

The Role of HCI in the Age of AI

Richard H.R. Harper^a

^aInstitute for Social Futures, Lancaster University, Lancaster, England

Room C56, InfoLab21, Lancaster University, Lancaster, LA1 4WA, UK.

r.harper@lancaster.ac.uk

Orcid ID: <https://orcid.org/0000-0001-8838-2012>

Richard Harper has written 14 books and collections, including *The Myth of the Paperless Office* (2003), *Texture: human expression in the age of communications overload* (2010) and *Skyping the Family* (2019). He is concerned with all aspects of HCI – from GUI design to systems architecture. He is Co-Director for the Institute of Social Futures (ISF) at the University of Lancaster.

The Role of HCI in the Age of AI

Abstract: This paper examines some of the mystique surrounding AI, including the interrelated notions of explainability and complexity, and argues that these notions suggest that designing human-centred AI is difficult. It explains how, once these are put aside, an HCI perspective can help define interaction between AI and users that can enhance rather than substitute one important aspect of human life: creativity. Key to developing such creative interactions are abstractions and grammars of action and other notions; the paper explores the history of these in HCI and how they are to be used in the contemporary interaction and design space, in relation to AI. The paper is programmatic rather than empirical though its argument uses real-world examples.

Keywords: AI; HCI; explainability; abstractions; grammars of action; creativity; usability.

Preamble

The past few years have shown a society-wide interest in the remarkable developments within machine learning and associated techniques that are enabling what has come to be called the *New AI*. This is said to supplement and even substitute human reasoning, with its powers being amply demonstrated in the capacity of AI machines to beat humans at even the most complex rule-based activities, such as the game of *Go*. In the longer term, AI will be at the heart of self-driving cars, human-less factories and service industries ‘populated’ by artificial assistants¹.

¹ The literature on this is immense and enormously varied. I do not seek to offer a literature review of it all but will point towards what seem to be representative contemporary examples at appropriate stages of my argument. But for a good introduction to the many points of view

The benefits that are seen in this are, of course, immense. But so are the concerns. If robots can do more work will that mean unemployment for the humans who used to be required to do that work, for instance?² In the long run, what will be the effect on human dignity if work is no longer the central currency of identity?³ If robots are more efficient, what will be the measures used to judge investment? Will robots themselves choose where money should go?⁴ More philosophically, if machines are able to reason more effectively than people, what will be the future of learning and further education? Why should society invest in people if machines are better learners? Ultimately will it be machines that do science and win Nobel Prizes for doing so?⁵

Much of these claims are hyperbole, some are simply overexcited. Many of those making the grandest assertions are computer scientists, it is worth noting, and while the excitement they feel about the advances of their field is justified, this does not mean their claims about the wider implications of this technology are well-judged or accurate⁶. Being able to build a *Turing Machine* does not necessarily qualify one on

that is not partisan see, Kaplan, J. (2016) *Artificial Intelligence*. In relation to games like Go, see Sadler and Regan, *Game Changer* (2019).

² See for instance, *The Glass Cage*, Carr (2015).

³ See Markoff's *Machines of Loving Grace* (2015).

⁴ See Kaplan's *Humans Need Not Apply* (2015).

⁵ See Kitano, *Artificial Intelligence to Win the Nobel Prize* (2016).

⁶ See for example Husain's *The Sentient Machine* (2017). Husain is a computer scientist who not unreasonably wants to make a business out of AI technologies but his claims extend well beyond computing. But see also Russell & Norvig's *Artificial Intelligence* (2016). It is just a textbook but its introduction suggests that AI is the most important thing ever invented.

moralties⁷. But by the same token, whilst many other disciplines have explored the implications of AI, very few have done so on the basis of careful examination of what the technology can actually do - the Turing theoretic models that underscore them. Instead, they adopt the excitement exuding from computer science and mix it with their own topics and concerns, creating a melange of claims that are often well removed from algorithms⁸. Meanwhile, government and policy makers hear this cacophony and, quite rightly, have sought to factor AI into their thinking about the future – even though they are all too aware that it is not quite clear what impact the technology will have. Finally, the general public is informed on these various issues by journalists who do not always investigate the claims in question with great care: tales about a robot-controlled future make better copy than one about more efficient production lines. The result of all this is that the true impact of AI is unclear, the hyperbole surrounding it is making careful policy analysis hard, and the full range of consequences that follow on from what the technology will provide remain, in many respects, unexamined. The future of AI, how it affects not only how machines function, what those machines can do, and how, in turn, this alters their role in society more generally, is largely muddled territory⁹.

In this paper, I want to explore what this future might be from a particular view: the view from Human Computer Interaction, HCI. This view focuses on how to design,

⁷ For a criticism of the AI community's understanding of such things as 'concepts', fundamental to understanding human affairs, see Shanker (1998:185-249).

⁸ One philosopher who has sought to be more careful in this regard is Boden (1977, 2016).

⁹ This has been a persistent problem. For example, Stanford University sought to bring clarity to this space with its *AI and Life in 2030* report written in 2015 (published the next year). The muddles it cites are very similar to those I list here, four years later.

evaluate and shape the interaction between computers and humans, and has been at the forefront of ensuring that human endeavour has been supplemented by computing, rather than replaced. This is because it has been especially effective at designing interaction that retains the discretion of the user whilst leveraging the powers of computing¹⁰.

HCI has achieved this through developing views on how people act and understanding how computers function when architected in particular ways. It is a trade that requires, then, both knowledge of the human (and their social practices), *and* knowledge of the digital (the code and the hardware). Its key value, to put it crudely, is grasping what can be done when these are brought together. It is a specialist perspective that allows researchers to shape what a ‘marriage of purposes’ between person and computing can achieve; given as I say, that the goal is always to shape those purposes so as to enable user discretion – to enhance their creativity. This is achieved through making the computer the resource for this creativity; this makes that creativity greater, more astounding. HCI can do this not only in work contexts, but at home, at play and all places in between¹¹.

¹⁰ There are of course many books in the area but I think the best history on why discretion has been so crucial to HCI is Grudin, J. (2017) *From Tool to Partner*. That HCI can nestle with the agenda of AI is clearly a central aspiration of this paper. Some commentators think this is not realistic, though. Indeed, some argue that AI and HCI are inimical. See Grudin again: *AI & HCI: Two fields divided* (2009).

¹¹ There are other perspectives on how to design computer-human systems even within HCI, though for the purposes of this paper I will ignore them. A more important distinction is between HCI and ergonomics which offers different benefits because it has different goals.

As it happens the role and status of HCI researchers has been diminished recently with AI engineers themselves claiming that many of the problems that HCI can solve having been solved (or handled) by such things as ‘natural interaction’ (such as speech-based). As a result of such techniques, it is often AI teams that are coming to design the interaction between person and machine. Though some of the designs that are resulting are innovative, I think in balance this has not resulted in good HCI solutions, when by that I mean interfaces and modes of interaction that lead person and machine (or machines and devices over networks) to produce creative opportunities for the human to achieve new ends. Certainly not in every case, and indeed, in my view, not in most cases¹². To make these interfaces better requires, I think, HCI. Its unique set of

In the case of ergonomics (or human factors as it sometimes known) the ambition is to make the overall person-machine symbiosis as efficient as possible, whatever the role of the human. It is not creativity that matters but optimised efficiency. Of course, in certain situations these concerns are quite close – making the use of hybrid methods appropriate. To be creative with text, for example, presupposes ease of data entry: the ergonomics of a keyboard underscoring the creative affordances of an editing tool. For a third time, Grudin is again good on this: see *Bridging HCI Communities* (2018).

¹² These inadequate interfaces have pedigree going back some years. A canonical example is with the Kinect camera, an AI vision system that was meant to enable natural (body) interaction but instead forced users to move their bodies in peculiar ways. The expectation and disappointment this created was reflected in the high sales of the system at first, the collapse of its sales once users realised how constraining the system was. In my view, this was a missed opportunity - if the technology had been designed from the outset with its affordances being treated as a resource for new, ‘peculiar’ forms of action, new grammars if you like, users may have been more delighted in what it could let them do. But HCI

tools and concerns seem to have been obscured by the excessive claims and excitements provoked by AI, though I think they are there to be used when ‘doing AI’. The purpose of this paper will be to show why this perspective is useful; how and with what key premises.

The paper will be made up of four main sections. In the first, I will note and explore the scope of claims made about AI – about what is intelligence, how machines ‘do it’ and so on. As I say, these are at once startling and so encompassing that they can make consideration of the HCI aspects of AI seem rather minor. They are also, often, rather mystifying – making out that what AI does and how it works so startling that everything we know about the relations between the human species and the tools they invent needing rethinking; we are left mystified at the prospect. But these are just that: *mystifications*; and, in my view, they need dissolving before one can think seriously not just about AI but also about what HCI skills need to be brought to bear to make sure AI technologies provide the usable benefits they surely can¹³. One needs to be aware in

researchers had insufficient role in its development. One reason for this was the mystifying language surrounding the technology suggesting that HCI would not be needed: after all, the system was ‘able to see’ what the user needed. (See Harper & Mentis, 2013; O’Hara et al, 2013). But beyond this, and part of the price paid with this language, is the notion that AI and HCI are inimical to each other: if AI succeeds, one won’t need HCI. But some, like myself, see this as misguided. See also Ren, *Rethinking the Relationship between Humans and Computers*, (2016); see also Ma, *Towards Human-Engaged AI*, (2018).

¹³ That this is so affects all sorts of attempts to explore what AI can do. Some of the better studies from, for example, the social perspective have to work their way around these mystifications before they can find out what the technology does in the real world and its

particular of how these mystifications can encourage us to start worrying about how AI works and not what an AI application can do for us, the user. A desire to comprehend AI draws away from understanding what are the purposes to which AI can be put; and, the paradoxical result is that the intelligent ends that AI can lead us to get obscured by interest in the mechanics of AI. I will tease this out by inverting the normal critique of fictional narratives about artificial intelligence and say, not that they are excessive in their claims about what AI can do, but they can serve to remind us that *doings with AI* are more important than the *mechanics of AI*, or at least should take precedence over the mechanics.

The second will explore a related issue, namely how to get the balance between purposes and means right, by exploring some of the claims made about AI that seem obvious areas where HCI might have had a role, though currently it seems not to. They are the interrelated questions of ‘complexity’ of AI systems and the problem of – or need of – ‘explainability’ partly related to that complexity. In many of the debates and research projects in these areas, HCI tools and techniques have tended to be ignored, the kinds of complexities and the kinds of explanations that AI require being thought to open up wholly new fields of inquiry and design. What is meant by explanation is said to do to with how AI tools are ‘intelligent’ in ways analogous to human intelligence and so need to account for their intelligence in a way a person might; they need to explain their conduct as a child might when rebuked, for example; how an expert explains to a novice, for another¹⁴. In the past computer systems were dumb, so to speak, and

consequences when seen from the social view. See Neyland, *The Everyday life of an Algorithm* (2019).

¹⁴ See for example Miller, (2018) *Explanation in Artificial Intelligence*.

therefore did not need to explain what they did (or how). I will show, however, that the intertwined issues of complexity and explainability have been central to HCI since its inception. The way that these issues have been approached by HCI might have been different, but the kinds of solutions that HCI has offered are applicable to AI.

This will lead me to the next section where I will look at some of the new things that AI tools can do, both in terms of their functioning, and in terms of the user experience they offer. I will be interested in what is called *backpropagation* and *greedy algorithms*, amongst other things, and illustrate what they label and what their use with ‘text-based’ interaction might result in, ‘overfitting’, for instance. Here I will show that such techniques are not to be thought of as black boxes beyond comprehension¹⁵, but rather as ones that need to be handled carefully if they are to deliver benefits to users; sometimes explainability is required, sometimes not; it only matters when explanation is relevant to the user’s purposes. Explanations are never to be thought of as generic, but as pertinent, in my view¹⁶.

In my final section, I will say that one of the key directions of HCI research in the AI space is for HCI researchers to uncover what are the reasons that new tools can offer users, such as facts, understanding, perspectives, and so on. Reasons in this

¹⁵ The use of the term ‘black box’ has become something of a mantra in this field, and not always in ways that are helpful. I have mentioned Neyland in this regard who writes about the mystifying effects of such language. Be that as it may, there are many papers that explore what gets defined as black box AI, distinguishing the sets of techniques deployed in any type, and the approaches to making those techniques ‘explainable’. See, for instance, Guidotti et al, *A Survey of Black Box Methods*, (2019).

¹⁶ This is a point I take from Wittgenstein: *Philosophical Investigations*, (1953).

encompassing sense guide how users might act; given the goal of HCI, they can also enable their discretion. This is where the action is with regard to the *New AI*, in my view – in offering new kinds of reasons. But to understand these without getting distracted and discombobulated is difficult, as the prior parts of the paper will have shown. So I will begin this section with illustrations of the kinds of reasons good search algorithms offer users, and how users have come to account for their use of these reasons in certain kinds of ways bound up with the kinds of grammars of action that search engines afford. I will then say that, as AI tools can come into play in new spaces of human machine interaction, the kinds of reasons they offer will need to be conceived of in terms of new *grammars of action*; as I will have explained by then, what these grammars might be needs sorting out and developing; they will need iteration, design, testing. And central to these are *abstractions* that act as the intermediary between the functioning of the system and the aspirations of the user. There are efforts to develop these, and the example I will take is a tool called LIME that abstracts text and image classifiers¹⁷. As it happens, LIME is far from perfect, as its inventors themselves admit. It could be enhanced with more explorations of its design and visual renderings, testing at greater depth, and further refinement of its scope and functions. All this could lead to new abstractions and even new grammars of action. For this to happen, though, HCI research is required not just in regard to LIME but across the board, in other places where AI could enhance human creativity. Though LIME might not be perfect it points the way to the kinds of research that needs to be done; the kind that can help create the marriage of purposes between user and machine that HCI excels in. In my view, the future is HCI, not AI.

¹⁷ Ribeiro *et al*, (2016) '*Why Should I trust You?*': *Explaining the Predictions of Any Classifier*.

From Norfolk to Silicon Valley

One might start any discussion of AI with the question: what is intelligence? A great deal has been written on this topic. But most of this has confined itself to a very narrow understanding of what the issues are. I alluded above to the idea that intelligence in search might be connected to understanding what some web page says; the kind of understanding that comes from reading it. PageRank does not do this, and so, in this view, is not intelligent in its operation.

Let us leave that line of thought aside for the moment. Most commentaries in the AI and intelligence debates start somewhere else. They presume, for example, that the meaning of intelligence when used in combination with the word artificial points towards ways of calculating, though not so much the kinds of calculations that an abacus can do, as in the calculating that is entailed when people play tightly ruled games – chess, for instance. This leads to research on the most complex game of all, the one mentioned at the outset – *Go*. Such research shows that the calculations are not simply based on the summings of an abacus, but weigh different outcomes given different choices and use learning sets as a reference to select, also probabilistically, ways forward, strategies chosen to succeed in the game. Elaborate statistical techniques are used for this process, most often *Bayesian*, named after the Norfolk vicar who invented them in the eighteenth century. New techniques of engineering have resulted in state-of-the art computers being able to do these sorts of endeavours, calculative, probabilistic, choice-making, in such a fashion that computers win at games when they play people – at *Go*, say. On this basis, some commentators have come to the conclusion that many things will be affected. It won't only be when some games need to be played, it will be whenever some rule-governed conduct is being undertaken; and there are many such activities. Indeed, what can be thought of as game-like can be quite

surprising. AI systems can often ‘see’ things in the visual field, for example, and so can be used to identify objects, even persons. This happens every time someone goes through certain airports, as a case in point. The type of technology, *computer vision*, uses probabilistic techniques to interrogate data it gets from its digital cameras and thereby comes up with labels for what objects are, objects which are defined (or *classified* as it is normally put) through various fairly clever statistics that allow aggregations of colour to be seen as edges, shapes, forms; as *classes*. However clever these techniques and however startling the power of the computers to label one shape over another (and hence one person from another), all they are doing, in effect, is treating that task of identification as one that can be *calculated in the manner of game-play*. The machines are being tasked, via the code, to function like amazingly sophisticated game players, when a game entails subtracting, subdividing and combining data sets in ways determined by elaborate, game-like rules related to the task of recognition. These presuppose what machines are to look for, how they might do this and how they might know when they have seen the things their calculations are designed to recognise – adequate distinctions between John, Fred, Harry, Sandra and Carolina who are queuing up at the passport gates and being seen by the computer system at the same time¹⁸.

Many commentators have started to argue that just as AI machines work this way, so must other ‘machines’ – even biological ones. Some have argued, for example,

¹⁸ One might note that to see, in this view, is not to *know* that it is Harry or Sandra or whoever; recognition is not familiarity, a cue to say ‘Hello!’; on the contrary, it is to behave like a Go player making one play rather than another; there is no interest in what is seen or why it is seen. The goal is to win, when in this case, to win is to recognise the right face.

that mechanisms inside the body are to be thought of as behaving in this sort of way, calculatively, probabilistically, with rules guiding their decision-making as in a game. When a cell confronts another, in this view, its reaction is determined by probability and game-like rules - the cell plays stratagems, so to say. This vision is used to explain how 'communication' between and across cells occurs, and ultimately within any system of cells. From this, these commentators come to assert that the human 'mind', consciousness in particular, emerges; it is the outcome of a vast, intricate system of probabilistically calculated stratagems. In the body these calculations are undertaken by enzymes and chemical processes whereas with an AI machine these calculations are done by logical gates carved in silicon by light; but those who hold this view think of both as more or less the same – the machine and the body. The material of the 'machine' in question is irrelevant; 'doing' intelligent activities in this way is common to one and all. In essence, this is the argument that gets called *singularity*¹⁹. Intelligence, consciousness, choice-making; all this have common roots; if there is a measure of intelligence it relates to this; the latest computers function this way; the age of AI has arrived. We are no different from AI machines, they no different from us²⁰.

¹⁹ This was originally formulated by John Van Neumann but has been popularised by Kurzweil (2005). But see Stanislaw (1958).

²⁰ So, from this view, while we might think of ourselves as singular – that is to say you and I might like to think of ourselves as such, that our minds are ours and ours alone – in fact, if one believes this view, our consciousness is the outcome of millions of little acts, little calculations and stratagems at the cellular (and system) level that produces this sense of self. Our sense of that self is now seen to be egregious. This is the view that Dennett argued for in his *Consciousness Explained* (1991).

There is much dispute about this. The dispute is not about whether cells react probabilistically, in a rule-governed way. It is whether one can say the same about a person calculating in rule-governed ways. Those who don't hold with the singularity argument would say what is meant in each case is *not the same*. To explain that cells behave this way is to account for the *outcomes* of their behaviour, it is not to say that they do, in and of themselves, choose (if one can summarise this view with one word); it is to say that their functioning *can be thought* of as being like that. Whether they choose or not is largely irrelevant. One might say it is a way of describing cells that accounts for their behaviour. In contrast, when one says a person chooses one is saying that they, the person, are aware that they do. In short, they choose to; one cannot say the same of cells or the systems they are part of. They choose nothing²¹.

From cells to Hollywood via soccer

This might seem merely a question of words, of conceptual distinctions that seem minor. But there are important issues here to do with what we mean by choice, and what we mean by the choices that people and machines might make and hence what we mean by intelligence. At the current time the accounts of intelligence that seem to dominate

²¹ This is most eloquently expressed by the physicist, R. Jones, in his (2004) *Soft Machines* - a much better book than Dennett's in my view, since it explores the consequence of this important distinction - the one between description of activities and action that is governed by self-awareness. For those interested in exploring this line of argument, they should go back to Anscombe's *Intention* (1957) which explains how motives distinguish human action. In this view, a machine cannot have a motive, though it might have 'reasons for doing what it does' - such as probabilistic reasons. But for an introduction see Harper *et al*, *Choice*. (op cit).

are of one particular kind, emphasising, as I say, one particular view about intelligence, its rule-like calculations leading to choices. But must we think about intelligence this way? Are we narrowing our notions to just one view? Is the singularity perspective making us think less about what intelligence might be rather than more?

A simple example can help us here. One might say that there are two ways of looking at the game of soccer. One looks at the way the muscles of a player functions when they play. Another looks at the game itself, at the strategies and skills used to win. In my view, those who hold the singularity view are looking at intelligence like those who look at soccer and see muscles acting. In my view, though it is true that muscles need to flex, to understand soccer is to look at the game, not at muscle movement. It is in the game, so to speak, that intelligence is to be found – in how people play it. To see it this way, AI notions about intelligence simply don't help. Indeed, they can make it hard to see – they can take you away from what you need to look at. Instead of letting one see play, the AI looks at muscles. And one doesn't look at muscles to find intelligence²².

This is not to say that one cannot explore what intelligence might be with reference to the artificial. Let me take two examples of artificial creatures as presented to the cultural imagination, one in a book and the other in a movie. What I will want to note is that there are different measures of intelligence applied in each case; and moreover, it is not just that the games that are being played are different (if one can put it that way), but the ways of accounting for success (or failure) are too. Indeed, one

²² This is of course an argument that derives from the ordinary language philosophers,

Wittgenstein (*op cit*) being the most regarded, if not the easiest to read.

might say that even using the game metaphor does not help in properly grasping the problems that the intelligence of these artificial creatures need to attend to. And it is *these* problems that one needs to focus on, not on the mechanics of the intelligence in question that I think is important.

The first artificial being I want to consider is the one imagined by Mary Shelley. It is the creature made by Frankenstein – unnamed in the book, being called the *creature* or the *monster* interchangeably. When she wrote *Frankenstein: or The Modern Prometheus* (1818) she was not thinking of artificial intelligence, not as we know it. Nevertheless, she was certainly wondering about how to make a man²³, and her interests were piqued by the technologies that were advancing at that period of time – in the machinery of clocks especially. So it was a kind of artificial intelligence that she had in mind. Her starting assumption was that one might be at the cusp of techniques that would make the *making of a man* practical. But she was not writing a theory of how man-the-machine might function – in a clock-like way, perhaps. She was writing in reference to other things of concern at that time, and her idea of making a novel about the manufacture of a man was designed as a vehicle to explore these. Central to her concerns were changing attitudes to God. At the time Shelley was writing, at the peak of the Romantic Period, the cultural imagination was focused on how the individual person was *special*. What they could do, because of what was in them – their talents, their skills, their ‘interior life’ - was more interesting than had been realised before; and indeed, altered the relationship to the divine. If, before Rousseau (often said to have invented the romantic era), individuals were thought to be constrained by their social

²³ At this time, the suffix man was meant to encompass all human kind, though whether that assumed all humans were equal is another matter, needless to say.

roles – as King, say, as Pope, as priest or scholar - ; these, though, were subordinate to the providence of God. It was God who gave them these roles, whatever they might be. Now, with romanticism, the idea emerged that the individual had powers and capacities within that broke the constraints of roles – and all the social conventions that went with them including the assumption that God choose a role for one. So great were these interior powers that the relationship between the world, God and the individual ought to change, so this view held. A social contract had to be made. It was this that led to Rousseau's *The Rights of Man*. These were to be contrasted with what had been taken for granted before, namely *The Rights of God* (as articulated in the Bible). It was man that mattered, not man as a focus of the arrangements of God. Shelly was not interested in the fate of faith, in God, so much as what ensues given the loss of faith. If power was now seen in the individual, what was the relationship between one individual and another? And perhaps more importantly, what was the relationship between a person and their sense of themselves? If the inner man was special, how did that make a man feel? Should they worship themselves? They no longer worshipped God, after all. That they might honour themselves seemed outrageous, evidently profane. But how were they to react to this newly realised sense of specialness, if not that way?²⁴

Shelley's fiction explored this. She proposed that when Frankenstein's 'creature' awoke he came to be shocked. We have come to think that it was 'his' ugliness that upset him. This interpretation derives not from the book, though, but from the Bela Legusi Hollywood movies: here the creature is appalled at his looks. As we are, when

²⁴ This is of course a massive simplification of a complex interweaving of ideas and trends; for an excellent overview related to the notion of the individual and the self, see Heehs, P.

(2013) *Writing the Self*.

we watch the films. Indeed, we have become all too familiar with this reading of Frankenstein. But this is not at all what Shelley writes about. In her story, the creature is appalled when he discovers he has been made by another person – it is not vanity about his body that matters, it is his *pride*. The creature asks what rights of Frankenstein were reflected in his manufacture and which of his own, the creature's? What kind of special interiority was he endowed with? He asked all this not because he thought the balance needed to be right – the balance of his own and Frankenstein's rights. He was frightened that his rights were zero: after all, what kind of man was *made by another*? Every man and woman is equal and unique was the presumption of the romantic era and hence in Shelley's narrative, and hence this interiority was not something to be made as if with putty. And yet, through new technology, here was a 'man' made. The creature came to decide that he would murderously pursue Frankenstein as revenge for condemning him, the creature, to being less than a man by being made by a man. This is the arc of Shelley's wonderful book.

I won't say anything more about *Frankenstein: or the Modern Prometheus*. But how very different this view is from how we see human-made, human-like creatures today, the machines that also have an artificial form, but perhaps made in different ways. As we look in our current cultural landscape, it is emphatically not the shocking ugliness that we associate with the artificial; we have moved on from Bella Legusi. For one thing, we imagine our technologies are far more sophisticated than those that Frankenstein had with his lightning-powered butchery; we can now use the almost magical powers of computers to make people. This can ensure that artificial creatures can reason better than us, they can be more handsome, too. They are, in a sense, so much more than us. They are Ava in *Ex Machina* written and directed by Alex Garland (2014) – ethereal, perfect, glorious. And yet here is the rub: the role Alicia Vikander

plays (Ava) is not about contentment. She is not opposite to Frankenstein's creature in being beautiful. For she suffers too. But hers is the suffering known as pathos. In *Ex Machina* it is in the discovery by her, acting as the cyborg Ava, that she might not be 'real' that drives the narrative arc. She has been programmed to think she is, but she slowly comes to realise this is false. She is not *real*. She is AI. If Shelley's monster was trapped by his romantic notion of human dignity – no man can make me – then the modern creature, in *Ex Machina*, is trapped by programmers who have not told 'her' everything. She is being made a fool of by her own code. This humiliation - and our sympathy - is driven by her growing self-awareness. Ultimately our sympathy deepens when we see her realise that she cannot be truly loved; she is loved only to the extent that she looks perfect. What she is within, so to speak, doesn't matter, it isn't even genuine; it is *artificial*. What Frankenstein's monster felt, indignation, righteous anger, is, then, not the same as the doubt and crushing insignificance that the cyborg feels in *Ex Machina*. How could any human choose to love Ava if they knew she was human-made? This is her angst, the angst of a cyborg.

This seems a long way from HCI and indeed it is. So where am I going? I am wanting to highlight the fact that how the creatures in question work, how they function is *not really salient* to the plots. It is what they put their *powers of thought to* that matters. Their intelligence is understood not in terms of how it functions but in terms that are external to the mechanics of that. To understand these two narratives requires us to focus not on how the creature that Frankenstein made works, nor on how Ava works; we need to look at the game(s) they are playing, to pursue my analogy. These games are plots, of course and don't lead to a win or a loss, but they are game-like nonetheless, insofar as they involve stratagems. They are to do with, on the one hand, the death of God and the powers of the inner human soul that must take up the resulting

responsibilities; and, on the other, the nature of love and the need for self-esteem that would emerge if that love is comprehensive – for the person and not just the body; for their entirety and not just for their looks. I am not denying that these concerns are partly bound up with the mechanics of the creatures – Ava’s code does not include scripts that instruct her that she is a cyborg - but it is the pathos of her predicament that is the centre of attention or *ought to be*. That is my point; what I am taking away from this nineteenth century novel and this contemporary movie is a moral. When we think about intelligence, we should ask what we want to do with intelligence more than we ask how it functions, what are its mechanics. When we want to understand a soccer game, we should look at the play, not at the muscles that enable it. So too with AI.

The insides of computers

It seems to me that narratives about AI can sometimes displace sensible discussion about what we want those applications to do. We end up thinking about how AI works (and to how explain *that*) instead of exploring what AI might help *us do*. We are offered muscles and explanations about how they work, if you like, and not games and their various purposes. As a result, we lose sight of why we might want to play the game, of the purposes that would make it worthwhile to play.

Part of the price paid with this narrowing of vision is, in my view, that terms used to describe AI and label related research agendas are themselves muddling, confounding our efforts to see. The terms *complexity* and its correlate, *explainability* are such. When these terms are used one isn’t sure whether we are being asked to think about what goes on inside an AI application or what is done with an application²⁵. And,

²⁵ *Op cit*, Neyland, 2019.

when these are bundled with the term black box, we can then find ourselves beginning to panic – though we might not be sure we really need to understand how AI works, this very term suggests it is beyond our understanding anyway! My view, as should be clear, is that I do not think it always matters – what happens inside a computer. Or rather, I think it might well matter when one is doing HCI, when one is designing the interaction between a computer system, an AI one say - and a user. But I also think that, often, the resulting design is such that understanding of the computer system is no longer required. What happens inside can be put aside. I do not mean it becomes invisible, so much as it is a matter determined by the interactional goals in question.

Let me illustrate by turning to what one might say is some basic features of about computing, and in particular one dimension of ‘computing’ and what the user knows about it. I am interested in what a basic computer, a desktop or a laptop say, does with *files* – presumably an elementary constituent of any computer, basic or otherwise.

A computer has to handle files, obviously. By handle I mean that a computer has a user’s files, so to say, and presumably makes them available for the user as and when desired by the user. The computer stores them on their behalf, if you like. But one might say that, while from a computer’s point of view (if one can put it so), a file is a label for the minimal digital entity that can be ‘persisted’ in its storage system, this isn’t what a user sees or *understands*. For a computer a file is a label for a bundle of data, but what that bundle consists in (i.e., what the binary data represent) doesn’t really matter nor is it implied in the use of the term ‘file’. From the computer’s view, what to a user is, let us say, a love letter, or correspondence from a Bank, are neither of these things; they are instead, the same - just bytes stored somewhere. And they are not even necessarily stored together, but only such that they can be retrieved (and aggregated if required) when asked to. One might add that a basic computer doesn’t only handle files in these

ways. It also chooses an application for interacting with those data in those files. This could be a word processing application like MS-Word, say. It is this combination – data and application - that start to look like the thing that user ‘sees’ when they interact with a computer. For the user, their files are data-as-seen-through-a-word-processing-application; it is this that they interact with it - as a thing they read, write, store, forward, print. These are of course their love letters or their correspondence from the bank; whatever. The long and short of it is that what a file is to the computer is not essentially the same as it is to the user. There is asymmetry here. But one might also say that the user *doesn't really care* what the computer thinks a file is, or where it is stored, or how its stored or *even* the application used to interact with that data; as long as the file and the things they can do with are what the user wants. Getting a match to that wanting is of course crucial, but if it is provided, it doesn't matter to the user how complicated or simple all that work that delivers the file to them is. It is *irrelevant*.

This is enormously consequential but I think it is the consequence of good HCI. It was HCI that made the insides – or these aspects of the insides to be more precise - irrelevant. These insides weren't irrelevant when the design was done; they mattered at that time. But the *output* of the design was to make them so. To be irrelevant, the design of the interaction with the computer had to result in something being afforded; some output of the interaction needed to satisfactorily address a need of the user so that those insides didn't matter or they didn't start to enquire into when they found themselves failing to achieve their ends. There were a number of ways this was achieved; graphical design, hardware, systems integration. But the thing I want to focus on is how this was achieved, in part, through an *abstraction*²⁶. An abstraction is not a way of ‘explaining’

²⁶ See Harper, et al (2013) *What is a File?*

complexity. In this case if files, an abstraction operates at the precise spot where the system and the user interact²⁷. It unifies the different sets of tasks I have just described: the management of files by the system, access to those files by an interactive application and through that, use of those files (and the data) by the user. The abstraction also functions in this two-sided way when a user ‘saves’: the system ‘writes’ that data (i.e., the elements constituted at an abstract level as ‘a file’) and stores it. The user doesn’t see this. They do not see where it is stored, though the abstraction gives an impression of this (on the desktop, say, or on some other location)²⁸.

One could go on. What I am drawing attention to is how interaction with basic computers entails the design of procedures where some complexities are hidden and modes of interaction foregrounded that turn around what users are thought to want to do given what the system in question can do. There might be explanation here, but it is not the explanation that some think is needed for AI: explanations for how computing works that is comprehensive and right. On the contrary, there is a design here, an abstraction in particular that enables the *unification of both the doings of the user and the computer*. This *harmonises* these doings.

These doings are not just enabled by the abstraction of course, but they are also articulated (or embodied if you prefer) in the symbols on a ‘desktop’ interface, in the

²⁷ Thereska & Harper, (2012) *Multi-structured redundancy*.

²⁸ When seen thus, in terms of what abstractions in Turing Theoretic machines do then it becomes clear that much of the social sciences critiques of AI that focus on the distance between abstraction and complexity miss the mark. Papers like Selbst *et al's Fairness and Abstraction* (2018) say more about social studies of science and technology (SST) than they do about computing despite the authors’ claim otherwise.

graphical representations that can be interacted with through a keyboard and mouse; through the whole assembly of components. This basic WIMP mode of interaction was devised by Xerox HCI researchers years ago – nearly forty years ago²⁹. The abstraction I have been focusing on is in regard to files, and in particular text files. There are of course other kinds of abstractions. My point is that with these abstractions, with the WIMP system it is part of, users could do things; it afforded what might call a grammar of action. This term has its roots elsewhere³⁰, not in computer system design, but was used by the Xerox researchers themselves to convey the idea that their particular designs and abstractions embodied particular meanings, and these actions and meanings come together - or afforded – a unit of action. Here it is with bundles of data stored and used in and through a particular application on a computer, via its interface so that users could do word processing. The Xerox researchers realised that any grammar of action needs to privilege certain tasks over other tasks if it is to enable users to get on – to avoid the mangling of fitting applications to data, to finding data, assembling it, writing and saving it; in short pulling all these complexities together. And they chose document creation and layout as the task that users could focus on through their WIMP interface on their Star machines. One might note that this was a big call; it meant that incredibly expensive and sophisticated computers were to be used to create documents. Up until then, documents were created by cheap technologies and lowly paid staff; now, with the new Star systems, Xerox gambled that the new creativity their systems enabled with the

²⁹ See Smith et al, (1982).

³⁰ It is associated with the later Wittgenstein, for example (1953), but probably the most important exploration of this concept was by Kenneth Burke in his *A Grammar of Motives* (1945).

written word would be leveraged by senior staff in organizations, one's whose role could justify the investment in the new machines and in learning the new grammar of action. CEOs would type in text, exploit the layout tools, delight in the wonders of the digitally mediated, organizational word³¹. Xerox chose this grammar, these doings with its revolutionary technology. As it happens, this choice did not make good business for Xerox; other, much less sophisticated desktop computers won the race – Apple and its MAC; Microsoft and its comparatively neanderthal MS-DOS running PCs. But these systems were vastly cheaper and sold because they pointed to what the Xerox Star machines could do. They offered hope and rather more than functionality; they afforded faltering copies of what might be. The future was made by Xerox and exploited by Bill Gates.

Be that as it may, this grammar of action - focused on the WIMP interface - is now a commonplace. New grammars of action have emerged with mobile phones, tablets, and of course recent applications, social media platforms. In regard to these, AI tools and applications are clearly going to enable new possibilities, new grammars. The relations between files shared over social media can be a resource for new ways of engaging with files, with metadata about authorship, viewings, and annotations all becoming part of what files come to mean. Whereas once files were the basic unit for the action between a user and a computer, now, crudely speaking, files with their

³¹ And finding out the role of this word was a task set me by Mark Weiser and William Newman at PARC. This led to the study of the world's first organisation to have a complete network of WIMP machines: the International Monetary Fund, in Washington DC. See Harper, 1998.

associated metadata are the currency of sociability³². There are lots of ways this sociability can be augmented and made startling through AI. For a simple example, the relationship between use of a shared file and user identity can be an indicator of other possibilities – such as the likelihood that two users of one shared file may find interest in some of the other files they individually use. There is a closeness here, a notion of ‘distance’ that can be used to make social connections³³. This is basically the ‘intelligence’ under the hood in Facebook. No-one seems to think that this needs *explaining* to the user. The mechanics of the processes of measuring distance, articulating it on an interface enabled by the grammar of Facebook, in its GUI and in the skills that users have come to learn, are not required when they do a post, or make a like. They don’t need to know. The insides of computing in these respects don’t need explaining. They have been designed out of the interaction.

From complexity to action in the age of AI

Users have to learn new grammars of action, needless to say; pointing and dragging, clicking and saving were all new once, just as liking and posting were, too. Leaving aside the history of these changes – certainly interesting in lots of ways – what we should remind ourselves of is that when new grammars are beginning to emerge, it is not always easy to find out what they are and to design good interaction around them. Doing so requires understanding both the user and the technology. HCI needs to look two ways if you like, as I said at the beginning.

³² See Odom, et al, (2012) *Lost in Translation*.

³³ Another way is of course to identify closeness between files themselves. See Harper et al, (Forthcoming) *Breaching the PC Data Store*.

Today, though, this looking is complex. The technologies at issue are not the desktop and its WIMP interface; it is the technology of AI in all sorts of places – yes, in the desktop, but also on the network, in the cloud; being used by different companies for many different ends, from marketing to transactions monitoring, from logistics to customer care³⁴. So finding out what the grammars of action with AI might be is turning out to be hard because it has so many forms.

But it is not just in the range of technologies that are entailed that is creating difficulties. There are distractions too. AI researchers are often keen to celebrate their technology by saying its inner functioning is somehow miraculous, almost beyond understanding; a black box. Though it might be ultimately *only* a question of probabilistic game play, quite how the game play in question is engineered in code can be opaque, they assert. Terms like deep learning are coined, phrases like Bayesian tipping points used and explanations offered that say that what happens inside the AI machine is too difficult to comprehend. All told, AI fails the test of ‘intuitive understanding’ they and those who believe then assert³⁵. Naturally this can encourage the notion that what AI does is beyond comprehension even for experts in the field, HCI ones, say, wanting to make AI tractable to human creativity. Instead the impression is given that AI is some dark science controlled by cabals in Silicon Valley. If this were true, then I would say that doing good HCI with AI would be hard – if not impossible.

³⁴ A point made by Farooq and Grudin in their *Human Computer Integration* (2016).

³⁵ For explorations of this apparent dilemma – the contrast between the elaborate complexity of AI tools and the desire for everyday understanding or ‘intuitive understanding’ - See Selbts & Barocas, *The Intuitive Appeal of Explainable Machines* (2018).

To say again, to do good design requires the HCI researcher understand both the technology and the user. So how does one do that? Can one look both ways – to the AI and to the human? Before I get to that, let me pause and reflect on the question of understanding AI in the general, and not just for HCI purposes.

Knowledge of AI, society, law

I do not want to suggest that questions of correct, ‘epistemic’ understanding of AI processes are never required beyond the moment of their devising and their referencing in HCI design work. The complexities of AI are not to be dismissed as merely a question of description, of words. They are important issues here, language notwithstanding. Statistics are always opaque, for example, and so that they produce outcomes that are not always expected is absolutely nothing special about AI or its processes. That statistical learning – as machine learning processes are called in the statistics community - is especially complex in its functioning, and thus almost guaranteed to produce outcomes that cannot be easily predicted through ‘back of the envelop reasoning’, is also nothing to be startled by. What is perhaps unique to AI is the kind of data encompassed in its statistics, not that AI *is* statistics. Moreover, the things that many AI tools are used for are often human activities in the aggregate and how these aggregates are used to articulate the single individual is quite rightly an interest to people especially if they are the single person highlighted in some instance³⁶. How insurance companies calculate premiums for one person over some other can be one

³⁶ Some of the more interesting and thoughtful work here can be found in the research of M. Hildebrandt. See for example *Profiling: From Data to Knowledge*, (2006); and *Smart technologies and the ends* (2015).

such case. Historically, these processes have always been at the cutting edge of statistics; but now, the legislative communities have sought ways of ensuring ever-greater accountability. Individuals are now entitled to have the way their activities are measured and calculated explained to them. These rights are embedded in consumer privacy regulations³⁷. This has not made the AI tools used in insurance explainable, though. For one thing, consumers of insurance policies, as a case in point, are only allowed to see the processing of data about themselves, and so cannot see how comparison with others affected the weighting given to their own case. This can make the account they are given almost useless. For another, to understand the statistics in some case requires expertise in statistics; that expertise simply does not equate to what is called the test of ‘intuitive logic’ – i.e. what some supposed ordinary persons would judge as a ‘reasonable logic’. One would not expect an untrained person to understand state of the art astrophysics, but for some reason, in the eyes of legislators, every citizen is expected to be able to comprehend machine learning. In effect, their capacity has become enshrined as a ‘right’³⁸, as if ability and rights were isomorphic. This is not to diminish the importance of accountability, nor indeed, the need to regulate and control the use of analytical tools that can deliver – all too often - social inequalities. But it is to say that the focus on AI as something extraordinarily difficult (and wonderful at the same time) is creating paradoxes and problems. Narratives about AI are creating unnecessary inflation of issues. And as they do so, so they diminish the potential role

³⁷ The canonical case is of course the EUs *General Data Protection Regulation* (GDPR) act that came into effect in 2018.

³⁸ See Goodman & Flaxman, (2016) *European Union regulation on algorithmic decision-making*.

of, for example, good HCI in the domain. It is to that we now return.

AI and types of interaction

One way we can begin to explore what HCI approaches to AI might do is to look at how some AI techniques work and might be used in some instances when brought alongside understanding of what the user might want to achieve with AI. For the purposes of this paper these instances can be real or simply thought experiments, as long as they are convincing – likely, as one might say (though we should remind ourselves that good HCI often requires ‘research in the wild’). Let us take the notions of *backpropagation* and relatedly, *greedy algorithms* – the latter being both the label for a process of optimization and a characterization of the consequences of algorithms. These are terms that conjure the mystique of AI; these terms are often too the magical components that make for black boxes. What do they entail?

To characterize them simply: backpropagation is in essence a step by step data analysis procedure that seeks to uncover patterns in data. The process is sequenced, with later processes revisiting outcomes of earlier ones; that is to say that outputs of later stages are taken back to *propagate* in earlier ones. This circular process is done iteratively. It often entails breaking up the data into subsets and analysing each set independently before bringing all the outcomes together; and iterating in the small again. Small data analysis activities can be undertaken in early stages, so to speak, before the outputs of these are aggregated and used to drive later stages of investigations and the results of these later stages are then returned (as it were) to early stages for a second (or more) run-through of the same data.

Gradually this procedure will begin to optimise on a set of patterns or classes. When this optimization starts to occur, the balance of interpretations (classifications)

may err toward one set over another. As this happens - assuming it does – this then leads to another run through of the data analysis task at the segmented, small data set level, where the ‘tipping point’ occurring in the data analysis process starts earlier (or more rapidly). This can change the resulting classifications in each subset and run through, which then might alter the overall balance of classifications. And so on and so forth.

This process of ‘learning’ depends on a reference set or a catalogue of things that might be found in the data in question. Reference to this is intrinsic to the classifications offered up in each cycle or iteration; various instances might be invoked in the iterative process, so at one stage an object X might seem to be identified, but by the conclusion, object Y is selected. Often times, though, an object that seems to have been identified early on can end up steering the overall process; when this happens the algorithm that found that object is sometimes said to ‘greedily’ search for and affirm proofs that that object is indeed the one. It *overfits* as the saying has it.³⁹

Now I do not want to claim that this is a perfect or accurate account of the processes in question; more like a caricature to be sure. But I hope it is good enough to grasp how a backpropagating system would work in real contexts – and why such a system might produce surprising results in some cases. In the literature, there are very many examples of the kinds of surprises this and related sorts of approaches deliver, and

³⁹ As it happens, greedy algorithms are more often associated with decision tree methods, where it becomes difficult for the process to return or go back to a prior junction in the tree structure of analysis, finding itself stuck in a line of interrogation that it cannot withdraw from. But the term greedy is also metaphorical, and that is how I am using it here.

most of these examples have to do with computer vision systems. Here, objects originally recognised in separated segments of data are redefined when all the segments are added up and a ‘balance’ made in ways that surprises. And the examples are of cases where these selections are wrong, the ‘greedy algorithm process’ choosing the incorrect optimisation. A husky dog identified in the first one through in a central segment of an image set might be redefined as a wolf when all the segments without a dog-like entity are processed and a second run through occurs; when backpropagation is imposed it is judged that snow is seen in these other segments. Wolves live in cold climates after all. And so a wolf it is - the greedy algorithm asserts itself, so to say.⁴⁰ What should be clear is that, in this example, it is not right or wrong that are being processed by the system but simply the additions of probability resulting in a call, a stratagem of interpretation given the data. The stratagem is to say (to assume it is) a wolf.

It seems reasonable to continue treating matters of classification this way; as strategic rather more than epistemic. Rather than think of issues about wrong and right, it is better to think in terms of doings and contexts – as an HCI researcher needs to. When looking at doings in the real world, people may often treat what they see in a similar manner to a greedy algorithm. After all, how easy it is to distinguish a husky from a wolf? Experts might be able to, but if one is skiing *off-piste*, to take one context, one doesn’t want to make a mistake. So one makes a choice, a practical one: one treats the creature as a wolf. Truth has to do with purposes, I am saying, practical purposes, ones to do with safety. More importantly, what I am doing is also justifying the use of

⁴⁰ This is in fact a very oft-used example and is selected because it seems uncontentious. But see Caruana, et al (2015) on the greedy algorithm problem in healthcare situations where the consequences are more worrisome

an ‘opaque’ process (backpropagation and how this might result in greedy algorithm behaviour, a black box scenario if ever there was one, and hence ‘opaque’) by giving a plausible situation where the use of this set of AI techniques *makes sense* – by locating them in a grammar of action where that sense is to be found. The grammar appropriate when skiing is one particular kind – and so wolves it is. In other situations, though, when other doings are at hand, the use of backpropagation, greedy algorithms, overfitting, needs careful thought – it might matter; it really might not be a wolf.

From wolves to text

One can illustrate this with the role of such techniques in assisting the creation of text by a user, when a user wants to create a message in a person-to-person communications scenario – via IM or SMS say; even on a social media platform where such communications tools are commonplace. In this situation, to put it simply, a user might start by typing – ‘entering’ - a word. One might conjecture that AI systems in this setting would be one that uses registers of likelihood to predict what a word would be as it is typed in. The AI would ‘offer up’ these likelihoods to the user at the appropriate time. Thus the word ‘Whatever’ would be predicted as a likely outcome of typing the following letters in sequence by the user ‘w, h, a, t, e.’ The system would then prompt the user with a ‘ver’ added to the end of their text, ‘whate’, and offer ‘whatever’ even as the user is typing. The user could then accept that assistance by allowing that prompted word to slip in as they type.

Without wanting to say too much about how these systems work, though, what one can say is that we are all familiar with the little dances between our fingers and text entry tasks that can result when these sorts of combinations between AI and ourselves occur. There is a grammar of action here, if you like, one in which AI tools and the

aspirations of the user nicely fit, they harmonise. There is an abstraction here too, insofar as the AI offers up an insert, a predicted word and not a picture of all the words in its engine; the user meanwhile agreeing to act on those pictured offerings by accepting or rejecting them at that moment when the object is rendered in the interface – the proposed word. The word is the point of abstraction, so to speak, combining the operations of the system with the actions of the user. Good interaction design would clarify how this process unfolds, without cluttering the screen say, offering alternate words with the optimal lingering time, and so on. Whether there is always good design in this space is another matter.

But moving on. These individual word solutions could also become part of a more complex AI set of solutions. It could become the first stage of a backpropagation process, for example, where the words become instances of a larger pattern – a syntax, a sentence. Here, the first set of words offered would be subject to second stage (or even more) of analysis, the output of which could create a different set of problems for the user, if not the system. In particular, when an iteration of backpropagation occurs and, as a phrase begins to appear, as the syntax emerges if you like, so the backpropagation process might seek to redefine words that have *already been shaped*. That is to say, the system could reinterpret what might be individual words and hence their spelling through reference to an emerging syntax.

So, for example, as the following phrase might appear (is typed in by the user), ‘Whatever I sa’, so the system might propose that what is about to be typed is *most likely to be* ‘What I say’. It might suggest this because ‘what I say’ offers what the system thinks is a closer fit to its pattern models than ‘whatever I say’. Accordingly, the system might alter the spelling of ‘whatever’ to ‘what’. Individual words would thus be redefined through backpropagation; meaning would be made backwards, so to speak.

The words, meanwhile, would have been predicted beforehand, ‘forwardly propagated,’ if you like. The important point is that the identification of meaning would be a function of the aggregation of separated calculations, first this, then that.

I am not claiming this a precise or accurate description of how text-based messaging systems do in fact offer support around syntax structures through AI. Some might, others might not. But bear with me – this is a thought experiment. What should be clear is that this overall process is fairly straight-forward – the workings of it. This doesn’t mean that the experience afforded is straight-forward to the user. Indeed, if this is what can happen, then one can imagine how this could be the source of enormous vexation to the user. One can imagine, for example, how this might create a greedy algorithm-type behaviour where a solution offered to the user fits the optimisation model but is not the *right* solution from the user’s point of view. A user creates words, shapes phrases but then a system can come to respell individual words – after they have been completed! - through reference to what it thinks is the likely clause being written – the syntax. And the result? Just take the imagined case above: instead of a user writing ‘Whatever I say’, the text that the (hypothetical) messaging app comes to produce is ‘What I say’. A husky becomes a wolf.

As I remark, I am not claiming that is these sorts of AI tools that are used in text communications platforms, but what one can note is that AI tools *are* embedded in most text messaging platforms and whatever they are, they do end up producing these sorts of muddles. They do not provide opportunities for the user to decide about meaning but instead *decide meaning for the user*. It is no wonder that they create vexations. I am not making this up. Terms like *textese* are now commonplace and label the ill-begotten product of millions of users attempting to express themselves *despite the muddling*

assistance of AI tools. ⁴¹ This is not to criticize backpropagation, greedy algorithms or AI in general, it is to criticise the design of systems where the grammar of action – the sought for, ideal grammar of action - has not been thought-through and hence the right role of AI has not been thought-through either. What we see in these imagined examples is that the tools can work wonderfully in the production of spelling but not so well in the production of meaning. The former task, in that ‘unit of action’, spelling words is well supported, but in the larger ‘units of action’ of which these are a part, the task of making sentences, it is not. Just why AI tools are so badly deployed in messaging contexts I cannot answer, except to say it would appear that the impressive powers of AI tools have dazzled developers and venders of messaging apps into thinking that their deployment will succeed *willy-nilly*. There is certainly no evidence of HCI research being used and if it has, it should not be lauded. Daily, users are confronted with proof that AI needs HCI – or least good HCI.

Messaging, meaning, purposes

I have chosen a person-to-person messaging examples since I think one needs to distinguish the kind of grammars of action for the behaviour that is sought for - the game being played to go back to an earlier simile - and other situations. I am not denying that AI tools can provide a resource *tout court*. I am saying that, in some cases,

⁴¹ I am not referring to the new vocabulary created by users, one that is at once playful and refined, despite its apparent abuse of good grammar. See Harper, *et al* (2005). I am thinking of the anacoluthia and plain loss of sense that users struggle with when interacting with their SMS tools, struggles they laugh about and mock. Saying wolf when you meant husky is but the least of their troubles. The analogies with AI translation tools are obvious, but there at least the AI provides plausible meaning. With messaging, meaning is often lost altogether.

what is required are not explanation of the complex ways that, for instance, the AI techniques deployed work, but rather abstractions that cohere both the user and the application in ways that lets them work hand in hand; that lets them get on with what they want to do. In these examples, the task has been *making meaning with words*⁴². But what about other tasks? If ‘meaning making’ in person to person communication is one kind of grammar, then a very different kind of activity is *seeing meaningful objects* in a visual field. To guess something is a wolf can be a reasonable thing to do in some situations we have suggested; and how a greedy algorithm-type process supports this might be assisting of the user and their needs. I am saying that, in the case of messaging, backpropagation and greedy algorithms may not be helpful, pushing the role of assistance one step too far; but elsewhere, in other situations, it might have a role. To determine this, though, requires both understanding of what a user is about and what a technology can afford when deployed by the user given those doings. Part of the background to my arguments about the need to understand what AI tools do alongside what a user wants to do is that I want to contend with the notion that AI tools are so opaque that they cannot be fathomed. But I think that they need to be fathomed to make good HCI. I think the above sketch of how some AI tools works shows how one can leverage such understanding, though I am not claiming my account is accurate. It is the need to understand that is my concern. With this understanding one can see misfits, but one can also begin to see new possibilities.

Grammars about reasons

In light of these examples one might say that a key direction of current HCI

⁴²I alluded to some of these issues many years ago. See *Texture* (2010).

research in the AI space is uncovering and clarifying what AI-enabled reasons can allow; what new user intentions can be achieved with the reasons AI affords? A reason can be many things, needless to say, embedded in particular acts, in what I am calling grammars of action. With AI one can learn spelling, for example, and thus a language (to continue the example from above) but to use language ‘on the fly’ to express oneself is another grammar altogether, and it might not be helpful for AI to offer ‘alternative’ spellings or syntax. The former kind of reason is useful, the latter less so. As I have intimated, in person-to-person communication, it is oneself that needs to express; the reasons for our choice in that expression should be our own, not a proxy’s.

The context of search offers other kinds of reason too. Here users have come to account for these reasons, reflecting, perhaps, a subtle awareness of the grammars these are part of. The PageRank algorithm doesn’t offer understanding of content of the web⁴³. It is a ‘frequency model’ that, as it were, offers ‘reasons’ that justify the use of a web page. PageRank, put simply, counts the traffic between sites - between pages – and gradually comes up with a method of weighting significance given that traffic; the variation in this weighting is then used to triage possible targets to offer to a user on the SERP – the search enquiry results page. The most weight, crudely speaking, or rather the most ‘weighty’ connections, are used to select pages out of the search engine index in some kind of order, a prioritised list. This is what ‘googling’ means as a *reason*. This is not an intelligent reason, in that some understanding of the content is implied. I

⁴³ It is worth noting that when the algorithm was patented it was not labelled AI; it was simply described as a technique. The current fashion for AI has meant that today it is often renamed as AI; the parent company of Google, Alphabet, is rather fond of saying all it does is ‘AI’. For them, AI is ABC, so to speak.

think that is what most users would imagine the word intelligence means here. Nevertheless, PageRank (as a verb) is reason enough to act. And, indeed, people recognise this in the way they talk about using search: they say they '*googled something*'. They do not say that they have undertaken a vast, comprehensive search, 'as a scientist might'. To google is a short-hand for what google itself does, namely short-hand techniques that are mostly good enough for what people are doing on the web – looking for something, buying something, searching for holidays to choose. They are not doing science.

So what happens when one is doing science? Google itself offers a scholarly version of its search which is essentially nothing more than a different index: the google crawlers have been through academic content on line, and not on web pages (though often these are interchangeable). It searches different stuff, if you like, but uses the same techniques, including PageRank, to triage and select targets to offer the user. One would imagine that even if they are seeking scientific reasons, scientists themselves would still find this procedure useful, offering 'reason enough' to get on. This is in large part because of the sheer scale of the 'scientific record' – the papers that scientists produce, to put it simply. Take the field of genomics: in 1995, there were 5 papers on 'genome wide association studies', 141 in 2005, and an incredible 3,633 papers in 2015.

Google can help here but does it help enough? Do scientists want more reasons, better ones, AI enabled ones, say? I think they do, but more work is required. I have remarked on how some of the latest AI tools function; backpropagation, for example. There are many other tools that one might want to apply to the scientific record. One might want to search through text to identify similarities between papers on the basis of phrases used or names cited; even the tone of a paper if one uses sentiment analyses. There are many techniques and procedures, even if they all start with the same material:

‘text’. Text is not the only thing a scientist might want to examine, of course; they might want to examine images, visual records of one thing or another. Here too there are numerous models, techniques and procedures; here too the jargon for all these is a kind of argot, distinguishing those on the inside of the tribe (the AI tribe) from those on the outside, who don’t understand. As should be clear by now, I think one should disregard the powers of this argot and its divisiveness.

Nevertheless, there are those who are seeing beyond these traps and seeking ways of offering up AI tools in the text and visual data fields in ways that points towards grammars of action. The researchers I have in mind look at the space here from both the systems and the user’s point of view. And, as they look in this way, so they do not think that the complexity of AI tools needs explaining except insofar as this pertains to what the user needs. What they find is that the user looks for trust in relation to the AI systems.

The trust here is of at least two basic types. The first is to do with whether one can trust the tools to do a good job at what they are supposed to do. That is to say, given a training set, does the application deliver the right matches to that set? This is, if you like, the engineers’ problem of wanting to make sure that the technology they are building does what they want it to. Ribeiro *et al* (2016) argue that measures of this can be afforded by making visible instances of the examples used in the data trawl (what can be called the ‘local set’) against examples from the training set. The engineers can make an evaluation by a shown comparison. Of course, this begs the question of which examples to compare. The interesting insight that Ribeiro *et al* provide is to use probabilistic techniques to select these; to use AI to allow humans to judge AI. In addition to this, they select the level of detail that is offered in the examples - the point of optimisation if you like – as a function of the time taken to examine them (by an

engineer). Their claim is that just as there is an optimum between correct outcome and speed of outcome inside AI systems (with one possible problem being greedy algorithm behaviours when this is not correctly set), so there is a similar need to balance speed with accuracy when user's test machines. Ribeiro *et al* determine this in a fairly simple way, and evaluate the outcomes of their model in laboratory settings, asking what the users feel about the balance – good or bad, too shallow or too deep, too long or too quick. In my view, Ribeiro *et al* are asking essentially the right questions, but I think the answers they seek would be better if based on real contexts of decision-making. Their research uses hypothesised scenarios, and these are experimented on in a university computer science dept. The ecological validity of all this seems fairly weak. I would go stronger than that and say that experiments are especially unsuited to answering the particular questions Ribeiro *et al* ask; fieldwork is required. In this case, this would be in the engineering sites where visual and text classifiers are being productised. But I don't want to rebuke Ribeiro and his colleagues; they are pointing the way.

The second type of trust that Ribeiro *et al* define is when people want to use the system to make decisions, to act with. This is not a question of whether a system does what it is supposed to, but whether the user can trust in the kind of data the system uses as its reference point given the different, new data used in the task they have at hand. Here questions have to do with whether the learning sets are relevant to the field that the user is inquiring in to. If a system is offering matches for quadrupeds, are the animals in question cows or dogs? And if dogs, does the system distinguish between types of dogs? Could it tell the difference between a husky and a wolf? And if it can, does it do so on the basis of the features of the creature or the setting in which the creature is 'seen'? Again, the solution Ribeiro *et al* suggest is offering representative examples of the learning sets against the examples of the data being examined; again they call the

latter 'local'.

In both cases of designing for (to give) trust Ribeiro *et al* offer what they call abstractions -; abstractions that show themselves in a visual display on a desktop but articulating underneath a quite complex set of operations; but all told, making a single point between what the technology is doing and what the user is doing. Their achievement is, thus, very analogous to what the Xerox researchers invented all those years ago. As it happens, these abstractions, in LIME on the one hand, and the other, in SP-LIME (the oddity of these names need not detain us here) are far from perfect, as Ribeiro *et al* admit. More explorations of their design and visual renderings could be undertaken and these might afford more finesse in both the rendering and in the evocativeness of the iconography chosen in the visualisations. More testing and at greater, more 'naturalistic' enquiries into the context could be undertaken too, of course. I do not want to explore what these suggestions might entail in any detail⁴⁴ but do want to note that *Why should I trust you?* is rare in the current literature; a paper by researchers who don't seem intoxicated by their own technologies and are instead interested in thinking about what those technologies could be put to. They inquire seriously in to this, and focus on two well-known areas: text and image classifiers. They combine their knowledge of tools for these with knowledge about what users might do

⁴⁴ The question of how one might enquire into the real world, into natural action for want of a phrase, has been a major concern in HCI since the turn to the social, with the emergence of CSCW and similar (See Randall *et al*, 2007). It is certainly something I have spent much time on, a key concern in my research being to distinguish such research in the wild for the purposes of HCI and doing so for social scientific reasons, for anthropology or sociology. These purposes are not the same and should not be muddled. See Harper *et al*, (2005).

with these technologies. The intersection of this produces the interaction design. This is HCI in my view.

Conclusion

Classifiers are to be found in all sorts of places, classifying all sorts of different phenomena: from classes in operations management – think of Amazon’s basic task; through to pricing models given market velocities – think of airline ticketing; through to media consumption and the edges on graphs that suggest advertising opportunities – think of social media platforms. The role of classifiers is also part of amalgams of different tools and systems. AI never works alone, even if it seems the most conspicuous component of the computer systems people use. Irrespective of that, the places in which AI can have role are immense and diverse. My thesis has been that to understand what these roles might be, to understand what the game in question is (to return to my analogy), requires, I think, not just understanding of AI but understanding of users too. It requires aswell methods for seeing how they might harmonise (or be harmonised). Ribeiro, Singh and Guestrin offer some ways in which this can be achieved. But from my reading of the literature – my limited reading of an immense literature so my understanding might be mistaken – their paper is the exception rather than the rule. It proves that AI needs more HCI than one currently finds. The limitations of LIME and SP-LIME point the way to the kinds of research that needs to be done, the kind that helps create the marriage of purposes between user and machine that HCI excels in; but as I look around, I do not see enough.

I have explored some of the reasons for this – the mystique surrounding AI for example. It might also be because of other reasons – HCI researchers have gotten very interested in the social contexts in which systems find themselves, as a case in point,

and this has perhaps led them away from ensuring their focus combines insights in to how systems work with insights into user activities. HCI might have become too concerned with use in new places at the very time when a revolutionary technology is altering the basis of the computing. The ‘turn to the social’ is emblematic of this.

Now while I acknowledge the benefits of this turn, one might pause and look at the current crop of textbooks in HCI and see that there is very little indeed about how systems work in these books. The emphasis is on contexts, not technology. But this focus is, it seems to me, unanchored without understanding of how technology works. I have argued that some knowledge of computing is required to enable affective exploration of what new things users can do with technology, and given the possibilities that AI affords, it is with AI that HCI researchers need to get to grips. So these textbooks need more on the insides of computers, especially AI systems and tools.

The purpose in this paper has been to offer some illustrations of the thinking that can result from this combined view and in particular what can be done with HCI through having some grasp of how procedures within AI systems work. Whether I have done a good job at that, or whether more care is needed in this getting to grips with AI and being creative with HCI are topics for future papers to ask. Doubtless they will improve upon my own.

Looking at the present, though, what I am saying is that HCI might have begun to take the insides of computers *too much for granted*. I am not offering an analysis of why that might be the case. I am just offering some possibilities for ways forward and away from that situation. What I do know is that the future is not AI; it can only be an AI enabled through HCI. But I also know that HCI researchers need to stand up and take on that labour. They need to engage with their own grammar of action if you like,

they need to understand what AI can do *and* see what it can let users do through some kind of collaboration, a joint working that can allow new things to be achieved. Only then can they make the future. But I do think it is there to be made. And if they are willing, it will thus be HCI researchers who will help make that future a creative one, a future where the artificial augments what people want to do rather than substitutes it. The future is HCI, not AI.

References

1. Anscombe, G.E.M. (1957) *Intention*, Harvard University Press, Cambridge.
2. Boden, M.A. (1977) *Artificial Intelligence and Natural Man* (New York, NY: Basic Books).
3. _____ (2016) *AI: Its Nature and Future* (Oxford: Oxford University Press), 2016.
4. Burke, K. (1945) *A Grammar of Motives*, University of California Press, Berkeley (reprinted 1969).
5. Carr, N. (2015) *The Glass Cage: who needs humans anyway?* Vintage Books, London.
6. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M. & Elhadad, N. (2015) *Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day re-admission*, Proc. 12th ACM International conference on Knowledge Discovery and Data Mining, pp 1721-1730.
7. Dennett, D. (1991) *Consciousness Explained*, The Penguin Press, London.
8. Farooq, U. & Grudin, J. (2016) *Human Computer Integrations: Implications for design as we enter a new phase in our relationship with technology*, ACM Interactions, 23,6,26-32.
9. Grudin, J. (2009) *AI and HCI: Two fields divided by a common focus*, AI Magazine 30, no. 4: 48-48.

10. _____ (2017) *From Tool to Partner: The Evolution of Human-Computer Interaction*, Morgan Claypool, New York.
11. _____ (2018) *Bridging HCI communities*, Interactions 25, no. 5: 50-53.
12. Farooq, U. & Grudin, J. (2016) *Human-Computer Integration*, ACM Interactions, 23, 6, 26-32.
13. Goodman, B. & Flaxman, F. (2017) *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"* ICML Workshop on Human Interpretability in Machine Learning, arXiv:1606.08813 (v3); 38.
14. Guidotti, R., Moreale, A., Ruggieri, S., Turnini, F., Pedreschi, D. & Giannotti, F. (2019) *A Survey of Black Box Methods*, ACM Computing Surveys, Vol 51, Issue 5 January, Article 93.
15. Harper, R.H.R., (1998) *Inside the IMF: An Ethnography of Documents, Technology and Organisational Action*, Academic Press, London & San Diego.
16. _____ (2010) *Texture: Human Expression in the Age of Communications Overload*, MIT Press, London and Boston.
17. Harper, R. & Mentis, H. (2013) *The Mocking Gaze: 'You are a poor controller!'*, CSCW2013, San Antonio, Texas, ACM Press.
18. Harper, R., Palen, L. & Taylor, A. (2005) (Eds) *The Inside Text: Social perspectives on SMS*, Kluwer, Dordrecht, Netherlands.
19. Harper, R., Randall, D. & Rouncefield, M. (2005) *Fieldwork and Ethnography in Design - The state of play from the CSCW Perspective*, Proceedings of EPIC, American Anthropology Association, Redmond, USA, pp81-100.
20. Harper, R. Thereska, E. Lindley, S. Banks, R., Gosset, P., Odom, W., Smyth, G., (2013) *What is a file?* CSCW13; also published as Microsoft Technical Report MSR-TR-2011-109.

21. Harper, R. Banks, R. Gosset, P. Lindley, S. Thereska, E. & Smyth, G.
(Forthcoming) *Breaching the PC data Store: What do Graphs tell us about files?* in
Chamberlain, A. (Ed) *Research in the Wild*, Springer, London.
22. Harper, R., Randall, D., & Sharrock, W. (2016), *Choice*, Polity Press, Cambridge.
23. Heehs, P. (2013) *Writing the Self: Diaries, Memoirs, and the History of the Self*,
Bloomsbury, London.
24. Hildebrant, M, (2006) *Profiling: From Data to Knowledge*, *Datenschutz und
Datensicherheit*, 30, 9.
25. _____ (2015) *Smart technologies and the ends of law*: Edward Elgar: Cheltenham.
26. _____ (2006) *Smart technologies and their ends* (2015).
27. Husain, A. (2017) *The Sentient Machine: the coming age of artificial intelligence*, Souvenir
Press, London.
28. Kaplan, J. (2015) *Humans Need Not Apply: a guide to health and work in the age of artificial
intelligence*, Yale University Press, New Haven.
29. _____ (2016) *Artificial Intelligence: what everyone needs to know*, Oxford University
Press, New York.
30. Jones, R. (2004) *Soft Machines: nanotechnology and life*, Oxford University Press,
Oxford.
31. Kitano, H. (2016) *Artificial Intelligence to Win the Nobel Prize*, *AI Magazine*, Spring,
pp39-49.
32. Kurzweil, R. (2005) *The Singularity is Near*, Penguin, London.
33. Ma, X. (2018) *Towards Human-Engaged AI*, In *IJCAI*, pp. 5682-5686.
34. Markoff, J., (2015) *Machines of Loving Grace*, Harper Collins, New York.
35. Miller, (2018) *Explanation in Artificial Intelligence: Insights from the Social Sciences*, under
review.
36. Neyland, D. (2019) *The Everyday life of an Algorithm*, Palgrave, London.

37. O'Hara, K., Harper, R., Mentis, H., Sellen, A. & Taylor, A. (2013) *On the naturalness of touch: putting the "interaction" back into NUI*, ToCHI, ACM Transactions on Human Computer Interactions Special issue, 'Embodied Interactions'. Volume 20 Issue 1, March, Article No. 5.
38. Odom, W., Harper R., Sellen, A. & Thereska, E. (2012) *Lost in translation: understanding the possession of digital things in the cloud*, In *Proc. CHI 2012*, ACM Press, 781-790.
39. Randall, D., Harper, R., & Rouncefield, M., (2007) *Fieldwork for Design: Theory and Practice*, Kluwer, Amsterdam.
40. Ren, X. (2016) *Rethinking the Relationship between Humans and Computers*, IEEE Computer 49, no.8: 104-108.
41. Ribeiro, M.T., Singh, S., & Guestrin, C. (2016) *Why Should I trust You? Explaining the Predictions of any Classifier*, Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA USA.
42. Russell, S & Norvig, P. (2017) *Artificial Intelligence: a Modern Approach*, Pearson: Boston.
43. Sadler, M & Regan, N. (2019) *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*, New Chess, The Netherlands.
44. Selbst, A. & Barocas, S. (2018) *The Intuitive Appeal of Explainable Machines*, Fordham Law Review 1085.
45. Selbst, A, Boyd, D., Friedler, S., Venkatasubramanian, S. & Vertesi, J. (2019) *Fairness and Abstraction in Socio-Technical Systems*, Proceedings of the Conference on Fairness, Accountability and Transparency, pp 59-68.
46. Shanker, S. (1998), *Wittgenstein's Remarks on the Foundations of AI*, London: Taylor & Francis.

47. Smith, D. C. (1982) *The Star Interface: An overview*, in Brown, R. & Morgan, H. (Eds) AFIPS'82, pp 515-52.
48. Stanford University (2016) *Artificial Intelligence and Life in 2030*, Stanford University, <https://ai100.Stanford.edu>
49. Stanislaw, U. (1958) *Tribute to John Van Neumann*, Bulletin of the American Mathematical Society: 64, #3, part 2: 5.
50. Thereska, E. & Harper, R., (2012) *Multi-structured redundancy*, in HotStorage'12: the 4th Workshop on Hot Topics in Storage and File Systems, Usenix, Boston 2012, June.
51. Wittgenstein, L. (1953/2009), *Philosophical Investigations*, 4th edition, tr. G.E.M. Anscombe, P.M.S. Hacker, & J. Schulte, Wiley-Blackwell, Oxford.