from indigenous communities, many of which might not have been traditionally recorded in written form. By using advanced recording technologies and AI-powered transcription tools, these oral histories could be captured in their native languages and translated into multiple languages for broader accessibility. The platform could allow community members to contribute directly, ensuring their stories are told authentically and as intended. Additionally, AI could be used to cross-reference these narratives with other historical data, providing a fuller picture of historical events and cultural practices.

The aim of this final chapter has been to navigate the intricate landscape of digital interaction with historical content, guided by the 'agents to agencies' theoretical framework and the practical dynamics of Software as a Service (SaaS) models. It delves into the profound transformation digitisation imparts on the preservation, interpretation, and dissemination of historical artefacts, acknowledging the spectrum of decisions and biases that shape public engagement with history. The discourse acknowledges that the digital realm, with its vast potential for inclusivity, also mirrors, and sometimes amplifies, existing societal biases, presenting both opportunities and challenges in the equitable representation of cultural and historical narratives.

The 'agents to agencies' framework is instrumental in dissecting the complexity of these digital interactions, moving beyond the notion of isolated actors to

reveal a comprehensive ecosystem of interconnected entities and influences. This perspective illuminates the roles and responsibilities of various stakeholders, from developers and institutions to end-users, each contributing to how history is perceived, interpreted, and reshaped in the digital sphere. The chapter underscores the necessity of a holistic approach to digitisation, one that critically evaluates and actively engages with the inherent power structures and biases within digital infrastructures.

In the realm of AI and advanced technologies, the discourse explores the transformative potential of multimodal generative approaches, particularly their capacity to enrich user interaction with historical content. However, it also cautions against the uncritical adoption of such technologies, highlighting the ethical, cultural, and factual accuracies at stake. The chapter advocates for interdisciplinary collaboration to navigate this ethical landscape, ensuring these technologies are harnessed responsibly and contribute to a nuanced and inclusive understanding of history.

Central to the discussion is the concept of a foundation model for history, an innovative approach that emphasises multidisciplinary collaboration, diversity, ethical consideration, and public engagement. This model represents a paradigm shift in historical research and engagement, promoting a more dynamic, inclusive, and participatory exploration of history. It recognises the necessity of adapting to the digital age's challenges and opportunities, advocating for strategies that not only leverage technological advancements but also address systemic inequalities and biases.

In summary, the chapter presents a forward-looking vision, one that acknowledges the complexities of digital transformation in historical engagement but remains optimistic about the prospects. It calls for collective responsibility, continuous dialogue, and an interdisciplinary approach in redefining our engagement with history in the digital age requiring the commitment of all users involved – from developers to researchers to GLAM professionals to community members. By embracing innovation and inclusivity while upholding ethical and scholarly rigour, we pave the way for a more diverse, equitable, and enriched understanding of our collective past.

7. Conclusion

This thesis embarked on a journey to unravel the intricate roles of algorithms and automated bots in the sphere of historical narrative, representation, and interpretation. Throughout the exploration, it became unequivocally clear that the digital agents, often perceived as impartial or neutral, are profoundly human in their genesis and function. They are created from communities that span a few individuals to billions, each member imprinting their biases, intentions, and cultural nuances onto the digital entities they create or interact with. This complex interplay blurs the lines between human and machine, making it essential to perceive algorithms and bots as extensions or manifestations of human agency and culture.

In this light, algorithms and bots serve as both mirrors and moulders of societal norms, biases, and power structures. They predominantly amplify Western and Anglo-centric narratives, often sidelining non-Western histories and perspectives in a digital echo of historical colonial practices. This not only perpetuates existing information inequalities but also underscores the urgent need for diversification and democratisation of digital spaces. For these automated entities to become trusted tools for sharing global history and heritage, there must first be a shift in the human landscapes they represent, ensuring inclusivity and diversity are intrinsic in the communities that build and interact with them.

This work firmly positioned itself within the interdisciplinary praxis of Digital Humanities research, being well suited to exploring the complex interactions inherent in sociotechnical systems. Previous and ongoing work in this area has highlighted different aspects of such systems, from the physical and digital infrastructures which underpin and shape interactions¹, to the effects of existing data structures, algorithms and machine learning models on researchers and users when trying to make sense of complex, messy and ontologically diverse

¹ Pereda, 'Where do I stand? Deconstructing digital collections [research] infrastructures: a perspective from Towards a National Collection'; Mullaney, 'QWERTY in China'.

information about the past². This work also stems from a growing call within scholarship exploring the influences of the sociotechnical systems inherent in the Web on information about the past towards decolonisation of these systems themselves. The complexities that emerge from this become clear when systems force ambiguity upon the roles of researchers and their methods, whilst simultaneously attempting to strip such nuance from the commercialised end product of information³. This thesis highlighted that issues of power, representation and perspective became even more pronounced as the influence of automated and AI-enabled bots reshaped and formed new sociotechnical networks.

The social machines that emerged from interactions, as seen in the case of the art bots on X (formerly Twitter) and Alexa Skills, highlighted that the realities of human-bot collaboration were far more complex and nuanced than the original model of social machines imagined. The boundaries between 'creativity' and 'administration' are blurred when automation reveals a distinct lack of interpretative information for cultural heritage objects, or when the churning out of history quiz questions for the chatbot becomes a creative chore. And yet when bots and algorithms promote the creation of human-bot communities around art and cultural heritage, it is the human administrative task of maintaining and nurturing them that becomes rewarding. These are true social machines, but their roles have been reimagined.

Whilst chatbots and education tools powered by the latest generative AI models are already sparking interest in the tech and GLAM sectors, especially when it comes to engaging new audiences, GLAM professionals will need to take a shifting set of realities into account when dealing with the coming wave of bots powered by generative AI. The potential for rapid and frequent changes in the underlying models, ambiguity about responsibility for the outputs of those models, copyright concerns and legal grey areas are all going to be part of the messy and mundane reality hidden behind the AI-chatbot hype. The exploration

² Murrieta-Flores, Favila-Vázquez, and Flores-Morán, 'Indigenous deep mapping'; Candela and others, 'An Ontological Approach for Unlocking the Colonial Archive'; Hacıgüzeller, Taylor, and Perry, 'On the Emerging Supremacy of Structured Digital Data in Archaeology'.

³ Graham, 'An Enchantment of Digital Archaeology'; Bowker and Star, 'Sorting Things Out'.

of ChatGPT and plugins (now part of 'GPT' bots soon to have their own store) also showed that just as today's GLAMs curate online communities as well as collections, in future their work will inevitably involve the curation of bots too.

Furthermore, the potential of bots and AI models as research tools was explored, revealing a landscape ripe with opportunities yet fraught with challenges. These digital entities, especially in their role in mass digitisation, unearth vast swathes of historical content, often highlighting under-researched areas and potentially democratising access to history. However, they also bring to light the inefficiencies, subjectivities, and biases inherent in digital work, emphasising that these digital processes are neither swift nor impartial. Researchers will need to engage critically with the ecosystems in which these tools operate and have been created before a detailed understanding of the tools themselves will be possible. In the future, it may be that this level of critical engagement is no longer seen as necessary, but for now it is vital if hasty, ill-judged and inaccurate historical research is to be avoided.

However, it is clear from this work that currently and in at least the near future human interpretation and curation of the past are still the key to its dissemination and power to engage. Wikipedia stood out as the most influential source for historical information, something which was only amplified when considering automated sharing of information and the biases of the training data fed into foundation models, and its editorial decisions are still very much human. It is also clear, for now, that automated and AI-powered approaches have significant limitations: existing online collections, primarily from larger institutions in the Global North, favour breadth over depth and are not rich enough in engaging information; AI models retain the ability to fabricate information and their very notion still exudes an aura of mistrust across populations – from researchers to GLAM professionals to public users. As the historical background to this work highlighted, these issues are nothing new and overcoming them will require far more than simply adopting the latest chatbot.

Given this increasing complexity, the proposed 'agents to agencies' theoretical framework aims to build upon the idea of a network of social machines to provide a more nuanced sociotechnical lens through which to examine the

multifaceted nature of digital historical representation. It compels researchers, GLAM professionals and developers to look beyond the surface interactions of users with digital content and delve into the underlying power structures, ethical considerations, and systemic biases that permeate the Web. It highlighted that the digitisation and digitalisation of the past are processes that are deeply cultural, ethical, and political, with far-reaching implications for how the past is understood, interpreted, and reshaped.

It has become clear throughout the thesis the tools of knowledge creation and sharing themselves, be they algorithms, bots, AI models or indeed humans, both act and are used within the wider context of the Web and the existing interactions between users and online content. One question at the start of this work was whether automated and algorithmic approaches could diversify both access to knowledge about the past and the interpretations of that knowledge shared online. This thesis has shown that while these approaches have the potential to achieve this, they also could easily amplify existing biases and filter bubbles at the same time – theoretical democratisation of access to knowledge, but practical constraint of interpretation. The latest generative AI models only serve to increase this potential for greater breadth of information to be funnelled into more singular interpretations, especially once the expected increase in synthetic data is taken into account. In many ways, this might increase the importance of GLAMs as places of diverse, engaging interpretations that could be produced and shared with the aid of such tools, highlighting the need for those within the sector to better understand such models and their ecosystems.

One central contribution of this work that has been gleaned from considering the influence on knowledge about the past from the use of automation on the Web is the understanding that decisions made around automating access to collections, enabling their reinterpretation and sharing, are often made by a range of people with varying levels of expertise about the past, and usually limited by the practical constraints of existing collections and Web ecosystems. The recurring issues of resource, cost, skill, risk and technical capability feature in every case study, from unofficial sharing of artworks on social media to collaborative efforts between researchers and GLAMs to create accurate and engaging chatbots. These points highlight the messy social nature of

sociotechnical interactions that often take place between humans away from online spaces, but which can have significant implications upon them, such as the drawing up of policies and procedures around GLAM collection digitalisation.

The different methodologies employed throughout this thesis, and the experimental approaches taken to developing and testing them, are in themselves important contributions to the field. The approaches described in this work required playful and creative processes in both their development and implementation in order to fully explore and understand the bots, algorithms and AI tools being investigated, along with their sociotechnical ecosystems. The results of this novel methodological work, from attempting to create qualitative interview schema for chatbots to the slippery trial and error of nudging large language models into generating synthetic archaeological datasets, highlight the fact that any research grounded in the unfathomable sociotechnical complexities of the Web, and especially generative AI models, will require approaches that are necessarily and unambiguously subjective.

Although this thesis has attempted to cover a breadth of areas related to bots, historical information and the GLAM sector, there are inevitably some gaps, especially as this work developed over the course of a six-year period during which the technological landscape changed rapidly and frequently. One area that was not explored in any detail was that of ‘deepfakes’, primarily because at the start of the thesis these were still predominantly manual human endeavours, although this has changed over the years with the rise in increasingly sophisticated AI models. Coupled with the predicted rise in synthetic content and data generated by the same models, this is undoubtedly an area for future work focusing on history.

Future work will also need to explore what calls for greater equity, inclusion and representativeness of content and community actually mean in practice for developers, researchers, GLAMs and users themselves. In an environment where the companies behind the latest AI models seem more focussed on rapid progress and market share than AI safety or ethics, researchers and professionals will need to find ways of ensuring that voices from academia and

cultural heritage can actually be heard. These are the gaps through which sociotechnical agencies permeate, and at the moment they are yawning chasms. Perhaps, for better and worse, a historical social machine might be the only way to fill them.

Appendices

Appendix 1: Alexa Skills' Example Transcripts

Example 1: World History Skill

[Utterance:] open world history - OK, Here's World History

[Utterance:] open world history - World History. What day do you want world history for?

[Utterance:] today - For October 8, In 314, "Constantine I defeats Roman Emperor Licinius, who loses his European territories. In 451, "The first session of the Council of Chalcedon begins. In 876, "Frankish forces led by Louis the Younger prevent a West Frankish invasion and defeat emperor Charles II ("the Bald"). Wanna go deeper in world history?

[Utterance:] yes - In 1075, "Dmitar Zvonimir is crowned King of Croatia. In 1200, "Isabella of Angoulême is crowned Queen consort of England. In 1322, "Mladen II of Bribir is deposed as the Croatian Ban after the Battle of Bliska. Wanna go deeper in world history?

[Utterance:] yes - In 1480, "The Great stand on the Ugra river puts an end to Tartar rule over Moscow In 1573, "End of the Spanish siege of Alkmaar, the first Dutch victory in the Eighty Years' War. In 1645, "Jeanne Mance open the first lay hospital in North America. Wanna go deeper in world history?

[Utterance:] no - Goodbye

[Utterance:] disable world history - Okay, I've disabled it.

Example 2: History Facts Skill

[Utterance:] open history facts - OK, Here's History Facts

[Utterance:] open history facts - Here's your fact: Similarly, Vikings did not drink out of the skulls of vanquished enemies. This was based on a mistranslation of the skaldic poetic use of $\tilde{A}^3r\ bj\tilde{A}^0gvi\tilde{A}^0um\ haus$ (branches of skulls) to refer to drinking horns.

[Utterance:] open history facts - Here's your fact: Vomiting was not a regular part of Roman dining customs. In ancient Rome, the architectural feature called a vomitorium was the entranceway through which crowds entered and exited a stadium, not a special room used for purging food during meals.

[Utterance:] open history facts - Here's your fact: There was never a bill to make German the official language of the United States that was defeated by one vote in the House of Representatives, nor has one been proposed at the state level. In 1794, a petition from a group of German immigrants was put aside on a procedural vote of 42 to 41, that would have had the government publish some laws in German. This was the basis of the Muhlenberg legend, named after the Speaker of the House at the time, Frederick Muhlenberg, a speaker of German descent who abstained from this vote.

Example 3: Ancient Greek History Quiz Skill

[Utterance:] open ancient greek history quiz - OK, Here's Ancient Greek History Quiz

[Utterance:] open ancient greek history quiz - Ancient Greek History Quiz. I will ask you 5 questions, try to get as many right as you can. Just say the number of the answer. Let's begin. Question 1. What king led the Spartans in their famous suicidal stand against Xerxes' Persian army? 1. Pericles. 2. Leonidas. 3. Lycurgus. 4. Alexander the Great.

[Utterance:] two - That answer is correct. Your score is 1. Question 2. Through what defensive alliance of Greek city-states, established just after the Persian wars, did the Athenian statesmen Pericles divert funds for the beautification of Athens? 1. Delian League. 2. Athenian Alliance. 3. League of Corinth. 4. Peloponnesian League.

[Utterance:] one - That answer is correct. Your score is 2. Question 3. Despite the power of their army, the Spartans lived in constant terror of an uprising of their slaves, who outnumbered them 8 to 1. What were these slaves called? 1. ethiopians. 2. persians. 3. ilotes. 4. tracians.

[Utterance:] three - That answer is correct. Your score is 3. Question 4. Name the poet who immortalized the Trojan War and its effects in 'The Iliad' and 'The Odyssey' 1. Homer. 2. Hippolytus. 3. Herodotus. 4. Plato.

[Utterance:] one - That answer is correct. Your score is 4. Question 5. In one of the pivotal battles of the ancient world, 300 Spartans held a mountain pass against a Persian army of thousands. At the end, every Spartan lay dead, but precious time had been gained for the rest of the Greek army. Name the battle. 1. alamis. 2. Thermopylae. 3. Zama. 4. Marathon.

[Utterance:] two - That answer is correct. You got 5 out of 5 questions correct.
Thank you for playing!

Glossary

Agent (AI): In artificial intelligence, an agent is a computer system programmed to act independently to perform tasks on behalf of users, such as answering questions or carrying out specific actions.

AI (Artificial Intelligence): AI refers to the creation of machines or software that can think and learn in a way that resembles human intelligence.

Algorithm: An algorithm is a set of rules or instructions given to a computer or machine to help it perform a task. It's like a recipe that tells a computer exactly what steps to take to solve a problem or achieve a goal.

Amazon Alexa: This is a virtual assistant technology developed by Amazon, which responds to voice commands to perform tasks, such as playing music or providing news updates.

(Amazon) Alexa Skills: These are custom features that extend the capabilities of Amazon Alexa devices, allowing them to perform additional tasks or interact with other apps and services.

Amazon AWS: Amazon Web Services (AWS) is a collection of services offered by Amazon that provide on-demand computing resources like storage, databases, and applications over the internet.

Anthropic: Anthropic is an artificial intelligence research company, focusing on developing AI systems with a strong emphasis on safety and interpretability. It is best known for developing the conversation AI model 'Claude'.

API (Application Programming Interface): An API is a set of tools and rules that allows different software programs to communicate with each other. It's like a translator that helps one program speak to another.

Application (app): An app is a type of software designed to help people carry out tasks or activities on computers, tablets, or mobile phones.

AT&T: AT&T is a large US-based telecommunications company. It provides services like mobile and landline telephone, broadband internet, and television.

Big Data: Big data refers to extremely large sets of digital information that are analysed by computers to reveal patterns, trends, and associations.

Blockchain: Blockchain is a digital ledger technology where data is recorded in a sequence of blocks, and each block is linked to the previous one through cryptography.

Bots: Bots are automated programs that can complete tasks online, like posting content or answering questions, often without the need for human intervention.

ChatGPT: ChatGPT is an advanced AI model developed by OpenAI that can chat with users and generate human-like text based on the prompts it receives.

Chatbots: Chatbots are computer programs designed to carry on conversations with human users, typically for customer service or information retrieval purposes.

Conversational Agent: A conversational agent is a type of AI that can engage in dialogue with humans, often used in customer service or information retrieval.

Corpus Linguistics: Corpus linguistics is the study of language as expressed in samples (corpora) of real world text. This method relies on analysing large databases of actual language use to study how words and phrases are used in different contexts.

Crawler Bots: Special bots designed to systematically browse the internet and index web page content for search engines.

Cultural Heritage: This term refers to the legacy of physical artefacts, buildings, and intangible attributes and customs inherited from past generations, considered important to preserve for future generations.

DALL-E: DALL-E is an AI model created by OpenAI that generates images from textual descriptions, allowing for the creation of visual content based on text prompts.

Dataset: A dataset is a collection of related sets of information that are composed of separate elements but can be manipulated as a unit by a computer.

Decentralised Web: A decentralised web refers to a network structure for the internet that distributes data across many different computers instead of relying on centralised servers.

Deep learning: Deep learning is an AI method involving neural networks with many layers, used for tasks such as image recognition and natural language processing.

Developer (software): A software developer is a professional who creates digital products like apps, websites, and computer programs.

Digital Platforms: Online services or tools that allow users to create, share, and engage with digital content.

Digitalisation: Digitalisation means integrating digital technologies into everyday life or work, transforming processes or services to be more efficient and accessible.

Digitisation: Digitisation refers to the process of converting information from a physical format into a digital one, such as turning pages of a book into an online PDF.

DIKW Pyramid: This is a model for representing the stages and relationships between Data, Information, Knowledge, and Wisdom.

Disinformation: Deliberately created to mislead or misinform, disinformation is false information that is spread intentionally, often for harmful purposes.

Eggdrop: Eggdrop is a bot which is used for managing channels on Internet Relay Chat (IRC) and assist in moderating and maintaining IRC chat rooms.

ELIZA (chatbot): ELIZA is one of the earliest examples of a chatbot, which was designed to mimic conversation with users in a way that could resemble talking to a therapist, with often simple and reflective prompts.

Europeana: A digital platform that provides access to millions of digitised items from museums, galleries, libraries, and archives across Europe, intended for education, research, and cultural engagement.

Facebook (see Meta): Facebook is a social media platform that connects people to share photos, thoughts, and updates with friends and family.

Foundation Model (AI): In the context of AI, a foundation model is a broad, adaptable AI system that serves as a base for multiple applications, able to generate or process diverse types of content.

Generative AI: AI that can produce content, such as text or images, which can mimic human-created content.

Generative Adversarial Network (GAN): Generative Adversarial Networks are a class of AI algorithms used in unsupervised machine learning and are particularly known for generating realistic images.

GLAMs (Galleries, Libraries, Archives, Museums): These are cultural institutions that preserve and share important historical and cultural works.

Google: Google is a multinational company providing internet-related services such as a search engine, online advertising, cloud computing, and software products.

GPT (Generative Pre-trained Transformer) model: GPT is a type of computer program designed to understand and generate human-like text. It works by analysing lots of examples of writing and learning patterns in the language.

GPT-4: GPT-4 is a highly advanced AI model developed by OpenAI that can generate text and process images, among other capabilities.

Hardcoded: When something is hardcoded, it is permanently written into the code of a software program and does not change through reflexive processes like machine learning.

Hypertext Markup Language (HTML): Hypertext Markup Language is the standard language used to create and design web pages and web applications.

Hypertext Transfer Protocol (HTTP): Hypertext Transfer Protocol (HTTP) is the foundation of data communication for the World Wide Web. It is a protocol used for transmitting hypermedia documents, such as HTML, and governs how messages are formatted and transmitted.

IBM (International Business Machines Corporation): a global company providing a wide range of hardware, software, and services to businesses and governments, known for its innovation in areas including artificial intelligence, and cloud computing.

IBM Watson: an artificial intelligence (AI) system developed by IBM that can process natural language and provide insights or answers by analysing large amounts of data.

Instagram (see Meta): Instagram is a social media app and service for sharing photos and videos, offering various filtering and editing tools.

Internet (The): The internet is the vast network infrastructure that connects computers all over the world, allowing them to communicate with each other. Through the internet, people can access information, send emails, share files, and more using various services such as the World Wide Web.

Internet Protocol: A set of rules governing the format of data sent over the internet or local network.

Internet Relay Chat (IRC): Internet Relay Chat, or IRC, is a form of real-time text-based communication through the internet.

LAION-5b: LAION-5b is a large and diverse text-to-image dataset used to train AI models in generating images from textual descriptions.

Large Language Models: A type of AI model designed to understand or generate human language on a large scale.

Linked Open Data (LOD): A method of publishing structured data to interconnect related data across the Web.

Machine learning: Machine learning is a field of artificial intelligence that uses statistical techniques to give computers the ability to "learn" from data.

Meta (company): Meta (formed in 2021, before this known as Facebook) is the parent company of the social platforms and messaging apps Facebook, Instagram and WhatsApp.

Metadata: Metadata is data about data. In other words, it's like a summary or description of the data's characteristics, such as author, date created, or contents.

Microsoft: Microsoft is a large multinational technology company known for its software, services and devices, most famously the Windows operating system.

Microsoft Azure: Microsoft Azure is a cloud computing platform and infrastructure created by Microsoft. It offers services such as computing power, storage and networking.

Midjourney: Midjourney is a company that specializes in developing advanced technologies and solutions in the realm of AI and machine learning, focussing on image generation.

Misinformation: Incorrect or misleading information that is not created with the intention of causing harm.

Mosaic (web browser): An early and important program for looking at web pages, launched in the early 1990s. It was one of the first web browsers that showed images and text on the same page.

Multimodal: Referring to systems that can handle different types of data, like text, images, audio and video, all at once.

Natural Language: Natural language refers to any language that has developed naturally in humans and is the way we communicate with each other verbally or in written form.

Natural Language Processing (NLP): This is a field of AI that focuses on helping computers understand and respond to human language as it is spoken or written.

Netscape: Netscape was a brand name associated with the development of the Netscape Navigator web browser.

Off The Easel (social bots): 'Off The Easel' refers to a group of social media bots, particularly on Twitter, that share content related to art and cultural heritage.

Open Source: Open source refers to something, typically software, where the source code is freely available for anyone to view, modify, and distribute.

OpenAI: OpenAI is an AI research lab and the company that produced the GPT AI models, including ChatGPT.

PaaS (Platform as a Service): PaaS is a cloud computing service model that provides a platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining the infrastructure typically associated with developing and launching an app.

Produser: An individual user of an online platform who simultaneously plays the role of both content producer and consumer. It highlights where traditional distinctions between producers (those who create content) and users (those who consume it) have blurred.

Prompt Engineering: The practice of crafting inputs and instructions in natural language to effectively interact with AI systems to produce desired outcomes.

RDF (Resource Description Framework): A standard framework structuring information and relationships within data using a set of rules. RDF uses a simple structure called "triples," which are made up of a subject, predicate, and object, much like a sentence in a language that has a subject, verb, and object.

Recommendation Algorithms: Algorithms that predict and suggest items to users based on past behaviour, preferences, or other criteria.

Recommender Systems: A recommender system is a type of information filtering system that seeks to predict and display preferences or recommendations, often algorithmically, that are likely to be of interest to the user.

Reddit: Reddit is a social news aggregation, web content rating, and discussion website where registered members submit content which is then voted up or down by other members.

SaaS (Software as a Service): SaaS is a way of delivering applications over the internet, as a service. It means users can access software and its functions remotely through the internet.

Scholarly Networks: Platforms or systems connecting researchers and scholars to collaborate and share knowledge.

Script (programming): In programming, a script is a series of commands that are executed by a computer to automate tasks.

Search Engine: A search engine is a software system that is designed to carry out web searches, which means to search the World Wide Web in a systematic way for information specified in a textual search query.

Semantic Web: An extension of the World Wide Web that aims to make internet data machine-readable to enable computers to understand and respond to complex human requests.

Social Media: Social media platforms are websites and applications that enable users to create and share content or to participate in social networking.

Social Network: A social network is a platform or service where people connect, interact, and share information with others online.

Sociotechnical Ecosystem: An approach to highlighting and seeking to better understand the complex interactions between humans and technology.

SPARQL: SPARQL (SPARQL Protocol and RDF Query Language) is a specialised programming language used to query databases stored in the Resource Description Framework (RDF) format, such as those holding Linked Open Data.

Stable Diffusion (company): Stable Diffusion is a technology company known for its work in machine learning and generative models, especially for image generation and enhancement.

Style Transfer: In AI and computer graphics, style transfer is a technique that applies the style of one image, such as an artwork, to another.

Synthetic Data: This is data that's artificially generated rather than obtained by direct measurement, often used for testing or training AI models.

Training Data: Training data is the dataset used to 'teach' or 'train' a machine learning model so that it can understand and perform specific tasks.

Transformer Models: A type of neural network model that uses an attention mechanism to improve the training process for tasks involving sequences, such as language processing.

Tumblr: Tumblr is a microblogging and social networking website where users can post multimedia and other content to a short-form blog.

Twitter: See X (Twitter)

Twitterbot: A Twitterbot is a type of bot software that automatically posts content, interacts with users, or performs actions on the Twitter social media platform.

UNESCO: The United Nations Educational, Scientific and Cultural Organization is a specialised agency of the United Nations that aims to promote world peace and security through international collaboration in education, the sciences, and culture.

URI (Uniform Resource Indicator): an address that points to a piece of information anywhere on the internet—not just web pages, but also images, videos or files.

URL (Uniform Resource Locator): a web address typed into a browser to visit a specific web page or image. It directs you to exactly where that item is located on the Web.

UX (User Experience): UX is how a person feels when interacting with a system, like a website or software. It includes the practical, experiential, and meaningful aspects of human-computer interaction.

VCA (Virtual Conversational Assistant): A Virtual Conversational Assistant (VCA) is a computer programme that can understand and respond to spoken or

written language from a person, and is designed to carry out tasks like answering questions or even controlling smart devices in your home.

Web Crawling: This is the process used by search engines to collect data from the internet by systematically browsing the web.

Web (World Wide Web): The World Wide Web, commonly known as the Web, is an information system where documents and other web resources are identified by Uniform Resource Locators (URLs), which may be linked via HTML, and are accessible over the Internet

Web 2.0: Web 2.0 refers to the second generation of the World Wide Web that is focused on the ability for people to collaborate and share information online. Web 2.0 is characterised by greater user interactivity and collaboration, often through wikis and social media platforms.

Web 3.0: Web 3.0 is often associated with the concept of a 'Semantic Web' and is envisioned as the next step in the evolution of the Web. It would be expected to feature more intelligent web search with advanced capabilities such as machine understanding of information.

Web Browser: A web browser is a software application used to access information on the World Wide Web by fetching web pages and displaying them on the user's device.

Wiki: A wiki is a website that allows collaborative editing of its content and structure by its users.

WikiArt: WikiArt is an online, user-editable database of visual arts, including paintings, sculptures, and other art forms.

WikiData: WikiData is a collaboratively edited knowledge base that can be read and edited by humans and machines alike.

Wikimedia: Wikimedia is a global movement whose mission is to bring free educational content to the world through its various projects, including Wikipedia.

Wikipedia: Wikipedia is a free online encyclopaedia that allows users to read and contribute to articles on virtually any topic.

X (Twitter): X (formerly Twitter until 2023) is a social media platform that allows users to post and interact with messages known as 'tweets'. Users can follow others to see their tweets, and engage with them through likes, retweets, and replies.

Bibliography

‘10,000+ Bots for Facebook Messenger’, *ChatBottle* <<https://chatbottle.co/>> [accessed 25 September 2019]

‘(20) Mechanical Curator (@MechCuratorBot) / Twitter’, *Twitter* <<https://twitter.com/mechcuratorbot>> [accessed 28 October 2019]

‘(20) Ukiyo-e Bot (@UkiyoeBot) / Twitter’, *Twitter* <<https://twitter.com/ukiyoebot>> [accessed 28 October 2019]

Abid, Abubakar, Maheen Farooqi, and James Zou, ‘Persistent Anti-Muslim Bias in Large Language Models’, *arXiv:2101.05783 [Cs]*, 2021 <<http://arxiv.org/abs/2101.05783>> [accessed 1 April 2021]

Abokhodair, Norah, Daisy Yoo, and David W. McDonald, ‘Dissecting a Social Botnet: Growth, Content and Influence in Twitter’, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW ’15* (ACM, 2015), pp. 839–51, doi:10.1145/2675133.2675208

‘Acquisition of Twitter by Elon Musk’, *Wikipedia*, 2023 <https://en.wikipedia.org/w/index.php?title=Acquisition_of_Twitter_by_Elon_Musk&oldid=1186289585> [accessed 3 December 2023]

Ada Lovelace Institute, *Regulating AI in the UK*, 2023 <https://www.adalovelaceinstitute.org/wp-content/uploads/2023/07/ADA_Regulating-AI-in-the-UK-Report_July2023_.pdf> [accessed 26 August 2023]

Ahmed, Wasim, Peter A. Bath, and Gianluca Demartini, ‘Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges’, 2017, doi:10.1108/S2398-601820180000002004

Alexandridis, Georgios, Angeliki Chrysanthi, George E. Tsekouras, and George Caridakis, ‘Personalized and Content Adaptive Cultural Heritage Path Recommendation: An Application to the Gournia and Çatalhöyük Archaeological Sites’, *User Modeling and User-Adapted Interaction*, 29.1 (2019), pp. 201–38, doi:10.1007/s11257-019-09227-6

Alexiev, Vladimir, ‘Museum Linked Open Data: Ontologies, Datasets, Projects’, *Digital Presentation and Preservation of Cultural and Scientific Heritage*, VIII, 2018, pp. 19–50

Al-Khalifa, Hend S., and Hugh C. Davis, ‘Towards Better Understanding of Folksonomic Patterns’, in *Proceedings of the Eighteenth Conference on Hypertext and Hypermedia*, HT ’07 (Association for Computing Machinery, 2007), pp. 163–66, doi:10.1145/1286240.1286288

Amazon Alexa Developer Documentation, ‘Certify and Publish Your Skill’, *Amazon (Alexa)* <<https://developer.amazon.com/en-US/docs/alexa/certify/certify-your-skill.html>> [accessed 18 May 2021]

———, ‘What Is the Alexa Skills Kit?’, *Amazon (Alexa)* <<https://developer.amazon.com/en-US/alexa/techdoc-template.html>> [accessed 24 May 2021]

‘Amazon Alexa Voice AI | Alexa Developer Official Site’, *Amazon (Alexa)* <<https://developer.amazon.com/en-US/alexa.html>> [accessed 24 May 2021]

‘Amazon Echo & Alexa Stats’, *Voicebot.Ai*, 2017 <<https://voicebot.ai/amazon-echo-alexa-stats/>> [accessed 22 October 2019]

‘Amazon.Com: American History Quiz: Alexa Skills’ <<https://www.amazon.com/dp/B01F601P7A/?ref=chatbottle>> [accessed 8 October 2019]

‘Amazon.Com: Black History Facts: Alexa Skills’ <<https://www.amazon.com/dp/B01MYC6MJG/?ref=chatbottle>> [accessed 2 October 2019]

‘Amazon.Com: Colonial History: Alexa Skills’ <<https://www.amazon.com/dp/B01LX57ZYF/?ref=chatbottle>> [accessed 2 October 2019]

‘Amazon.Com: Fake History: Alexa Skills’ <<https://www.amazon.com/Tellables-Fake-History/dp/B074DKP2BK?ref=chatbottle>> [accessed 2 October 2019]

‘Amazon.Com: Hardcore History: Alexa Skills’ <<https://www.amazon.com/dancarlin-com-Hardcore-History/dp/B01N8TOMI9?ref=chatbottle>> [accessed 2 October 2019]

‘Amazon.Com: History Facts: Alexa Skills’ <<https://www.amazon.com/dp/B01NB9Y3C5/?ref=chatbottle>> [accessed 8 October 2019]

‘Amazon.Com: History Trivia: Alexa Skills’ <<https://www.amazon.com/sawjax-History-trivia/dp/B0754Q8MML?ref=chatbottle>> [accessed 2 October 2019]

‘Amazon.Com: This Day in History Top Story: Alexa Skills’ <<https://www.amazon.com/This-Day-History-Top-Story/dp/B06XH9VWV5?ref=chatbottle>> [accessed 12 October 2019]

‘Amazon.Com: Today in History: Alexa Skills’ <<https://www.amazon.com/dp/B01EBCW532/?ref=chatbottle>> [accessed 8 October 2019]

‘Amazon.Com: Ultimate HISTORY Quiz: Alexa Skills’ <<https://www.amazon.com/dp/b075zs916k/?ref=chatbottle>> [accessed 8 October 2019]

‘Amazon.Com: US History Quiz: Alexa Skills’ <<https://www.amazon.com/Quyen-US-History-quiz/dp/B07432FYRY?ref=chatbottle>> [accessed 2 October 2019]

‘Amazon.Co.Uk Help: Alexa and Your Privacy’ <<https://www.amazon.co.uk/gp/help/customer/display.html?nodeId=GVP69FUJ48X9DK8V>> [accessed 3 January 2021]

Andrei Taraschuk, ‘Tate’s Collection Includes Hundreds of Sketchbook Pages like This One. I Am Not Sure Whether It Makes Sense to Include Those into the #artbot. What Do You Think?’, *@andreitr*, 2020 <<https://twitter.com/andreitr/status/1266030314108801025>> [accessed 3 March 2021]

‘Andrei Taraschuk on Twitter: “I Stopped Sharing Art on Facebook a While Ago but They Just Restricted My Account for Something @artist_dali Shared in 2017... That Artwork Tho, Lol [Https://T.Co/94Wixwstpv](https://T.Co/94Wixwstpv)” / Twitter’, *Twitter* <<https://twitter.com/andreitr/status/1254645595760783360>> [accessed 22 May 2020]

‘@andreitr/Art Bots on Twitter / Twitter’, *Twitter* <<https://twitter.com/andreitr/lists/art-bots>> [accessed 1 October 2019]

‘@andreitr/Museum Bots on Twitter / Twitter’, *Twitter* <<https://twitter.com/andreitr/lists/museum-bots>> [accessed 1 October 2019]

Ardissono, Liliana, Tsvi Kuflik, and Daniela Petrelli, ‘Personalization in Cultural Heritage: The Road Travelled and the One Ahead’, *User Modeling and User-Adapted Interaction*, 22.1–2 (2012), pp. 73–99, doi:10.1007/s11257-011-9104-x

Arora, A., M. Barrett, E. Lee, E. Oborn, and K. Prince, ‘Risk and the Future of AI: Algorithmic Bias, Data Colonialism, and Marginalization’, *Information and Organization*, 33.3 (2023), p. 100478, doi:10.1016/j.infoandorg.2023.100478

Artemisia Gentileschi, ‘@andreitr Please Help Me to Make This Account Better - Add More Art to Gentileschi’s Wikiart Profile <Https://T.Co/HK1eMdi0k7>’, *@gentileschi_art*, 2019 <https://twitter.com/gentileschi_art/status/1184320197806542848> [accessed 3 March 2021]

Ashmolean Museum, University of Oxford, ‘Ashmolean – The Elements of Drawing, John Ruskin’s Teaching Collection at Oxford’ (Ashmolean Museum, University of Oxford), World <http://ruskin.ashmolean.org/collection/8995/per_page/25/offset/0/sort_by/relevance/object/13559> [accessed 3 March 2021]

Association of Internet Researchers, *Internet Research: Ethical Guidelines 3.0* Association of Internet Researchers, 2019 <<https://aoir.org/reports/ethics3.pdf>> [accessed 5 March 2020]

Ballatore, Andrea, Mark Graham, and Shilad Sen, ‘Digital Hegemonies: The Localness of Search Engine Results’, *Annals of the American Association of Geographers*, 107.5 (2017), pp. 1194–1215, doi:10.1080/24694452.2017.1308240

Barassi, Veronica, and Emiliano Treré, ‘Does Web 3.0 Come after Web 2.0? Deconstructing Theoretical Assumptions through Practice’, *New Media & Society*, 14.8 (2012), pp. 1269–85, doi:10.1177/1461444812445878

Bastos, Marco, and Dan Mercea, ‘The Public Accountability of Social Platforms: Lessons from a Study on Bots and Trolls in the Brexit Campaign’, *Phil. Trans. R. Soc. A*, 376.2128 (2018), p. 20180003, doi:10.1098/rsta.2018.0003

Bastos, Marco T., and Dan Mercea, ‘The Brexit Botnet and User-Generated Hyperpartisan News’, *Social Science Computer Review*, 37.1 (2019), pp. 38–54, doi:10.1177/0894439317734157

Bausch, Paul, Matthew Haughey, Meg Hourihan, Paul Bausch, Matthew Haughey, and Meg Hourihan, *We Blog: Publishing Online with Weblogs* (Wiley, 2002)

<<http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0764549626>> [accessed 11 February 2019]

BBC, 'Twitter Bot Purge Prompts Backlash', 21 February 2018, section Technology <<https://www.bbc.com/news/technology-43144717>> [accessed 15 October 2018]

'BBC History Magazine Quiz: Amazon: Alexa Skills' <https://www.amazon.co.uk/Immedia-Media-Co-History-Magazine/dp/B07D9TXPH7/ref=sr_1_2?dchild=1&keywords=bbc+history+magazine&qid=1622623317&s=digital-skills&sr=1-2> [accessed 2 June 2021]

Beaumont, Romain, 'LAION-5B: A NEW ERA OF OPEN LARGE-SCALE MULTIMODAL DATASETS', *LAION*, 2022 <<https://laion.ai/blog/laion-5b>> [accessed 8 December 2023]

Beer, David, 'The Social Power of Algorithms', *Information, Communication & Society*, 20.1 (2017), pp. 1–13, doi:10.1080/1369118X.2016.1216147

Behera, Himansu Sekhar, Janmenjoy Nayak, Bighnaraj Naik, and Danilo Pelusi, 'Finding Correlation Between Twitter Influence Metrics and Centrality Measures for Detection of Influential Users', in *Computational Intelligence in Data Mining: Proceedings of the International Conference on ICCIDM 2018* (Springer, 2019)

Ben-David, Anat, and Hugo Huirudeman, 'Web Archive Search as Research: Methodological and Theoretical Implications', *Alexandria*, 25.1–2 (2014), pp. 93–111, doi:10.7227/ALX.0022

Benkler, Yochai, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press, 2006)

Bergendorff, Clara Lindh, 'From The Attention Economy To The Creator Economy: A Paradigm Shift', *Forbes*, 2021 <<https://www.forbes.com/sites/claralindhbergendorff/2021/03/12/from-the-attention-economy-to-the-creator-economy-a-paradigm-shift/>> [accessed 3 December 2023]

Berners-Lee, Tim, and Mark Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor* (Harper San Francisco, 1999)

Berners-Lee, Tim, Jim Hendler, and Ora Lassila, 'The Semantic Web', *Scientific American*, 2001

Bijker, Wiebe E., Thomas Parke Hughes, and Trevor Pinch, *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (MIT Press, 2012)

'Black History Facts Chatbot on BotList' <<https://botlist.co/bots/black-history-facts>> [accessed 14 February 2019]

Bloch, Francis, Matthew O. Jackson, and Pietro Tebaldi, 'Centrality Measures in Networks', 2016 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2749124> [accessed 9 June 2017]

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, 'Fast Unfolding of Communities in Large Networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2008.10 (2008), p. P10008, doi:10.1088/1742-5468/2008/10/P10008

Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez, 'Recommender Systems Survey', *Knowledge-Based Systems*, 46 (2013), pp. 109–32, doi:10.1016/j.knosys.2013.03.012

Bogle, Ariel, 'New York Times, CNN and Australia's ABC Block OpenAI's GPTBot Web Crawler from Accessing Content', *The Guardian*, 25 August 2023, section Technology <<https://www.theguardian.com/technology/2023/aug/25/new-york-times-cnn-and-abc-block-openais-gptbot-web-crawler-from-scraping-content>> [accessed 15 December 2023]

Boiano, Stefania, Ann Borda, Giuliano Gaia, Stefania Rossi, and Pietro Cuomo, 'Chatbots and New Audience Opportunities for Museums and Heritage Organisations', 2018, doi:10.14236/ewic/EVA2018.33

Bommasani, Rishi, Dilara Soylu, Thomas I. Liao, Kathleen A. Creel, and Percy Liang, 'Ecosystem Graphs: The Social Footprint of Foundation Models' (arXiv, 2023), doi:10.48550/arXiv.2303.15772

Bonacchi, Chiara, Mark Altaweel, and Marta Krzyzanska, 'The Heritage of Brexit: Roles of the Past in the Construction of Political Identities through Social Media', *Journal of Social Archaeology*, 2018, p. 1469605318759713

Bonadio, Enrico, and Luke McDonagh, *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity* (Social Science Research Network, 2 June 2020) <<https://papers.ssrn.com/abstract=3617197>> [accessed 19 August 2020]

Boniface, Michael, Bassem Nasser, Juri Papay, Stephen C. Phillips, Arturo Servin, Xiaoyu Yang, and others, 'Platform-as-a-Service Architecture for Real-Time Quality of Service Management in Clouds', in *2010 Fifth International Conference on Internet and Web Applications and Services*, 2010, pp. 155–60, doi:10.1109/ICIW.2010.91

Borchers, Conrad, Dalia Sara Gala, Benjamin Gilbert, Eduard Oravkin, Wilfried Bounsi, Yuki M. Asano, and others, *Looking for a Handsome Carpenter! Debiasing GPT-3 Job Advertisements* (arXiv, 23 May 2022), doi:10.48550/arXiv.2205.11374

Borda, Ann, 'Engaging Museum Visitors with AI: The Case of Chatbots', *Museums and Digital Culture, Springer Series on Cultural Computing*, 2018 <https://www.academia.edu/40437562/Engaging_Museum_Visitors_with_AI_The_Case_of_Chatbots> [accessed 21 February 2020]

Boshmaf, Yazan, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu, 'The Socialbot Network: When Bots Socialize for Fame and Money', in *Proceedings of the 27th Annual Computer Security Applications Conference on - ACSAC '11* (presented at the the 27th Annual Computer Security Applications Conference, ACM Press, 2011), p. 93, doi:10.1145/2076732.2076746

Bot, The Vinny G., 'Order a Pizza, Through Slack, Today!', *The Vinny G Bot*, 2016 <<https://medium.com/@VinnyGBot/order-a-pizza-through-slack-today-3f371764d417>> [accessed 14 February 2019]

BotFrens, 'BotFrens', 2023 <<https://botfrens.com/>> [accessed 8 December 2023]

'BotList' <<https://botlist.co>> [accessed 25 September 2019]

Bowker, Geoffrey C., and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (MIT Press, 2000)

Bradshaw, Samantha, and Philip N Howard, 'Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation', 2017, p. 37

Brandes, Ulrik, *Network Analysis: Methodological Foundations* (Springer Science & Business Media, 2005)

Brandtzaeg, Petter Bae, and Asbjørn Følstad, 'Why People Use Chatbots', in *Internet Science*, ed. by Ioannis Kompatsiaris, Jonathan Cave, Anna Satsiou, Georg Carle, Antonella Passani, Efstratios Kontopoulos, and others, Lecture Notes in Computer Science (Springer International Publishing, 2017), pp. 377–92

Brezina, Vaclav, Tony McEnery, and Stephen Wattam, 'Collocations in Context: A New Perspective on Collocation Networks', *International Journal of Corpus Linguistics*, 20.2 (2015), pp. 139–73, doi:10.1075/ijcl.20.2.01bre

Brin, Sergey, and Lawrence Page, 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', *Computer Networks and ISDN Systems*, 30.1–7 (1998), pp. 107–17, doi:10.1016/S0169-7552(98)00110-X

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, and others, 'Language Models Are Few-Shot Learners', *arXiv:2005.14165 [Cs]*, 2020 <<http://arxiv.org/abs/2005.14165>> [accessed 1 April 2021]

Bruns, Axel, 'After the "APIcalypse": Social Media Platforms and Their Fight against Critical Scholarly Research', *Information, Communication & Society*, 22.11 (2019), pp. 1544–66, doi:10.1080/1369118X.2019.1637447

———, 'The Future Is User-Led: The Path towards Widespread Produsage', in *FibreCulture Journal* (presented at the PerthDAC: Digital Arts & Culture, Fibreculture Publications, 2008) <http://journal.fibreCulture.org/issue11/issue11_bruns.html> [accessed 11 February 2019]

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, and others, 'Sparks of Artificial General Intelligence: Early Experiments with GPT-4' (arXiv, 2023), doi:10.48550/arXiv.2303.12712

Bucher, Taina, 'About a Bot: Hoax, Fake, Performance Art', *M/C Journal*, 17.3 (2014) <<http://journal.media-culture.org.au/index.php/mcjourn/article/view/814>> [accessed 29 October 2018]

Bukht, Rumana, and Richard Heeks, *Defining, Conceptualising and Measuring the Digital Economy* (Social Science Research Network, 3 August 2017), doi:10.2139/ssrn.3431732

Burke, Robin, 'Hybrid Recommender Systems: Survey and Experiments', *User Modeling and User-Adapted Interaction*, 12.4 (2002), pp. 331–70, doi:10.1023/A:1021240730564

Cahn, Jack, 'CHATBOT: Architecture, Design, & Development' (University of Pennsylvania, 2017)

Candela, Gustavo, Javier Pereda, Dolores Sáez, Pilar Escobar, Alexander Sánchez, Andrés Villa Torres, and others, 'An Ontological Approach for Unlocking the Colonial Archive', *Journal on Computing and Cultural Heritage*, 16.4 (2023), pp. 1–18, doi:10.1145/3594727

Carah, Nicholas, 'Algorithmic Brands: A Decade of Brand Experiments with Mobile and Social Media', *New Media & Society*, 19.3 (2017), pp. 384–400, doi:10.1177/1461444815605463

Carroll, Stephanie Russo, Maui Hudson, Jarita Holbrook, Simeon Materechera, and Jane Anderson, 'Working with the CARE Principles: Operationalising Indigenous Data Governance', 2020
<<https://www.adalovelaceinstitute.org/blog/care-principles-operationalising-indigenous-data-governance/>> [accessed 8 December 2023]

Caruth, N, and S Bernstein, 'Building an On-Line Community at the Brooklyn Museum: A Timeline', 2007
<<https://www.archimuse.com/mw2007/papers/caruth/caruth.html>> [accessed 3 December 2023]

Chabert, Jean-Luc, Évelyne Barbin, and Chris Weeks, *A History of Algorithms: From the Pebble to the Microchip* (Springer, 1999)

Chan, Carylyne, 'The Chatbot Landscape, 2017 Edition', *Medium*, 2017
<<https://medium.com/keyreply/the-chatbot-landscape-2017-edition-ff2e3d2a0bdb>> [accessed 26 May 2020]

'ChatGPT Creator Withholds Latest AI over Fears It's Too Powerful', *The Independent*, 2023 <<https://www.independent.co.uk/tech/openai-gpt-4-facial-recognition-b2377890.html>> [accessed 20 July 2023]

Christiano, Paul F, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei, 'Deep Reinforcement Learning from Human Preferences', in *Advances in Neural Information Processing Systems* (presented at the Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017), xxx, 1–9
<https://proceedings.neurips.cc/paper_files/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html> [accessed 20 July 2023]

Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia, 'Who Is Tweeting on Twitter: Human, Bot, or Cyborg?', in *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10 (ACM, 2010), pp. 21–30, doi:10.1145/1920261.1920265

Clark, Peter, Oyvind Tafjord, and Kyle Richardson, 'Transformers as Soft Reasoners over Language', *arXiv:2002.05867 [Cs]*, 2020 <<http://arxiv.org/abs/2002.05867>> [accessed 16 October 2020]

'Claude Monet @artistmonet' <<https://twitter.com/artistmonet?lang=en>> [accessed 14 February 2019]

Clément, Maxime, and Matthieu J. Guitton, 'Interacting with Bots Online: Users' Reactions to Actions of Automated Programs in Wikipedia', *Computers in Human Behavior*, 50 (2015), pp. 66–75, doi:10.1016/j.chb.2015.03.078

Clough, Paul, Timothy Hill, Monica Lestari Paramita, and Paula Goodale, 'Europeana: What Users Search for and Why', in *Research and Advanced Technology for Digital Libraries*, ed. by Jaap Kamps, Giannis Tsakonas, Yannis Manolopoulos, Lazaros Iliadis, and Ioannis Karydis, Lecture Notes in Computer Science (Springer International Publishing, 2017), pp. 207–19

CMA: Chinese Art, 'Tripod with Cover, Late 1700s [#museumarchive #cmaopenaccess](https://T.Co/6mg3pX3R43) [@cma_chinese](https://T.Co/1DfI0I7P1A)', 2021 <https://twitter.com/cma_chinese/status/1363670251595042818> [accessed 3 March 2021]

CMA: Greek and Roman Art, 'Torso of a Youth, 400-375 BC [#museumarchive #clevelandartmuseum](https://T.Co/IY2Elg4NF2) [@cma_greekroman](https://T.Co/cPcPEy3Tng)', 2021 <https://twitter.com/cma_greekroman/status/1363598191376162819> [accessed 3 March 2021]

CMA: Medieval Art, 'Hours of Queen Isabella the Catholic, Queen of Spain: Fol. 24r, c. 1500 [#museumarchive #clevelandartmuseum](https://T.Co/lz1FTIUiI) [@cma_medieval](https://T.Co/ev1BrwYeau)', 2021 <https://twitter.com/cma_medieval/status/1362514454869663758> [accessed 3 March 2021]

Colby, Kenneth Mark, 'Modeling a Paranoid Mind', *Behavioral and Brain Sciences*, 4.4 (1981), pp. 515–34, doi:10.1017/S0140525X00000030

Coleman, Gabriella, 'Phreaks, Hackers and Trolls: The Politics of Transgression and Spectacle', in *The Social Media Reader*, ed. by Michael Mandiberg (New York University Press, 2012)

Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson, 'Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data', *Journal of Communication*, 64.2 (2014), pp. 317–32, doi:10.1111/jcom.12084

Cook, Katherine, 'EmboDIYing Disruption: Queer, Feminist and Inclusive Digital Archaeologies', *European Journal of Archaeology*, 22.3 (2019), pp. 398–414, doi:10.1017/eaa.2019.23

Crosman, Penny, 'Mad about Erica: Why a Million People Use Bank of America's Chatbot', *American Banker; New York, N.Y.* (14 June 2018) <<http://search.proquest.com/docview/2054327212/abstract/8855C1FD21E0437FPQ/1>> [accessed 14 February 2019]

Cusumano, Michael, 'Cloud Computing and SaaS as New Computing Platforms', *Communications of the ACM*, 53.4 (2010), pp. 27–29, doi:10.1145/1721654.1721667

Dai, Haixing, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, and others, 'AugGPT: Leveraging ChatGPT for Text Data Augmentation' (arXiv, 2023), doi:10.48550/arXiv.2302.13007

Dale, Robert, 'The Return of the Chatbots', *Natural Language Engineering*, 22.5 (2016), pp. 811–17, doi:10.1017/S1351324916000243

Daniel, Florian, Maristella Matera, Vittorio Zaccaria, and Alessandro Dell'Orto, 'Toward Truly Personal Chatbots: On the Development of Custom Conversational Assistants', in *Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services*, SE4COG '18 (ACM, 2018), pp. 31–36, doi:10.1145/3195555.3195563

Dataethics, 'Big Tech Is Cutting Down on Ethics · Dataetisk Tænkehandletank', 2023 <<https://dataethics.eu/big-tech-is-cutting-down-on-ethics/>> [accessed 2 December 2023]

Davis, Clayton Allen, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer, 'BotOrNot: A System to Evaluate Social Bots', in *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion (International World Wide Web Conferences Steering Committee, 2016), pp. 273–74, doi:10.1145/2872518.2889302

De Angelis, Alessio, Fabio Gasparetti, Alessandro Micarelli, and Giuseppe Sansonetti, 'A Social Cultural Recommender Based on Linked Open Data', in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17 (ACM, 2017), pp. 329–32, doi:10.1145/3099023.3099092

De Roure, David, and Pip Willcox, 'Scholarly Social Machines: A Web Science Perspective on Our Knowledge Infrastructure', in *Proceedings of the 12th ACM Conference on Web Science*, WebSci '20 (Association for Computing Machinery, 2020), pp. 250–56, doi:10.1145/3394231.3397915

De Virgilio, Roberto, Fausto Giunchiglia, and L. Tanca, eds., *Semantic Web Information Management: A Model-Based Perspective* (Springer, 2010)

Deladienne, L., and Y. Naudet, 'A Graph-Based Semantic Recommender System for a Reflective and Personalised Museum Visit: Extended Abstract', in *2017 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, 2017, pp. 88–89, doi:10.1109/SMAP.2017.8022674

Deuschel, Tilman, Timm Heuss, Bernhard Humm, and Torsten Frohlich, 'Finding without Searching - A Serendipity-Based Approach for Digital Cultural Heritage', *Digital Intelligence*, 2014, p. 8

D'Ignazio, Catherine, and Lauren F. Klein, *Data Feminism*, Strong Ideas (Penguin Random House, 2023) <<https://www.penguinrandomhouse.com/books/654051/data-feminism-by-catherine-dignazio-and-lauren-f-klein/9780262358538/>>

van Dijck, Jose, *The Culture of Connectivity: A Critical History of Social Media* (Oxford University Press, 2013) <<http://site.ebrary.com/id/10686674>> [accessed 11 February 2019]

van Dijck, José, 'Users like You? Theorizing Agency in User-Generated Content', *Media, Culture & Society*, 31.1 (2009), pp. 41–58, doi:10.1177/0163443708098245

'Discord - Free Voice and Text Chat', *Discord* <<https://discordapp.com>> [accessed 14 February 2019]

Donovan, Kevin, 'The Best of Intentions: Public Access, the Web & the Evolution of Museum Automation', 1997
<https://www.museumsandtheweb.com/mw97/speak/donovan.html?_gl=1*1c718qy*_ga*MTEwNDMzMDIwOC4xNjgzODcxNjkz*_ga_4799M3BP4Z*MTcwMTM3OTM0My4xLjEuMTcwMTM3OTcwNi4wLjAuMA..&_ga=2.85551641.1065217541.1701379344-1104330208.1683871693> [accessed 30 November 2023]

'DPLA Bot @DPLABot' <<https://twitter.com/DPLAbot>> [accessed 14 February 2019]

Drotner, Kirsten, and Kim Christian Schrøder, *Museum Communication and Social Media: The Connected Museum* (Routledge, 2014)

Edwards, Benj, 'ChatGPT Is One Year Old. Here's How It Changed the Tech World.', *Ars Technica*, 30 November 2023 <<https://arstechnica.com/information-technology/2023/11/chatgpt-was-the-spark-that-lit-the-fire-under-generative-ai-one-year-ago-today/>> [accessed 3 December 2023]

'Eggheads' <<https://www.eggheads.org/>> [accessed 13 February 2019]

Ekstrand, Michael D., John T. Riedl, and Joseph A. Konstan, 'Collaborative Filtering Recommender Systems', *Foundations and Trends® in Human–Computer Interaction*, 4.2 (2011), pp. 81–173, doi:10.1561/1100000009

El Baz, Didier, 'Can We Trust Chatbots for Now? Accuracy, Reproducibility, Traceability; a Case Study on Leonardo Da Vinci's Contribution to Astronomy' (arXiv, 2023), doi:10.48550/arXiv.2304.11852

Engler, Tobias H., Patrick Winter, and Michael Schulz, 'Understanding Online Product Ratings: A Customer Satisfaction Model', *Journal of Retailing and Consumer Services*, 27 (2015), pp. 113–20, doi:10.1016/j.jretconser.2015.07.010

Epstein, J, and W. D Klinkenberg, 'From Eliza to Internet: A Brief History of Computerized Assessment', *Computers in Human Behavior*, 17.3 (2001), pp. 295–314, doi:10.1016/S0747-5632(01)00004-8

EQUITY, BRAIN [BRN AI] CODE FOR, 'Chatbot Trends Report 2021', *Medium*, 2021 <<https://chatbotsjournal.com/chatbot-trends-report-2021-b15479c404e4>> [accessed 2 June 2021]

European Commission, *REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the Evaluation of Europeana and the Way Forward* (European Commission, 2018) <<https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1536235661093&uri=COM:2018:612:FIN>> [accessed 23 August 2019]

Evans, David S, 'The Online Advertising Industry: Economics, Evolution, and Privacy', *Journal of Economic Perspectives*, 23.3 (2009), pp. 37–60, doi:10.1257/jep.23.3.37

Ferrara, Emilio, 'Should ChatGPT Be Biased? Challenges and Risks of Bias in Large Language Models' (arXiv, 2023), doi:10.48550/arXiv.2304.03738

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini, 'The Rise of Social Bots', *Communications of the ACM*, 59.7 (2016), pp. 96–104, doi:10.1145/2818717

Ferrer, Xavier, Tom van Nuenen, Jose M. Such, and Natalia Criado, 'Discovering and Categorising Language Biases in Reddit', *arXiv:2008.02754 [Cs]*, 2020 <<http://arxiv.org/abs/2008.02754>> [accessed 15 December 2020]

First Versions, 'Google', *First Versions*, 2023 <<https://www.firstversions.com/2015/02/google-search-engine.html>> [accessed 3 December 2023]

Fitzpatrick, L. Kelly, 'Anatomy of a Museum Twitter Bot', *Berkman Klein Center Collection*, 2017 <<https://medium.com/berkman-klein-center/anatomy-of-a-museum-twitter-bot-2311d81de243>> [accessed 16 November 2018]

Floridi, Luciano, *The Cambridge Handbook of Information and Computer Ethics* (Cambridge University Press, 2010)

Floridi, Luciano, and Massimo Chiriatti, 'GPT-3: Its Nature, Scope, Limits, and Consequences', *Minds and Machines*, 30.4 (2020), pp. 681–94, doi:10.1007/s11023-020-09548-1

Floridi, Luciano, Mariarosaria Taddeo, and Matteo Turilli, 'Turing's Imitation Game: Still an Impossible Challenge for All Machines and Some Judges—An Evaluation of the 2008 Loebner Contest', *Minds and Machines*, 19.1 (2009), pp. 145–50, doi:10.1007/s11023-008-9130-6

Ford, H., E. Dubois, and C. Puschmann, 'Keeping Ottawa Honest—One Tweet at a Time? Politicians, Journalists, Wikipedians and Their Twitter Bots', ed. by P. Howard and S. Woolley, *International Journal of Communication*, 10 (2016), pp. 4891–4914

Ford, Heather, *Writing the Revolution: Wikipedia and the Survival of Facts in the Digital Age* (The MIT Press, 2022)

Franklin, Stan, and Art Graesser, 'Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents', in *Intelligent Agents III Agent Theories, Architectures, and Languages*, ed. by Jörg P. Müller, Michael J. Wooldridge, and Nicholas R. Jennings, Lecture Notes in Computer Science (Springer Berlin Heidelberg, 1997), pp. 21–35

Freberg, Karen, Kristin Graham, Karen McGaughey, and Laura A. Freberg, 'Who Are the Social Media Influencers? A Study of Public Perceptions of Personality', *Public Relations Review*, 37.1 (2011), pp. 90–92, doi:10.1016/j.pubrev.2010.11.001

Frida Kahlo, 'Juan Gris @artist_gris #followart #juangris <https://t.co/3Ik37J3C4y>', @artfridakahlo, 2021 <<https://twitter.com/artfridakahlo/status/1361009060427018249>> [accessed 3 March 2021]

Fuchs, Christian, *Nationalism on the Internet: Critical Theory and Ideology in the Age of Social Media and Fake News* (Routledge, 2019)

Gaia, Giuliano, Stefania Boiano, and Ann Borda, 'Engaging Museum Visitors with AI: The Case of Chatbots', in *Museums and Digital Culture: New Perspectives and Research*, ed. by Tula Giannini and Jonathan P. Bowen, Springer Series on Cultural Computing (Springer International Publishing, 2019), pp. 309–29, doi:10.1007/978-3-319-97457-6_15

Gaia, Giuliano, Stefania Boiano, Jonathan P. Bowen, and Ann Borda, 'Museum Websites of the First Wave: The Rise of the Virtual Museum' (presented at the Proceedings of EVA London 2020, BCS Learning & Development, 2020), pp. 24–31, doi:10.14236/ewic/EVA2020.4

Garber, Megan, 'When PARRY Met ELIZA: A Ridiculous Chatbot Conversation From 1972', *The Atlantic*, 2014
<<https://www.theatlantic.com/technology/archive/2014/06/when-parry-met-eliza-a-ridiculous-chatbot-conversation-from-1972/372428/>> [accessed 14 February 2019]

Giaccardi, Elisa, ed., *Heritage and Social Media: Understanding Heritage in a Participatory Culture*, 1st ed (Routledge, 2012)

Gieck, Robin, Hanna-Mari Kinnunen, Yuanyuan Li, Mohsen Moghaddam, Franziska Pradel, Peter A. Gloor, and others, 'Cultural Differences in the Understanding of History on Wikipedia', in *Designing Networks for Innovation and Improvisation*, ed. by Matthias P. Zylka, Hauke Fuehres, Andrea Fronzetti Colladon, and Peter A. Gloor, Springer Proceedings in Complexity (Springer International Publishing, 2016), pp. 3–12

Gilani, Zafar, Reza Farahbakhsh, Gareth Tyson, Liang Wang, and Jon Crowcroft, 'Of Bots and Humans (on Twitter)', in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17* (presented at the the 2017 IEEE/ACM International Conference, ACM Press, 2017), pp. 349–54, doi:10.1145/3110025.3110090

Giomelakis, Dimitrios, and Andreas A. Veglis, 'Search Engine Optimization', *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*, 2019, pp. 1789–1800, doi:10.4018/978-1-5225-7598-6.ch132

Gittinger, Juli L., *Hinduism and Hindu Nationalism Online* (Routledge, 2018)

Goel, Sharad, Jake M. Hofman, and M. Irmak Sirer, 'Who Does What on the Web: A Large-Scale Study of Browsing Behavior', in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012
<<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4660>> [accessed 26 May 2020]

'Google Home Added 600,000 More U.S. Users in 2018 Than Amazon Echo, But Amazon Echo Dot Is Still the Most Owned Smart Speaker', *Voicebot.Ai*, 2019
<<https://voicebot.ai/2019/03/07/google-home-added-600000-more-u-s-users-in-2018-than-amazon-echo-but-amazon-echo-dot-is-still-the-most-owned-smart-speaker/>> [accessed 25 October 2019]

'Google's Arts & Culture App Taps AI to Transform Your Photos into Works of Art', *VentureBeat*, 2020 <<https://venturebeat.com/2020/04/02/google-arts-culture-app-taps-ai-to-transform-your-photos-into-works-of-art/>> [accessed 6 April 2020]

Gorwa, Robert, *Computational Propaganda in Poland: False Amplifiers and the Digital Public Sphere* — Oxford Internet Institute (2017) <<https://www.ox.ac.uk/blog/computational-propaganda-in-poland-false-amplifiers-and-the-digital-public-sphere/>> [accessed 13 February 2019]

Gorwa, Robert, and Douglas Guilbeault, 'Unpacking the Social Media Bot: A Typology to Guide Research and Policy', *arXiv:1801.06863 [Cs]*, 2018 <<http://arxiv.org/abs/1801.06863>> [accessed 15 August 2018]

'GPT3 Hunt | Prompts, Examples, & Demo's for OpenAI's GPT-3 API.' <<http://www.buildgpt3.com/>> [accessed 28 December 2020]

Graham, Shawn, *An Enchantment of Digital Archaeology: Raising the Dead with Agent-Based Models, Archaeogaming and Artificial Intelligence* (Berghahn Books, 2020)

———, 'The Resurrection of Flinders Petrie', 2019 <<https://electricarchaeology.ca/2019/10/23/the-resurrection-of-flinders-petrie/>> [accessed 6 April 2020]

'Great Fire of London', *Wikipedia*, 2019 <https://en.wikipedia.org/w/index.php?title=Great_Fire_of_London&oldid=919888587> [accessed 8 October 2019]

Groves, Michael, and Klaus Mundt, 'Friend or Foe? Google Translate in Language for Academic Purposes', *English for Specific Purposes*, 37 (2015), pp. 112–21, doi:10.1016/j.esp.2014.09.001

Guilbeault, Douglas, 'Automation, Algorithms, and Politics| Growing Bot Security: An Ecological View of Bot Agency', *International Journal of Communication*, 10.0 (2016), p. 19

Hacıgüzeller, Piraye, James Stuart Taylor, and Sara Perry, 'On the Emerging Supremacy of Structured Digital Data in Archaeology: A Preliminary Assessment of Information, Knowledge and Wisdom Left Behind', *Open Archaeology*, 7.1 (2021), pp. 1709–30, doi:10.1515/opar-2020-0220

Hand, David J., 'Principles of Data Mining', *Drug Safety*, 30.7 (2007), pp. 621–22, doi:10.2165/00002018-200730070-00010

Hao, Karen, 'OpenAI Is Giving Microsoft Exclusive Access to Its GPT-3 Language Model', *MIT Technology Review*, 2020 <<https://www.technologyreview.com/2020/09/23/1008729/openai-is-giving-microsoft-exclusive-access-to-its-gpt-3-language-model/>> [accessed 7 December 2023]

Haraway, Donna, 'A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20th Century', in *The International Handbook of Virtual Learning Environments*, ed. by Joel Weiss, Jason Nolan, Jeremy Hunsinger, and Peter

Trifonas (Springer Netherlands, 2006), pp. 117–58, doi:10.1007/978-1-4020-3803-7_4

Harpring, Patricia, 'Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA', *Art Documentation: Journal of the Art Libraries Society of North America*, 29.1 (2010), pp. 67–72, doi:10.1086/adx.29.1.27949541

Harris, John, 'Tim Berners-Lee: "We Need Social Networks Where Bad Things Happen Less"', *The Guardian*, 15 March 2021, section Life and style
<<https://www.theguardian.com/lifeandstyle/2021/mar/15/tim-berners-lee-we-need-social-networks-where-bad-things-happen-less>> [accessed 2 December 2023]

Harrison, Maggie, 'Microsoft Patents AI-Powered Therapy App', *Futurism*, 2023
<<https://futurism.com/microsoft-patent-ai-therapy>> [accessed 18 November 2023]

Hays, Kali, 'Twice as Many Companies Block OpenAI's GPTbot, Other AI Web Crawlers', *Business Insider*, 28 September 2023, section Tech
<<https://www.businessinsider.com/openai-gptbot-ccbot-more-companies-block-ai-web-crawlers-2023-9>> [accessed 2 December 2023]

Hazan, Susan, 'The Dance of the Doppelgängers: AI and the Cultural Heritage Community' (presented at the Proceedings of EVA London 2023, BCS Learning & Development, 2023), pp. 77–84, doi:10.14236/ewic/EVA2023.13

Heikkilä, Melissa, 'Inside a Radical New Project to Democratize AI', *MIT Technology Review*, 2022 <<https://www.technologyreview.com/2022/07/12/1055817/inside-a-radical-new-project-to-democratize-ai/>> [accessed 3 December 2023]

Hendler, Jim, and Tim Berners-Lee, 'From the Semantic Web to Social Machines: A Research Challenge for AI on the World Wide Web', *Artificial Intelligence*, Special Review Issue, 174.2 (2010), pp. 156–61, doi:10.1016/j.artint.2009.11.010

Herman, Laura, 'For Who Page? TikTok Creators' Algorithmic Dependencies', in *IASDR 2023: Life-Changing Design* (presented at the IASDR 2023: Life-Changing Design, Design Research Society, 2023), doi:10.21606/iasdr.2023.576

Hern, Alex, 'New AI Fake Text Generator May Be Too Dangerous to Release, Say Creators', *The Guardian*, 14 February 2019, section Technology
<<https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction>> [accessed 7 December 2023]

'History Buddy Chatbot on BotList' <<https://botlist.co/bots/history-buddy>> [accessed 14 February 2019]

'Home', *Appbly.Com* <<https://appbly-com.weebly.com/>> [accessed 25 October 2019]

Hoppe, Anett, Rudolf Seising, Andreas Nurnberger, and Constanze Wenzel, 'Wisdom - the Blurry Top of Human Cognition in the DIKW-Model?', in *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-2011)* (presented at the 7th conference of the European Society for

Fuzzy Logic and Technology, Atlantis Press, 2011),
doi:10.2991/eusflat.2011.91

‘How to Use Amazon Alexa in Non-Supported Countries’, *The Ambient*, 2019
<<https://www.the-ambient.com/how-to/use-alexa-non-supported-countries-855>>
[accessed 25 October 2019]

Howard, Philip N., Aiden Duffy, Deen Freelon, M. M. Hussain, Will Mari, and Marwa Maziad, *Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?* (Social Science Research Network, 2011)
<<https://papers.ssrn.com/abstract=2595096>> [accessed 13 February 2019]

Howard, Philip N., and Bence Kollanyi, ‘Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum’, *arXiv:1606.06356 [Physics]*, 2016
<<http://arxiv.org/abs/1606.06356>> [accessed 13 February 2019]

———, *Bots, #StrongerIn, and #Brexit: Computational Propaganda During the UK-EU Referendum* (Social Science Research Network, 20 June 2016)
<<https://papers.ssrn.com/abstract=2798311>> [accessed 13 August 2018]

Howard, Philip N., Samuel Woolley, and Ryan Calo, ‘Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration’, *Journal of Information Technology & Politics*, 15.2 (2018), pp. 81–93,
doi:10.1080/19331681.2018.1448735

Howcroft, Debra, ‘After the Goldrush: Deconstructing the Myths of the Dot.Com Market’, *Journal of Information Technology*, 16.4 (2001), pp. 195–204,
doi:10.1080/02683960110100418

Hoy, Matthew B., ‘Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants’, *Medical Reference Services Quarterly*, 37.1 (2018), pp. 81–88,
doi:10.1080/02763869.2018.1404391

‘Internet Archive: About IA’ <<https://archive.org/about/>> [accessed 23 April 2019]

‘Introducing the Bots Landscape: 170+ Companies, \$4 Billion in Funding, Thousands of Bots | VentureBeat’ <https://venturebeat.com/2016/08/11/introducing-the-bots-landscape-170-companies-4-billion-in-funding-thousands-of-bots/?utm_content=buffereddf6&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer> [accessed 26 May 2020]

‘Japanese Woodblock Print Search - Ukiyo-e Search’ <<https://ukiyo-e.org/>> [accessed 2 March 2021]

Jiao, Wenxiang, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu, ‘Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine’ (arXiv, 2023),
doi:10.48550/arXiv.2301.08745

Jones, Nicola, ‘What the OpenAI Drama Means for AI Progress — and Safety’, *Nature*, 623.7989 (2023), pp. 898–99, doi:10.1038/d41586-023-03700-4

Just, Natascha, and Michael Latzer, ‘Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet’, *Media, Culture & Society*, 39.2 (2017), pp. 238–58, doi:10.1177/0163443716643157

Kaiser, Katja, Ina Heumann, Tahani Nadim, Hagit Keysar, Mareike Petersen, Meryem Korun, and others, 'Promises of Mass Digitisation and the Colonial Realities of Natural History Collections', *Journal of Natural Science Collections*, 11 (2023), pp. 13–25

Kaplan, Andreas M., and Michael Haenlein, 'Users of the World, Unite! The Challenges and Opportunities of Social Media', *Business Horizons*, 53.1 (2010), pp. 59–68, doi:10.1016/j.bushor.2009.09.003

Karaman, Svebor, Andrew D. Bagdanov, Lea Landucci, Gianpaolo D'Amico, Andrea Ferracani, Daniele Pezzatini, and others, 'Personalized Multimedia Content Delivery on an Interactive Table by Passive Observation of Museum Visitors', *Multimedia Tools and Applications*, 75.7 (2016), pp. 3787–3811, doi:10.1007/s11042-014-2192-y

Kennedy, Krista A, 'Textual Curators and Writing Machines: Authorial Agency in Encyclopedias, Print to Digital' (University of Minnesota, 2009)

'Kik' <<https://www.kik.com/>> [accessed 14 February 2019]

Kimpton, M, and J Ubois, 'Year by Year: From an Archive of the Internet to an Archive on the Internet', in *Web Archiving*, by J Masanes (Springer-Verlag Berlin Heidelberg, 2006) <<http://dx.doi.org/10.1007/978-3-540-46332-0>> [accessed 23 April 2019]

Klopfenstein, Lorenz, and Alessandro Bogliolo, 'THE QUIZ-MASTER BOT: A PERSISTENT AUGMENTED QUIZ DELIVERED THROUGH ONLINE MESSAGING' (presented at the International Technology, Education and Development Conference, 2017), pp. 9806–11, doi:10.21125/inted.2017.2328

Klopfenstein, Lorenz Cuno, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo, 'The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms', in *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17 (ACM, 2017), pp. 555–65, doi:10.1145/3064663.3064672

Knight, Will, 'A Plague of Dumb Chatbots', *Technology Review*; Cambridge, August 2016, 28

_____, 'Amazon Upgrades Alexa for the ChatGPT Era', *Wired*, 2023 <<https://www.wired.com/story/amazon-upgrades-alex-for-the-chatgpt-era/>> [accessed 1 December 2023]

Krzyzanska, Marta, and Chiara Bonacchi, 'Digital Heritage Research Re-Theorised: Ontologies and Epistemologies in a World of Big Data (2019)', *International Journal of Heritage Studies*, 2019 <https://www.academia.edu/38273188/Digital_heritage_re-theorised_Ontologies_and_epistemologies_in_a_world_of_big_data_2019_> [accessed 4 December 2019]

Lamo, Madeline, and Ryan Calo, *Regulating Bot Speech* (Social Science Research Network, 16 July 2018) <<https://papers.ssrn.com/abstract=3214572>> [accessed 13 February 2019]

Latour, Bruno, and Centre de Sociologie de l'Innovation Bruno LaTour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (OUP Oxford, 2005)

Latzer et al, Michael, *The Economics of Algorithmic Selection on the Internet* (University of Zurich, 2014), doi:<http://dx.doi.org/10.5167/uzh-100400>

Latzko-Toth, Guillaume, 'The Socialization of Early Internet Bots: IRC and the Ecology of Human-Robot Interactions Online', *Socialbots and Their Friends*, 2016, doi:10.4324/9781315637228-10

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, and others, 'The Science of Fake News', *Science*, 359.6380 (2018), pp. 1094–96, doi:10.1126/science.aoa2998

Leistert, Oliver, 'Social Bots as Algorithmic Pirates and Messengers of Techno-Environmental Agency', in *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*, ed. by Jonathan Roberge and Robert Seyfert (Taylor & Francis, 2016)

Leonard, Andrew, *Bots: The Origin of New Species* (Penguin Books Limited, 1998)

Leuf, Bo, and Ward Cunningham, *The Wiki Way: Quick Collaboration on the Web* (Addison-Wesley Professional, 2001)
<http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/020171499X> [accessed 11 February 2019]

Li, Zihao, 'The Dark Side of ChatGPT: Legal and Ethical Challenges from Stochastic Parrots and Hallucination' (arXiv, 2023), doi:10.48550/arXiv.2304.14347

Lightman, Hunter, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, and others, 'Let's Verify Step by Step' (arXiv, 2023), doi:10.48550/arXiv.2305.20050

Lih, Andrew, 'Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Resource', *5th International Symposium on Online Journalism*, 2004, pp. 1–31

Ling, Rich, and Chih-Hui Lai, 'Microcoordination 2.0: Social Coordination in the Age of Smartphones and Messaging Apps', *Journal of Communication*, 66.5 (2016), pp. 834–56, doi:10.1111/jcom.12251

Lipsman, Andrew, Graham Mudd, Mike Rich, and Sean Bruich, 'The Power of "Like": How Brands Reach (and Influence) Fans Through Social-Media Marketing', *Journal of Advertising Research*, 52.1 (2012), pp. 40–52, doi:10.2501/JAR-52-1-040-052

'Loebner Prize', AISB - The Society for the Study of Artificial Intelligence and Simulation of Behaviour <<http://www.aisb.org.uk/events/loebner-prize>> [accessed 14 February 2019]

Lops, Pasquale, Marco de Gemmis, and Giovanni Semeraro, 'Content-Based Recommender Systems: State of the Art and Trends', in *Recommender Systems Handbook*, ed. by Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Springer US, 2011), pp. 73–105, doi:10.1007/978-0-387-85820-3_3

Lucy, Li, and David Bamman, 'Gender and Representation Bias in GPT-3 Generated Stories', in *Proceedings of the Third Workshop on Narrative Understanding*

(presented at the NAACL-NUSE 2021, Association for Computational Linguistics, 2021), pp. 48–55, doi:10.18653/v1/2021.nuse-1.5

‘Magicpedia’ <https://geniimagazine.com/wiki/index.php?title=Main_Page> [accessed 25 October 2019]

Mak, Aaron, ‘When Is Technology Too Dangerous to Release to the Public?’, *Slate Magazine*, 2019 <<https://slate.com/technology/2019/02/openai-gpt2-text-generating-algorithm-ai-dangerous.html>> [accessed 6 April 2019]

Manovich, Lev, ‘What Is Visualisation?’, *Visual Studies*, 26.1 (2011), pp. 36–49, doi:10.1080/1472586X.2011.548488

Marc Chagall, ‘Lid: “Meeting of Isaac and Rebecca” #naïveart #chagall [Https://T.Co/HttCYTP2hh](https://T.Co/HttCYTP2hh)’, [@artistchagall](https://twitter.com/artistchagall/status/1361849925961711618), 2021 <<https://twitter.com/artistchagall/status/1361849925961711618>> [accessed 3 March 2021]

Marden, Julia, Carolyn Li-Madeo, Noreen Whysel, and Jeffrey Edelstein, ‘Linked Open Data for Cultural Heritage: Evolution of an Information Technology’, in *Proceedings of the 31st ACM International Conference on Design of Communication*, SIGDOC ’13 (ACM, 2013), pp. 107–12, doi:10.1145/2507065.2507103

Maus, Gregory, ‘A Typology of Socialbots (Abbrev.)’, in *Proceedings of the 2017 ACM on Web Science Conference - WebSci ’17* (presented at the the 2017 ACM, ACM Press, 2017), pp. 399–400, doi:10.1145/3091478.3098860

Mayer, Allana, ‘Linked Open Data for Artistic and Cultural Resources’, *Art Documentation: Journal of the Art Libraries Society of North America*, 34.1 (2015), pp. 2–14, doi:10.1086/680561

McDonough, Katie, ‘You Cannot Ground Truth the Past: Computational Approaches to Historical Maps’ (2023)

McEnery, Tony, and Andrew Hardie, *Corpus Linguistics: Method, Theory and Practice*, Cambridge Textbooks in Linguistics (Cambridge University Press, 2012)

Megahed, Fadel M., Ying-Ju Chen, Joshua A. Ferris, Sven Knoth, and L. Allison Jones-Farmer, ‘How Generative AI Models Such as ChatGPT Can Be (Mis)Used in SPC Practice, Education, and Research? An Exploratory Study’, *Quality Engineering*, 2023, pp. 1–29, doi:10.1080/08982112.2023.2206479

Mesgari, Mostafa, Chitu Okoli, Mohamad Mehdi, Finn Årup Nielsen, and Arto Lanamäki, “The Sum of All Human Knowledge”: A Systematic Review of Scholarly Research on the Content of Wikipedia.’, *Journal of the Association for Information Science and Technology*, 66.2 (2015), pp. 219–45, doi:10.1002/asi.23172

‘Messenger’, *Facebook* <<https://www.messenger.com/>> [accessed 14 February 2019]

Miao, Fengchun, and Wayne Holmes, *Guidance for Generative AI in Education and Research* (UNESCO, 2023) <<https://unesdoc.unesco.org/ark:/48223/pf0000386693>>

Milmo, Dan, 'The Guardian Blocks ChatGPT Owner OpenAI from Trawling Its Content', *The Guardian*, 1 September 2023, section Technology <<https://www.theguardian.com/technology/2023/sep/01/the-guardian-blocks-chatgpt-owner-openai-from-trawling-its-content>> [accessed 3 September 2023]

Mohan, Sriram, 'Locating the "Internet Hindu": Political Speech and Performance in Indian Cyberspace', *Television & New Media*, 16.4 (2015), pp. 339–45, doi:10.1177/1527476415575491

Møller, Anders Giovanni, Jacob Aarup Dalsgaard, Arianna Pera, and Luca Maria Aiello, 'Is a Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks' (arXiv, 2023), doi:10.48550/arXiv.2304.13861

Montgomery, Blake, 'OpenAI Offers to Pay for ChatGPT Customers' Copyright Lawsuits | Artificial Intelligence (AI) | The Guardian', *The Guardian*, 6 November 2023 <https://amp.theguardian.com.cdn.ampproject.org/v/s/amp.theguardian.com/technology/2023/nov/06/openai-chatgpt-customers-copyright-lawsuits?amp_gsa=1&_js_v=a9&usqp=mq331AQIUAKwASCAAgM%3D#amp_tf=From%20%251%24s&aoh=16993418795294&referrer=https%3A%2F%2Fwww.google.com&share=https%3A%2F%2Fwww.theguardian.com%2Ftechnology%2F2023%2Fnov%2F06%2Fopenai-chatgpt-customers-copyright-lawsuits> [accessed 18 November 2023]

'More than 9 Million Broken Links on Wikipedia Are Now Rescued | Internet Archive Blogs' <<https://blog.archive.org/2018/10/01/more-than-9-million-broken-links-on-wikipedia-are-now-rescued/>> [accessed 23 April 2019]

Motoki, Fabio, Valdemar Pinho Neto, and Victor Rodrigues, 'More Human than Human: Measuring ChatGPT Political Bias', *Public Choice*, 2023, doi:10.1007/s11127-023-01097-2

Mullaney, Thomas S., 'QWERTY in China: Chinese Computing and the Radical Alphabet', *Technology and Culture*, 59.4 (2018), pp. S34–65

Murrieta-Flores, Patricia, Mariana Favila-Vázquez, and Aban Flores-Morán, 'Indigenous Deep Mapping: A Conceptual and Representational Analysis of Space in Mesoamerica and New Spain', in *Making Deep Maps* (Routledge, 2021)

Murthy, Dhiraj, 'Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital', 2016, p. 20

'Museum Bot - About' <<https://www.facebook.com/pg/catalhoyukbot/about/>> [accessed 31 March 2020]

'MuseumGenerator @MuseumGenerator' <<https://twitter.com/MuseumGenerator>>

My Modern Met, 'Frida Kahlo Is an Iconic Artist Whose Legacy Lives on. Learn More about This Legend. [Https://T.Co/f2vR9OZiHk](https://T.Co/f2vR9OZiHk)', @mymodernmet, 2021 <<https://twitter.com/mymodernmet/status/1364805996196069376>> [accessed 3 March 2021]

Natale, Simone, and Andrea Ballatore, 'Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence', *Convergence: The International Journal of Research into New Media Technologies*, 2017, p. 135485651771516, doi:10.1177/1354856517715164

Naughton, John, 'ChatGPT Exploded into Public Life a Year Ago. Now We Know What Went on behind the Scenes | Artificial Intelligence (AI) | The Guardian', *The Guardian*, 9 December 2023 <https://amp-theguardian-com.cdn.ampproject.org/v/s/amp.theguardian.com/commentisfree/2023/dec/09/chatgpt-ai-pearl-harbor-moment-sam-altman?amp_gsa=1&_js_v=a9&usqp=mq331AQIUAKwASCAgM%3D#amp_p_ct=1702590810071&_tf=From%20%251%24s&aoh=17025906609618&eferrer=https%3A%2F%2Fwww.google.com&share=https%3A%2F%2Fwww.theguardian.com%2Fcommentisfree%2F2023%2Fdec%2F09%2Fchatgpt-ai-pearl-harbor-moment-sam-altman> [accessed 15 December 2023]

———, 'The Evolution of the Internet: From Military Experiment to General Purpose Technology', *Journal of Cyber Policy*, 1.1 (2016), pp. 5–28, doi:10.1080/23738871.2016.1157619

Nayyar, Rhea, 'Is This the End of Twitter's Beloved "Art Bots"?' , *Hyperallergic*, 8 February 2023 <<http://hyperallergic.com/799103/twitter-api-beloved-art-bots/>> [accessed 1 December 2023]

Neff, Gina, and Peter Nagy, 'Automation, Algorithms, and Politics| Talking to Bots: Symbiotic Agency and the Case of Tay', *International Journal of Communication*, 10.0 (2016), p. 17

Nguyen, Nhai Thi, and Yeow-Tong Chia, 'Decolonizing Research Imagination: A Journey of Reshaping Research Epistemology and Ontology', *Asia Pacific Education Review*, 24.2 (2023), pp. 213–26, doi:10.1007/s12564-023-09822-8

'#nyplremix: Get Creative With the Public Domain', *The New York Public Library*, 2016 <<https://www.nypl.org/blog/2015/01/06/creativity-public-domain>> [accessed 2 March 2021]

Offert, Fabian, 'On the Concept of History (in Foundation Models)', *IMAGE*, 37.1 (2023), pp. 121–34, doi:<https://doi.org/10.1453/1614-0885-1-2023-15462>

Offert, Fabian, and Peter Bell, 'Generative Digital Humanities', 2020

OpenAI, 'ChatGPT', 2023 <<https://chat.openai.com>> [accessed 26 August 2023]

———, 'ChatGPT Can Now See, Hear, and Speak', 2023 <<https://openai.com/blog/chatgpt-can-now-see-hear-and-speak>> [accessed 29 September 2023]

———, 'ChatGPT-User', *OpenAI Platform*, 2023 <<https://platform.openai.com>> [accessed 2 December 2023]

———, 'Custom Instructions for ChatGPT', 2023 <<https://openai.com/blog/custom-instructions-for-chatgpt>> [accessed 2 September 2023]

———, 'DALL·E 2', 2022 <<https://openai.com/dall-e-2>> [accessed 20 July 2023]

———, 'GPT-4 Technical Report' (arXiv, 2023), doi:10.48550/arXiv.2303.08774

———, 'GPTBot', *OpenAI Platform*, 2023 <<https://platform.openai.com>> [accessed 2 December 2023]

———, 'Introducing ChatGPT', 2022 <<https://openai.com/blog/chatgpt>> [accessed 20 July 2023]

———, 'Microsoft Invests in and Partners with OpenAI to Support Us Building Beneficial AGI', 2019 <<https://openai.com/blog/microsoft-invests-in-and-partners-with-openai>> [accessed 7 December 2023]

———, 'OpenAI API', 2023 <<https://platform.openai.com>> [accessed 26 August 2023]

———, 'Our Approach to AI Safety', 2023 <<https://openai.com/blog/our-approach-to-ai-safety>> [accessed 20 July 2023]

———, 'Plugin Terms', 2023 <<https://openai.com/policies/plugin-terms>> [accessed 2 September 2023]

———, 'Teaching with AI', 2023 <<https://openai.com/blog/teaching-with-ai>> [accessed 1 September 2023]

openarchief, 'The Image Collection of the @MuseumRotterdam Is Offline Due to Ruling of a Court about Copyrights. @oa_mrotterdam Stream Is Modified.',
 @openarchief, 2015
 <<https://twitter.com/openarchief/status/615099232807190528>> [accessed 10 February 2021]

'openArtBrowser', 2023 <<https://openartbrowser.org/en/>> [accessed 20 July 2023]

O'Reilly, Tim, 'What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software', 65, 2007, p. 22

O'Sullivan, James, *The Bloomsbury Handbook to the Digital Humanities* (Bloomsbury Publishing, 2022)

Paananen, Siiri, Mari Suoheimo, and Jonna Häkkilä, 'Decolonizing Design with Technology in Cultural Heritage Contexts - Systematic Literature Review', in [] *With Design: Reinventing Design Modes*, ed. by Gerhard Bruyns and Huixin Wei (Springer Nature, 2022), pp. 1839–55, doi:10.1007/978-981-19-4472-7_119

Painter, Deryc T., Bryan C. Daniels, and Jürgen Jost, 'Network Analysis for the Digital Humanities: Principles, Problems, Extensions', *Isis*, 110.3 (2019), pp. 538–54, doi:10.1086/705532

Paolacci, Gabriele, and Jesse Chandler, 'Inside the Turk: Understanding Mechanical Turk as a Participant Pool', *Current Directions in Psychological Science*, 23.3 (2014), pp. 184–88, doi:10.1177/0963721414531598

Pasikowska-Schnass, Magdalena, and Lim Young-Shin, *Artificial Intelligence in the Context of Cultural Heritage and Museums: Complex Challenges and New Opportunities* (European Parliamentary Research Service, 2023)
 <[https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2023\)747120](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)747120)> [accessed 18 November 2023]

'Paul Cezanne @cezanneart' <<https://twitter.com/cezanneart?lang=en>> [accessed 14 February 2019]

Pavis, Mathilde, and Andrea Wallace, *Response to the 2018 Sarr-Savoy Report: Statement on Intellectual Property Rights and Open Access Relevant to the Digitization and Restitution of African Cultural Heritage and Associated Materials* (Social Science Research Network, 25 March 2019) <<https://papers.ssrn.com/abstract=3378200>> [accessed 28 October 2019]

Pereda, Javier, 'Where Do I Stand? Deconstructing Digital Collections [Research] Infrastructures: A Perspective from Towards a National Collection' (presented at the Open and Engaged 2023: Community over Commercialisation, British Library, 2023), doi:10.23636/9fh3-ze12

Perry, Sara, 'The Enchantment of the Archaeological Record', *European Journal of Archaeology*, 22.03 (2019), pp. 1–18

Petsas, Thanasis, Antonis Papadogiannakis, Michalis Polychronakis, Evangelos P. Markatos, and Thomas Karagiannis, 'Rise of the Planet of the Apps: A Systematic Study of the Mobile App Ecosystem', in *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13 (ACM, 2013), pp. 277–90, doi:10.1145/2504730.2504749

Pfeifle, Anne, 'Alexa, What Should We Do about Privacy: Protecting Privacy for Users of Voice-Activated Devices', *Washington Law Review*, 93 (2018), p. 421

Piscopo, Alessandro, 'Wikidata: A New Paradigm of Human-Bot Collaboration?', *arXiv:1810.00931 [Cs]*, 2018 <<http://arxiv.org/abs/1810.00931>> [accessed 11 October 2018]

Prachi Bhardwaj, Business Insider, 'Fiber optic wires, servers, and more than 550,000 miles of underwater cables: Here's what the internet actually looks like', *Business Insider Deutschland*, 2018 <<https://www.businessinsider.de/how-internet-works-infrastructure-photos-2018-5>> [accessed 19 August 2019]

Prescott, Andrew, 'Mixed Methods and the Digital Humanities', in *Mixed Methods and the Digital Humanities* (Bielefeld University Press, 2023), pp. 27–42, doi:10.1515/9783839469132-004

'Privacy Policy' <<https://twitter.com/content/twitter-com/legal/en/privacy.html>> [accessed 14 October 2019]

Pu, Pearl, Li Chen, and Rong Hu, 'Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art', *User Modeling and User-Adapted Interaction*, 22.4–5 (2012), pp. 317–55, doi:10.1007/s11257-011-9115-7

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever, 'Language Models Are Unsupervised Multitask Learners', 2019

Radziwill, Nicole, and Morgan Benton, 'Evaluating Quality of Chatbots and Intelligent Conversational Agents', p. 21

Ratkiewicz, Jacob, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and others, 'Truthy: Mapping the Spread of Astroturf in Microblog Streams', in *Proceedings of the 20th International Conference Companion on World Wide Web - WWW '11* (presented at the the 20th

international conference companion, ACM Press, 2011), p. 249, doi:10.1145/1963192.1963301

‘RDF 1.1 Primer’ <<https://www.w3.org/TR/rdf11-primer/>> [accessed 26 May 2020]

Riedl, J., and A. Halfaker, ‘Bots and Cyborgs: Wikipedia’s Immune System’, *Computer*, 2012, 79–82

Risam, Roopika, ‘Decolonizing the Digital Humanities in Theory and Practice’, in *The Routledge Companion to Media Studies and Digital Humanities* (Routledge, 2018)

‘Robots.Txt Meant for Search Engines Don’t Work Well for Web Archives | Internet Archive Blogs’ <<https://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/>> [accessed 23 April 2019]

Roussou, Maria, Sara Perry, Akrivi Katifori, Stavros Vassos, Angeliki Tzouganatou, and Sierra McKinney, ‘Transformation through Provocation?’, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19 (Association for Computing Machinery, 2019), pp. 1–13, doi:10.1145/3290605.3300857

Rtology, Eva, ‘Become an Art Curator: How Chat GPT Plugins Are Shaping the Art World’, *Medium*, 2023 <<https://medium.com/data-driven-fiction/become-an-art-curator-how-chat-gpt-plugins-are-shaping-the-art-world-7e78bebe553d>> [accessed 19 July 2023]

Ruotsalo, Tuukka, Krister Haav, Antony Stoyanov, Sylvain Roche, Elena Fani, Romina Deliai, and others, ‘SMARTMUSEUM: A Mobile Recommender System for the Web of Data’, *Journal of Web Semantics*, 20 (2013), pp. 50–67, doi:10.1016/j.websem.2013.03.001

Samaroudi, Myrsini, Karina Rodriguez Echavarria, and Lara Perry, ‘Heritage in Lockdown: Digital Provision of Memory Institutions in the UK and US of America during the COVID-19 Pandemic’, *Museum Management and Curatorship*, 35.4 (2020), pp. 337–61, doi:10.1080/09647775.2020.1810483

Sample, Mark, ‘A Protest Bot Is a Bot so Specific You Can’t Mistake It for Bullshit’, *Medium*, 2015 <<https://medium.com/@samplereality/a-protest-bot-is-a-bot-so-specific-you-cant-mistake-it-for-bullshit-90fe10b7fbaa>> [accessed 27 March 2020]

Sanchez Bocanegra, Carlos Luis, Jose Luis Sevillano Ramos, Carlos Rizo, Anton Civit, and Luis Fernandez-Luque, ‘HealthRecSys: A Semantic Content-Based Recommender System to Complement Health Videos’, *BMC Medical Informatics and Decision Making*, 17.1 (2017), p. 63, doi:10.1186/s12911-017-0431-7

Savage, Saiph, Andres Monroy-Hernandez, and Tobias Höllerer, ‘Botivist: Calling Volunteers to Action Using Online Bots’, in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW ’16 (ACM, 2016), pp. 813–22, doi:10.1145/2818048.2819985

Schmidhuber, Juergen, '2006: Celebrating 75 Years of AI - History and Outlook: The Next 25 Years', *arXiv:0708.4311 [Cs]*, 2007 <<http://arxiv.org/abs/0708.4311>> [accessed 11 October 2018]

Schweibenz, Werner, 'The Work of Art in the Age of Digital Reproduction', *Museum International*, 70.1–2 (2018), pp. 8–21, doi:10.1111/muse.12189

Sengupta, Anasuya, and Brooke A. Ackerly, 'Wikipedia Edit-A-Thons: Sites of Struggle, Resistance, and Responsibility', *PS: Political Science & Politics*, 55.2 (2022), pp. 434–38, doi:10.1017/S1049096521001220

Sequeda, Juan, Dean Allemang, and Bryon Jacob, 'A Benchmark to Understand the Role of Knowledge Graphs on Large Language Model's Accuracy for Question Answering on Enterprise SQL Databases' (arXiv, 2023), doi:10.48550/arXiv.2311.07509

Seymour, Tom, Dean Frantsvog, and Satheesh Kumar, 'History Of Search Engines', *International Journal of Management & Information Systems (IJMIS)*, 15.4 (2011), pp. 47–58, doi:10.19030/ijmis.v15i4.5799

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, and others, 'Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks', *Genome Research*, 13.11 (2003), pp. 2498–2504

Shao, Chengcheng, Pik-Mai Hui, Lei Wang, Xinwen Jiang, Alessandro Flammini, Filippo Menczer, and others, 'Anatomy of an Online Misinformation Network', *PLoS ONE*, 13.4 (2018), pp. 1–23, doi:10.1371/journal.pone.0196087

Shiningayamwe, Agnes Sofia Mpingana, 'Decolonizing Heritage Management Systems: New Directions in Digital Heritage Management from Namibia', *Conservation and Management of Archaeological Sites*, 0.0 (2023), pp. 1–21, doi:10.1080/13505033.2023.2287894

Shum, Heung-yeung, Xiao-dong He, and Di Li, 'From Eliza to Xiaolce: Challenges and Opportunities with Social Chatbots', *Frontiers of Information Technology & Electronic Engineering*, 19.1 (2018), pp. 10–26, doi:10.1631/FITEE.1700826

Sinclair, Stefan, and Geoffrey Rockwell, 'Voyant Tools' (Voyant Tools, 2016) <<https://voyant-tools.org/>> [accessed 22 March 2020]

Slack, 'Where Work Happens', *Slack* <<https://slack.com/>> [accessed 14 February 2019]

Smith, Brent, and Greg Linden, 'Two Decades of Recommender Systems at Amazon.Com', *IEEE Internet Computing*, 21.3 (2017), pp. 12–18, doi:10.1109/MIC.2017.72

Smith, Laurajane, *Uses of Heritage*, Reprinted (2009)

Solon, Olivia, 'The Rise of "Pseudo-AI": How Tech Firms Quietly Use Humans to Do Bots' Work', *The Guardian*, 6 July 2018, section Technology <<https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies>> [accessed 3 February 2019]

Spawning.ai, 'What Is Ai.Txt?', *Spawning.Ai*, 2023 <<https://www.spawning.ai>> [accessed 2 December 2023]

Srinivasan, Ramesh, 'Re-Thinking the Cultural Codes of New Media: The Question Concerning Ontology', *New Media & Society*, 15.2 (2013), pp. 203–23, doi:10.1177/1461444812450686

———, *Whose Global Village?: Rethinking How Technology Shapes Our World* (NYU Press, 2018)

Standage, Tom, *The Mechanical Turk: The True Story of the Chess-Playing Machine That Fooled the World* (Penguin, 2003)

'Stefano Della Bella (@artdellabella) / Twitter', *Twitter* <<https://twitter.com/artdellabella>> [accessed 3 March 2021]

Steiner, Thomas, 'Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata', in *Proceedings of The International Symposium on Open Collaboration*, OpenSym '14 (ACM, 2014), p. 25:1-25:7, doi:10.1145/2641580.2641613

Stieglitz, Stefan, Florian Brachten, Björn Ross, and Anna-Katharina Jung, 'Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts', *arXiv:1710.04044 [Cs]*, 2017 <<http://arxiv.org/abs/1710.04044>> [accessed 23 January 2019]

Subrahmanian, V.S., Amos Azaria, Skylar Durst, Vadim Kagan, Aram Galstyan, Kristina Lerman, and others, 'The DARPA Twitter Bot Challenge', *Computer*, 49.6 (2016), pp. 38–46, doi:10.1109/MC.2016.183

SYNCED, 'The Staggering Cost of Training SOTA AI Models', 27 June 2019 <<https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/>> [accessed 7 December 2023]

Tandoc Jr, Edson C., Zheng Wei Lim, and Richard Ling, 'Defining "Fake News"', *Digital Journalism*, 6.2 (2018), pp. 137–53, doi:10.1080/21670811.2017.1360143

Taraschuk, Andrei, 'Bots Going Artsy – Strategies for Sharing Large Art Collections with Social Media Bots and AI' (presented at the SharingisCaring X Stockholm, 2019)

'Telegram – a New Era of Messaging', *Telegram* <<https://telegram.org/>> [accessed 14 February 2019]

Terras, Melissa, 'Opening Access to Collections: The Making and Using of Open Digitised Cultural Content', *Online Information Review*, 39.5 (2015), pp. 733–52, doi:10.1108/OIR-06-2015-0193

The Met: Islamic Art, 'Fragment <https://t.co/eBsTx2aCdr> #metmuseum #themet <https://t.co/g5Eud5tJ6H>', @met_islamicart, 2021 <https://twitter.com/met_islamicart/status/1364204331843649539> [accessed 3 March 2021]

The Met: Medieval Art, 'Glass Fragment [Https://T.Co/Owx45OJ0Qg](https://T.Co/Owx45OJ0Qg) #MedievalArt #metmuseum [Https://T.Co/SHtUxAK1Ng](https://T.Co/SHtUxAK1Ng)', @met_medievalart, 2021

<https://twitter.com/met_medievalart/status/1364062121181540352> [accessed 3 March 2021]

'The Web Robots Pages' <<http://www.robotstxt.org/orig.html>> [accessed 11 April 2019]

Thelwall, Mike, 'Homophily in MySpace', *Journal of the American Society for Information Science and Technology*, 60.2 (2009), pp. 219–31, doi:10.1002/asi.20978

Thielges, Andree, Orestis Papakyriakopoulos, Juan Carlos Medina Serrano, and Simon Hegelich, 'Effects of Social Bots in the Iran-Debate on Twitter', *arXiv:1805.10105 [Cs]*, 2018 <<http://arxiv.org/abs/1805.10105>> [accessed 15 August 2018]

'This Day in History', *HISTORY* <<https://www.history.com/this-day-in-history>> [accessed 27 October 2019]

'This Day in Naval History - Oct. 25' <https://www.navy.mil/search/display_history.asp> [accessed 25 October 2019]

'This Day in Women's History: Amazon.Co.Uk: Alexa Skills' <https://www.amazon.co.uk/This-Day-in-Womens-History/dp/B07CXFX2N7/ref=sr_1_1?keywords=women+in+history&qid=1572770243&s=digital-skills&sr=1-1> [accessed 3 November 2019]

Trichopoulos, Georgios, Markos Konstantakis, George Caridakis, Akrivi Katifori, and Myrto Koukouli, 'Crafting a Museum Guide Using GPT4' (Preprints, 2023), doi:10.20944/preprints202306.1618.v1

'Trove Goldfields News Bot @TroveGoldfields' <<https://twitter.com/trovegoldfields>> [accessed 14 February 2019]

Tsvetkova, Milena, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri, 'Even Good Bots Fight: The Case of Wikipedia', *PLOS ONE*, 12.2 (2017), p. e0171774, doi:10.1371/journal.pone.0171774

'Twice the Number of U.S. Adults Have Tried In-Car Voice Assistants as Smart Speakers - Voicebot.Ai' <<https://voicebot.ai/2019/01/15/twice-the-number-of-u-s-adults-have-tried-in-car-voice-assistants-as-smart-speakers/>> [accessed 25 October 2019]

Tzouganatou, Angeliki, 'Can Heritage Bots Thrive? Toward Future Engagement in Cultural Heritage', *Advances in Archaeological Practice*, 6.4 (2018), pp. 377–83, doi:10.1017/aap.2018.32

Valtysson, Bjarki, 'Europeana', *Information, Communication & Society*, 15.2 (2012), pp. 151–70, doi:10.1080/1369118X.2011.586433

Vandenbergh, Frederic, 'Reconstructing Humans: A Humanist Critique of Actant-Network Theory', *Theory, Culture & Society*, 19.5–6 (2002), pp. 51–67, doi:10.1177/026327602761899147

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, and others, 'Attention Is All You Need' (arXiv, 2017), doi:10.48550/arXiv.1706.03762

Veale, Tony, 'Exploring the Placebo Effect in Computational Creativity', *Proceedings of the Sixth International Conference on Computational Creativity*, 2015

Vincent, James, 'Twitter Taught Microsoft's Friendly AI Chatbot to Be a Racist Asshole in Less than a Day', *The Verge*, 2016
<<https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>>
[accessed 14 February 2019]

'Vincent van Gogh (@vangoghartist) / Twitter', *Twitter*
<<https://twitter.com/vangoghartist>> [accessed 24 May 2020]

Viraf, 'Create A Synthetic Image Dataset — The "What", The "Why" and The "How"', *Medium*, 2020 <<https://towardsdatascience.com/create-a-synthetic-image-dataset-the-what-the-why-and-the-how-f820e6b6f718>> [accessed 12 October 2020]

Viscount Jonkeer. , 'Okay Twitter Is Getting Truly Weird. Apparently They Have Permabanned Medieval Death Bot???? Twitter Really Needs Some Humans to Keep an Eye on Their AI before It Becomes All Powerful...', *@carlclare*, 2021
<<https://twitter.com/carlclare/status/1351831333870444552>> [accessed 3 March 2021]

Vrandecic, D., 'The Rise of Wikidata', *IEEE Intelligent Systems*, 28.4 (2013), pp. 90–95, doi:10.1109/MIS.2013.119

Vrandečić, Denny, and Markus Krötzsch, 'Wikidata: A Free Collaborative Knowledgebase', *Commun. ACM*, 57.10 (2014), pp. 78–85,
doi:10.1145/2629489

Vukovic, D. R., and I. M. Dujlovic, 'Facebook Messenger Bots and Their Application for Business', in *2016 24th Telecommunications Forum (TELFOR)*, 2016, pp. 1–4,
doi:10.1109/TELFOR.2016.7818926

van der Wagen, Wytske, and Wolter Pieters, 'From Cybercrime to Cyborg Crime: Botnets as Hybrid Criminal Actor-Networks', *The British Journal of Criminology*, 55.3 (2015), pp. 578–95, doi:10.1093/bjc/azv009

Wahlster, Wolfgang, Andreas Dengel, Deutsche Telekom, Wahlster Dengel, Wolfgang Wahlster, Andreas Dengel, and others, 'Web 3.0: Convergence of Web 2.0 and the Semantic Web', in *In Technology Radar, Feature Paper, 2nd Ed.; Deutsche Telekom Laboratories*, 2006, pp. 1–23

Wallace, Andrea, 'Decolonization and Indigenization', *Open GLAM*, 2021
<<https://openglam.pubpub.org/pub/decolonization/release/1>> [accessed 27 April 2024]

Wallace, Andrea, and Mathilde Pavis, 'Finding the Nuance in Open Access During the Repatriation of Cultural Heritage' (presented at the Museums+Tech2019: Openness, 2019)

'WeChat for Web' <<https://web.wechat.com/>> [accessed 14 February 2019]

Weidinger, Laura, Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Juan Mateos-Garcia, and others, 'Sociotechnical Safety Evaluation of Generative AI Systems' (arXiv, 2023), doi:10.48550/arXiv.2310.11986

Weilenmann, Alexandra, Thomas Hillman, and Beata Jungsellius, 'Instagram at the Museum: Communicating the Museum Experience Through Social Photo Sharing', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13 (ACM, 2013), pp. 1843–52, doi:10.1145/2470654.2466243

Weizenbaum, Joseph, 'ELIZA—a Computer Program for the Study of Natural Language Communication Between Man and Machine', *Commun. ACM*, 9.1 (1966), pp. 36–45, doi:10.1145/365153.365168

'WhatsApp', *WhatsApp.Com* <<https://www.whatsapp.com/>> [accessed 14 February 2019]

White, Jules, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, and others, 'A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT' (arXiv, 2023), doi:10.48550/arXiv.2302.11382

Whitt, Richard S., and Stephen Schultze, *The New 'Emergence Economics' of Innovation and Growth, and What It Means for Communications Policy* (Social Science Research Network, 1 September 2008)
<<https://papers.ssrn.com/abstract=1311904>> [accessed 28 February 2019]

'Wikidata:WikiProject Sum of All Paintings - Wikidata'
<https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings>
[accessed 16 April 2019]

Wikimedia, 'Future Audiences/Experiments: Conversational/Generative AI', 2023
<https://meta.wikimedia.org/wiki/Future_Audiences/Experiments:_conversational/generative_AI> [accessed 26 August 2023]

Wilkie, Alex, Mike Michael, and Matthew Plummer-Fernandez, 'Speculative Method and Twitter: Bots, Energy and Three Conceptual Characters', *The Sociological Review*, 63.1 (2015), pp. 79–101, doi:10.1111/1467-954X.12168

Wise, J. Macgregor, 'Intelligent Agency', *Cultural Studies*, 12.3 (1998), pp. 410–28, doi:10.1080/095023898335483

'Wookieepedia | FANDOM Powered by Wikia'
<https://starwars.fandom.com/wiki/Main_Page> [accessed 25 October 2019]

Wooldridge, Michael, and Nicholas R. Jennings, 'Intelligent Agents: Theory and Practice', *The Knowledge Engineering Review*, 10.2 (1995), pp. 115–52, doi:10.1017/S0269888900008122

Woolley, Samuel, 'The Political Economy of Bots: Theory and Method in the Study of Social Automation', in *The Political Economy of Robots*, International Political Economy Series (Palgrave Macmillan, Cham, 2018), pp. 127–55, doi:10.1007/978-3-319-51466-6_7

Wu, Steven, and Philip Chua, 'Museum Collection Management On-Demand', in *Proceedings of the 2nd International Conference on Theory and Practice of Electronic Governance*, ICEGOV '08 (Association for Computing Machinery, 2008), pp. 310–15, doi:10.1145/1509096.1509161

Wusteman, Judith, 'RSS: The Latest Feed', *Library Hi Tech*, 22.4 (2004), pp. 404–13, doi:10.1108/07378830410570511

Xie, Jian, Kai Zhang, Jiangjie Chen, Renze Lou, and Yu Su, 'Adaptive Chameleon or Stubborn Sloth: Unraveling the Behavior of Large Language Models in Knowledge Clashes' (arXiv, 2023), doi:10.48550/arXiv.2305.13300

Y (Yiwen), Wang, 'Semantically-Enhanced Recommendations in Cultural Heritage', 2011, doi:10.6100/ir694408

Yan, Nan, and Cheng Xu, 'Decolonizing African NLP: A Survey on Power Dynamics and Data Colonialism in Tech Development', 2024
<<https://openreview.net/forum?id=D698BEfwgv>> [accessed 29 April 2024]

Zeifman, Igal, 'Bot Traffic Report 2016', *Incapsula Blog*, 2017
<<https://www.incapsula.com/blog/bot-traffic-report-2016.html>> [accessed 24 August 2018]

Zhang, Aijia, *Attributing Digitalization Decisions in Museums : A Multiple Case Study of Swedish Public Museums*, 2023
<<https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-510841>> [accessed 29 April 2024]