



# Using Game Engines to Design Digital Workshops for AI Legibility

First name Last name<sup>a</sup>, First name Last name<sup>b</sup>, First name Last name\* **LEAVE BLANK UNTIL FINAL ACCEPTANCE**

<sup>a</sup>Affiliation name **LEAVE BLANK UNTIL FINAL ACCEPTANCE**

<sup>b</sup>Affiliation name **LEAVE BLANK UNTIL FINAL ACCEPTANCE**

\*Corresponding author e-mail: **LEAVE BLANK UNTIL FINAL ACCEPTANCE**

## Abstract:

Like many researchers responding to the pandemic, we have had to adapt design practices traditionally done face-to-face to online experiences. While online services provide adequate support for communication and sharing, they do not readily support the physical tools designed for workshop activities. This paper presents our experience of turning a face-to-face workshop into a digital experience that sustained the primary research goals relating to AI legibility and took advantage of the online world, rather than merely adapting to it, by utilising the game engine *Godot*. This paper explores the theoretical scaffolding that led to the creation of the workshops, which explore AI legibility through iconography and the transition of the workshop experience from face-to-face to online. The workshop's conception followed the original approach of Research through Design and allowed participants to fully engage with our research during the pandemic.

**Keywords:** AI Legibility; Digital Workshops; Game Engines; Iconography

## Introduction

Faced with an increasing inability to run face-to-face workshops, we designed a series of workshops to empirically evaluate AI iconography created to address the challenge of Artificial Intelligence (AI) legibility. The workshops were developed as a playful – Ludic – activity (Huizinga, 1980; Gaver, 2002; Rodriguez, 2006), as the use of play was a way to put participants of all knowledge levels at ease when discussing potentially complex ideas outside their experience (Bogost, 2016). Continuing the research during the pandemic required adapting the workshop from face-to-face to an online experience. After a short investigation of online tools available through third parties, such as *Zoom*, *Miro*, etc., we concluded that none supported the Ludic design of the original workshops and the collection of data rich enough to analyse.

It is important to note that while the paper will feature elements of the data from the digital workshops, the focus is reflecting on redesigning the workshops to maximise the online experience by utilising a game engine. First, the paper provides a brief overview of the theoretical background on AI legibility and the resultant design of the AI iconography. Second, a snapshot of the face-to-face workshop, acting as a blueprint for the digital workshop detailing the workshop's aims and hypothesis of testing the icons. Third, the digital workshop's design and conception through the game-engine Godot will be explained while describing the different workshop exercises and unpacking their various research aims. Fourth, the creation of a custom-made data acquisition tool or 'Analyser'; created to operate in tandem with the digital workshop application to convey a live visual account of the data collected from an 'in progress' workshop to the participants. This section will give an overview of the data from the first iteration of the workshop and how this will inform the design of the second iteration of icons. We will conclude by reflecting on the transition of the workshop when faced with the need to radically reconsider how we do research and the future direction of the research.

## 2. Researching AI legibility through Design

Any discussion of AI with non-experts is challenging as their perception is often heavily influenced by science-fiction renderings of Artificial General Intelligence (AGI) as sentient machines rather than the reality of the Machine Learning we encounter daily in the plethora of AI-infused products such as Alexa and Netflix. This dichotomy has been defined as the 'Definitional Dualism of AI' (Author., 2020), highlighting the misconception AI materialises, exposing the significant knowledge gap between AI-experts and Non-AI-experts (Burrell 2016). This gap is exacerbated through deliberate design choices that further obfuscate the operation of AI in products: with the intent of avoiding overloading the user with auxiliary information by simplifying the interaction to focus solely on the task (Norman, 1998); institutional self-protection in concealing corporate intellectual property; and in some cases the distinct feature of implementing deceptive strategies (Burrell, 2016), for instance, when data is collected without explicit consent (Bridle, 2018) propelling us into a zeitgeist of 'surveillance capitalism' (Zuboff, 2019).

Data is a vital component of any AI discussion as it is the primary driver for the current resurgence of AI research. Therefore, to improve user agency when using or being affected by the consequential results of AI programming requires our relationship with AI technology and data utilisation to be made more legible. Legibility is seen in many frameworks promoting 'better' strategies of implementing AI (Fjeld et al., 2020) and is one of the three challenges raised within the emerging concept of Human Data Interaction, with negotiability and agency completing the triad. Legibility is considered the precursor mechanism to give users the agency to understand data-driven systems and their implications, ensuing support mechanisms for user negotiability to act and re-evaluate their decisions and forge opportunities for change (Mortier et al., 2015). Contemporaneously, the budding research field of 'eXplainable AI' (XAI) has been defined as 'one that produces details or reasons to make its functioning clear or easy to understand' (Arrieta et al., 2020), though as a fledgling concept, its literature is often contradictory. Two examples of contingent XAI principles are; Transparency, making available AI models (Lipton, 2018), though extracted in configurations only legible to AI experts; Interpretability, noted as a 'slippery' term and often confused with transparency (Ibid), has been defined as 'the mapping of an abstract concept into a domain that the human can make sense of' (Montavon et al., 2018), although the method of *mapping* is in the *domain* of an AI-expert, for example using saliency maps to highlight data features utilised by a neural network. While XAI captures the need to unpack the AI black-box, we concur with Gunning et

al. (2019) that the current guidance overlooks end-users' different background knowledge, specifically those with no AI knowledge. In contrast *legibility* is concerned with ensuring AI systems and their operations are understandable for non-AI experts.

To initiate our research, we performed an online survey of AI imagery and found that it lacked effective communication or the semantics to relay the operational remit and reality of the working parameters of AI (Author., 2020). The majority of AI imagery also succumbed to 'AI's dualism', underscoring the need to develop a visual language to enhance AI legibility. Research through Design (Frayling, 1993) (RtD) was the adopted approach as it is inherently generative (Gaver, 2012), allowing us to thread together various hypotheses: such as Semiotics, for example, Peirce's philosophical study on signs (Peirce, 1991); theories such as Human-Computer Interaction relating to both; richer concepts for inter-relationships between users and computer (Bowers & Rodden, 1993; Ferreira et al., 2006); and merging disciplines such as Design and AI, for instance, questioning how design can implement trust in AI services (Arnold et al., 2019) while considering Machine Learning bias (Angwin et al., 2016).

Following the survey, we designed a set of seventeen graphical icons (Figure 01) to individually, and in their respected grouping, detail key operational factors our research identified, such as *Learning Scope* to expose how the AI adapts for better human-AI interaction (Amershi et al., 2019), *Data Provenance* to note the source of training data implementing trust (Arnold et al., 2019), *Processing Location* for the perception of accountability (Rader et al., 2018), and type of *Data Training* used, reducing bias and increasing trust (d'Alessandro et al., 2017). Together, the system of graphical icons in different combinations can map and communicate an AI's 'ontological constitution', making the AI in question more legible to the user (Author., 2021). The icons are abstract in nature which conformed to the principles of semiotics whereby icons hybridise symbolic, indexical and iconic categories together to communicate the intended concept. Though in the instance of a digital thing having no 'conventional representation' (Ferreira et al., 2006), the designs had to materialise the concept or implication being communicated via abstract symbology to develop a convention between reader, icon and concept being communicated (Nakamura & Zeng-Treitler, 2012). Nevertheless, icon designs require some degree of intuitiveness that needs to be embodied in the design through visual cues to interpret (Ferreira et al., 2006).

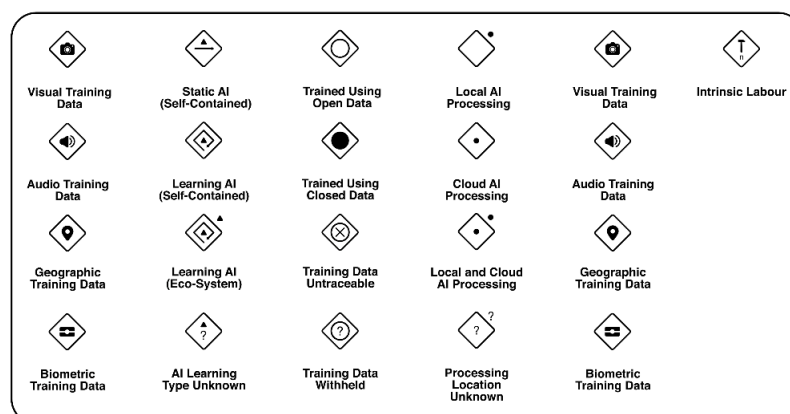


Figure 01. Iteration one of the AI iconography.

### 3. Intuitive Testing through workshops

The next stage of the research was to empirically test the intuitiveness and usability (Ferreira et al., 2006) and, therefore, the legibility of these icons. Barr, Biddle and Noble (2002) explain that an icon works if the user can match the interpretant to the intended object, concept, or implication. This set the precedence for our workshops by designing a range of Ludic exercises whereby participants matched AI functions to the correct icons and offering opportunities to discuss and question the icons' practical use.

For the face-to-face workshops, we had designed a deck of cards depicting either an icon or their associated text descriptors acting as tools to complete game-like exercises and engage in a tangible manner with the intangible operations of AI. We deliberately instigated the idea of embedding Ludic methodologies into the workshop exercises to ignite the participants' 'playful curiosities' (Gaver, 2002) for completing the tasks rather than overloading them with convoluted AI theory. The application of playfulness has been described as 're-ambiguat[ing] the world ... through the characteristics of play, it makes it less formalized, less explained, open to interpretation and wonder and manipulation' (Sicart, 2014) - exactly the experience we wanted for the participants to intuitively test the icons as representations of the AI functions, and if the icons acted as a bridge towards further knowledge about AI.

#### 3.1 The Digital workshop

While the face-to-face workshops acted as a guide as expected, a 'carbon copy' would not be achievable within the constraints of the typical online services (Zoom etc.). The critical factor was the recreation of the 'playful' workshop experience, which serendipitously foregrounded the idea of using a game engine to produce the digital workshop. The workshop was programmed using GDScript, a simplified variant of Python. The code and the associated build of a Graphical User Interface (GUI) that 'acted out' the code's logic was implemented through the open-source game engine *Godot*.

Creating games in *Godot* involves building individual 'scenes' or mini-worlds and then stitching them together through code to 'run' a complete game. Advanced graphics were deemed excessive; therefore, we opted to use the 2D editor in *Godot* to build a 2D GUI with programable building blocks, known as nodes. Specific nodes can be used as direct interactive components of the game, such as sprite nodes and text-editor nodes. The game engine's operating format promoted a design whereby each exercise was a self-contained scene. Hence, each exercise had its own unique coded GUI comprised of curated nodes exclusive to each exercise or the series of mini-research games. As our face-to-face workshop employed a playful game-like interaction with the design of the icon cards and exercises, it was essential to reproduce the notion of digitally handling and moving the replicated icon cards. Thus, the creation of the digital cards was accomplished by importing Portable Network Graphics (.png images) of each icon which 'textured' a sprite node and could then be coded to be manipulated by the user such as move, place, or change colour depending on the task in hand. Building a digital workshop in a game engine offered the unique opportunity to quickly make many iterations and test while still in the design phase.

We opted to package, export, and publish the digital workshop on a dedicated research webpage, thereby preventing the need for participants to download the workshop onto their systems, which would most likely lead to difficulties. We were also keen to facilitate the workshop via conference calls, with participants completing the exercises individually but simultaneously, thus enabling the participants to discuss the recently completed exercise in interceded discussion segments. These

sections were an essential part of the data analysis on why an icon was intuitive or not and to gain qualitative data regarding AI legibility.

### 3.2 Workshop Exercises

The digital workshop consisted of four exercises adapted from the face-to-face workshop; the first exercise *Making Connections*, where participants were individually tasked to intuitively match the digital icon cards to their associated text descriptors (Figure 02). The second exercise, called *What's in My AI?*, presented participants with three moderately speculative scenarios of AI products conducting a distinctive operation (Figure 03). Here, participants tested the icon's concepts by selecting the icons that best described and made legible the functions considered to be transpiring.

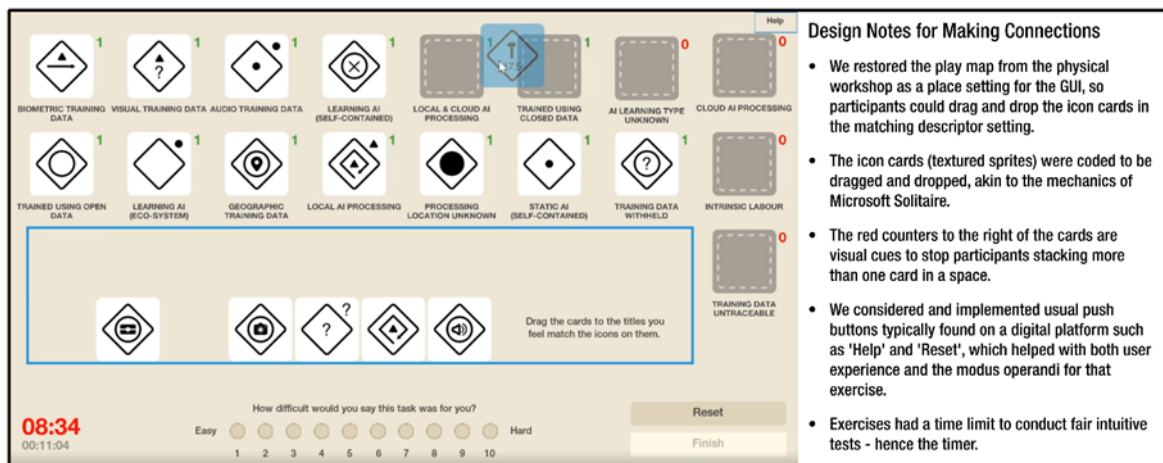


Figure 02: A screenshot of the *Making Connections* exercise with detailed notes regarding the design and build of this digital exercise.

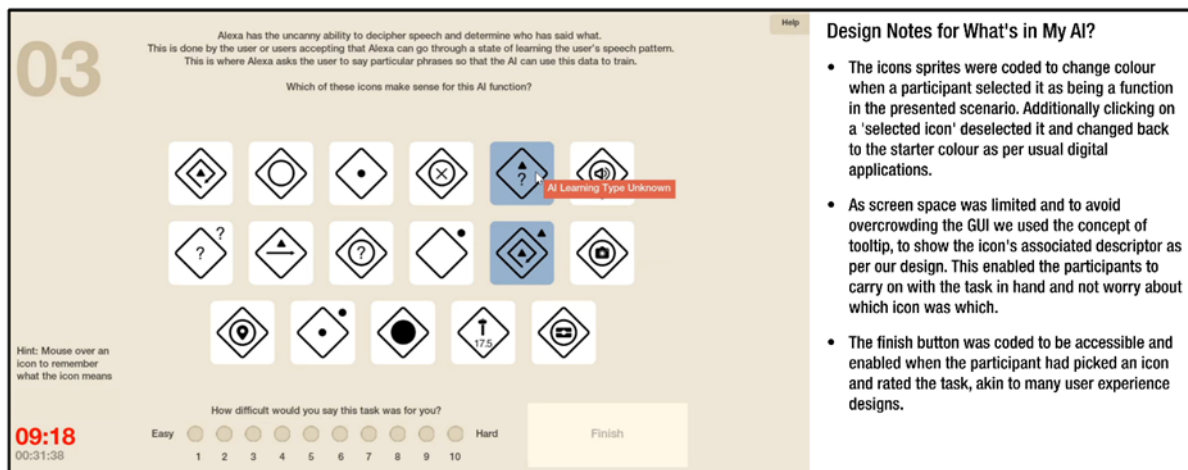


Figure 03: A screenshot of *What's in My AI*.

The third exercise, called *Draw Your Own*, tasked participants to design their own icons using a digital canvas and drawing tools reminiscent of the program *Microsoft Paint* (Figure 04), testing the range of icons and the potential for alternative unaccounted icons.

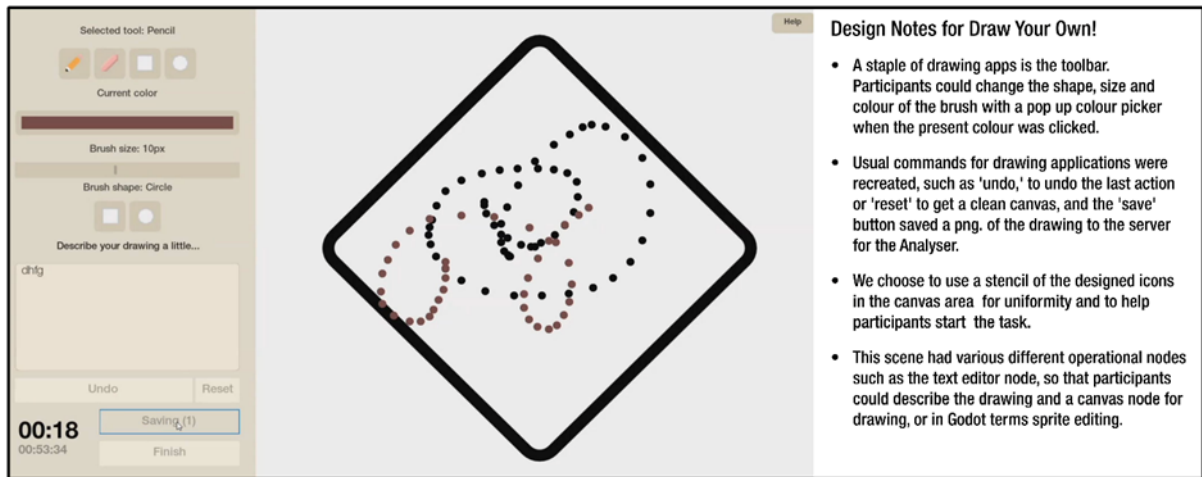


Figure 04: A screenshot of Draw Your Own.

The final exercise called *What's an AI's Intrinsic Labour* allowed participants to hypothesise the meaning of the icon *Intrinsic Labour*. This icon was a speculative icon that attempted to provide a semantic, perhaps tangible interpretation of the unambiguous costs of using AI technology beyond monetary value, for instance, 'how much data would need to be captured from a user for the AI to work efficiently?'. This exercise gave participants an occasion to theorise the current ambiguous impacts of using AI that conceivably need to be made legible (Figure 05).

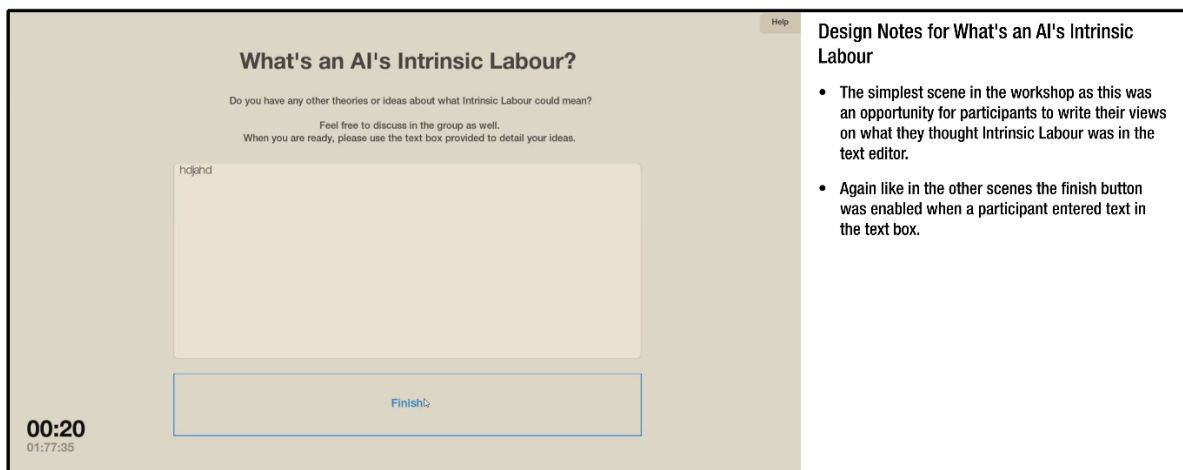


Figure 05: A screenshot of What's an AI's Intrinsic Labour.

## 4. The Analyser

The digital workshop was successful for two reasons; the workshop's design and build, and the instant visual feedback of the participant's exercise results relayed and visualised to the web-based *Analyser*, which doubled as a tool to conduct research analysis after the workshops (Figure 06). Leveraging *Godot's* networking capacity, we connected and synced data to a server, where a *RESTful API* saved and sorted the participants' data into relevant service tables. This enabled a flexible approach for exploration and interpretation to a web-based application in a predetermined and visual manner, corresponding to the research aims, such as the icons displayed in order of most to least correctly matched. The participant's data was displayed onto separate webpages reflective of each exercise. We coded various ways the *Analyser* would display and sort the data during the final analysis, such as 'the most common words associated with intrinsic labour'. By developing our own



data analysis tool, we were able to incorporate it into the workshops so that the participants could be part of the analysis, and therefore the growing conversation on legible AI, strengthening the quality of the data amassed. After the workshops, transferring the data into a third-party analytical tool that would have essentially analysed the data in the same way seemed redundant.



Figure 06: The Analyser on the left was programmed to relay all the participants' icon contributions from the task Draw Your Own in one space, forming an AI icon gallery. On the right is the visual display of the Making Connections exercise.

## 4.1 Results overview towards second iteration

Potential stakeholders, including end-users, academics and industry practitioners, attended the first round of workshops we ran with a total of 47 participants. While it is not the paper's focus, it is important to note that analysing the data from the Analyser and from the discussions, we found the majority of icons could be intuitively deciphered. The *Training Data* and *Processing Location* icons were the most intuitive, with the highest correct matches, and the least intuitive were *Data Provenance* and the *Learning Scope* icons, which were more abstract icons.

The success of the *Training Data* icons was because they utilised both 'iconic' signifiers such as an audio speaker for audio training data and 'symbolic signifiers' such as the geographic pin for geographic data, which over time have been concretised in meaning through use elsewhere. With the *Processing Location* icons, participants commented that they could decipher these as the pattern displayed by the icons suggested importance to position.

We assess that the more abstract icons would, over time, become familiar. However, testing did highlight specific icons and AI terms as misleading and will need to be rectified in the second iteration of the icons. For instance, the term *Data Provenance* was ambiguous as to what the icons were attempting to communicate; therefore, we are working on the idea that this should be changed to *Training Data Origin*. Furthermore, in discussions with participants, another icon was envisioned for the second iteration to communicate an AI is 'trained using user data', which was the overriding concern founded in the workshops.

The icon categories were generally designed to have their own visual and symbolic patterns, with participants using non-verbal reasoning tactics to match. The icons that were often not correctly matched opposed our own implemented pattern and, or, conflicted with the icons intended concept. For instance, the *Static AI* icon (Figure 01) with the arrow implies movement rather than static in

learning. Further, confusion was also noted between *Training Data Untraceable* and *Training Data Withheld* as the cross and the question mark symbol were chosen interchangeably between the two. For the second iteration, these icons will be redesigned, and the intended concept reconsidered as many participants queried what 'untraceable' and 'withheld data' meant and their implications. Testing the icons as we did affirms the need to be legible about how AI functions are communicated, striking a balance between convoluted expert information and information that will improve user agency when using AI technology.

## 5. Conclusion, Reflections and Future Research

The next stage is to design the second iteration of icons and test the icons via an updated version of the workshop. We also plan on extending the workshop to include an exercise that will enable participants to rank the icons questioning what is important for a user to know about their devices and what level of information they require to make a conscious choice.

There are several nuances of designing, facilitating, and working with physical artefacts that are missing from digital fabrication and vice-versa. For example, participants had greater freedom in a physical setting to collaborate, whereas in the digital experience, participants only 'came together' in the discussion sections of the workshop. When comparing the two types of workshops, we found a stark difference between data collection, as it was difficult to record participants results in the face-to-face workshops in a quantifiable manner; in the digital version, data collection was greatly amplified. We can also say the digital workshop had a better opportunity to reach far more participants on a global scale, adding a greater depth to the research and creating perhaps a more holistic test for icon intrusiveness.

This account of transitioning research to a digital experience is not intended to be a prescribed methodology. In our own experience, the state of already having designed and piloted the face-to-face workshop assisted with time and was an invaluable launch pad to begin the process of translating to a digital experience. By employing an RtD approach, we had the liberty to follow the research by creating research artefacts that are not predetermined but shaped by the research. Such as the workshops themselves and the bespoke tools that transcended the capacity to gather data. The intention of circulating our conceivably unique process of utilising a game engine to design and build digital workshops for design research aims to shine a light on one of many distinctive yet off-the-cuff processes for research to continue in the digital realm, coinciding with the need for flexibility in these unpredictable times.

## References

Amershi, S., Weld, D., Vorvoreanu, M., Founney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N.,

Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). *Guidelines for Human-AI Interaction*. 13.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica.

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>



- Arnold, M., Bellamy, R. K. E., Hind, M., Houde, S., Mehta, S., Mojsilovic, A., Nair, R., Ramamurthy, K. N., Reimer, D., Olteanu, A., Piorkowski, D., Tsay, J., & Varshney, K. R. (2019). FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity. *ArXiv:1808.07261 [Cs]*.  
<http://arxiv.org/abs/1808.07261>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gill-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>.
- Barr, P., Noble, J., & Biddle, R. (2002). *Icons R Icons: User interface icons, metaphor and metonymy* (CS-TR-02/20). Victoria University of Wellington School of Mathematical and Computing Sciences Computer Science.
- Bogost, I. (2016). *Play Anything: The pleasure of limits, the uses of boredom, and the secret of games*. Basic Books.
- Bowers, J., & Rodden, T. (1993). Exploding the interface: Experiences of a CSCW network. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*, 255–262.  
<https://doi.org/10.1145/169059.169205>
- Bridle, J. (2018). *New Dark Age, Technology and the End of the Future*. Verso.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 205395171562251. <https://doi.org/10.1177/2053951715622512>
- d'Alessandro, B., O'Neil, C., & LaGatta, T. (2017). Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification. *Big Data*, 5(2), 120–134. <https://doi.org/10.1089/big.2016.0048>
- Ferreira, J., Noble, J., & Biddle, R. (2006). *A Case for Iconic Icons*. 50, 87–90.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*.  
<https://doi.org/10.2139/ssrn.3518482>
- Frayling, C. (1993). *Research in Art and Design*. 1.
- Gaver, W. (2002). DESIGNING FOR HOMO LUDENS. *13 Magazine*.

Gaver, W. (2012). What should we expect from research through design? *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*, 937.

<https://doi.org/10.1145/2207676.2208538>

Huizinga, J. (1980). *Homo Ludens: A Study of the Play-Element in Culture*. (2nd ed.). Redwood Burn Ltd.

Author. (2020). Conference Paper

Lipton, Z. C. (2018, June). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31–57.

Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1–15.

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., & Crowcroft, J. (2015). Human-Data Interaction: The Human Face of the Data-Driven Society. *ArXiv:1412.6159 [Cs]*. <http://arxiv.org/abs/1412.6159>

Nakamura, C., & Zeng-Treitler, Q. (2012). A taxonomy of representation strategies in iconic communication. *International Journal of Human-Computer Studies*, 70(8), 535–551.

<https://doi.org/10.1016/j.ijhcs.2012.02.009>

Norman, D. (1998). *The Invisible Computer: Why Good Products Can Fail, the Personal Computer is So Complex, and Information Appliances are the Solution*. MIT.

Peirce, C. S. (1991). On a New List of Categories. In J. Hoopes (Ed.), *Peirce on Signs* (pp. 23–33). University of North Carolina Press; JSTOR. [http://www.jstor.org/stable/10.5149/9781469616810\\_hoopes.7](http://www.jstor.org/stable/10.5149/9781469616810_hoopes.7)

Author. (2021). Journal Article

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as Mechanisms for Supporting Algorithmic Transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

<https://doi.org/10.1145/3173574.3173677>

Rodriguez, H. (2006). The Playful and the Serious: An approximation to Huizinga's Homo Ludens. *The International Journal of Computer Game Research*, 6(1).

<http://gamestudies.org/0601/articles/rodrigues>

Sicart, M. (2014). *Play Matters*. The MIT Press.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Profile Books Ltd.

Author Bios:

**Author 1** add an Author bio that describes your background and any other achievements in a maximum of 40 words. [LEAVE BLANK UNTIL FINAL ACCEPTANCE] [\_SH Author Bio and Acknowledgements]

**Author 2** add an Author bio that describes your background and any other achievements in a maximum of 40 words.

**Author 3** add an Author bio that describes your background and any other achievements in a maximum of 40 words.

**Acknowledgements:** this section is optional. You can use this section to acknowledge support you have had for your research from your colleagues, student's participation, internal or external partners' contribution or funding bodies. [LEAVE BLANK UNTIL FINAL ACCEPTANCE]