Human Centered Artificial Intelligence: Weaving UX into Algorithmic Decision Making

Raymond R. Bond
Ulster University
School of Computing
rb.bond@ulster.ac.uk

Maurice Mulvenna
Ulster University
School of Computing
md.mulvenna@ulster.ac.uk

Hui Wang
Ulster University
School of Computing
h.wang@ulster.ac.uk

Dewar Finlay
Ulster University
School of Engineering
d.finlay@ulster.ac.uk

Alexander Wong
University of Waterloo
Systems Design Engineering
a28wong@uwaterloo.ca

Ansgar Koene
University of Nottingham
Horizon DER institute
ansgar.koene@nottingham.ac.uk

Rob Brisk
Ulster University
School of Computing
brisk-r@ulster.ac.uk

Jennifer Boger
University of Waterloo
Systems Design Engineering
jboger@uwaterloo.ca

Tameem Adel
University of Cambridge
Department of Engineering
tah47@cam.ac.uk

ABSTRACT
Artificial intelligence (AI) is changing many industries as intelligent systems assist humans with algorithmic decision making, machine learning, computer vision, natural language processing, chatbots and robotics. Deep learning has been a key development that has helped propel the widespread interest in AI. Given AI is becoming more prevalent and maybe regarded as the new 'electricity', it is imperative that human needs, interests and values are centered around all future AI developments. Given Human Computer Interaction (HCI) researchers created user centered design principles and practices based on human factors psychology, they are best placed to work with AI engineers to establish Human-Centered AI (HAI) or Human-AI interaction solutions. This paper presents the challenges that would provide HAI solutions, including the need for explainable AI (XAI) in algorithmic decision making and the need to mitigate automation bias which is when humans over trust and accept computer-based advice. Other ethical HAI challenges are outlined including algorithmic bias and ethical chatbots as well as a discussion on the likely democratisation of AI with the aid of end-user machine learning. The challenges that are discussed herein need addressed in order to properly calibrate human-machine trust and to provide a responsible basis for the future of human-machine symbiosis.

Author Keywords
Human centered artificial intelligence, artificial intelligence, explainable AI, automation bias, algorithmic bias, chatbots, democratization of AI, ethical AI

ACM Classification Keywords
Human-centered computing

General Terms
Human Factors

INTRODUCTION
Artificial intelligence (AI) is likely the next 'electricity' [1] and is affording the fourth industrial revolution [2]. With the promise and even prevalence of AI in healthcare, education and personal coaching, the time has come to put humans at the center of AI engineering lifecycles. This has been referred to as Human Centered AI or HAI for short [3]. Given Human-Computer Interaction (HCI) and User Experience (UX) specialists are familiar with user-centered design practices, they are needed to enhance the human-centeredness of AI solutions. Putting humans in the spotlight ensures that AI systems will be ethical, adopted, usable and used, and will help avoid harmful unintended consequences of AI systems. This paper is speculative and positional but draws on published research when necessary. The paper provides a brief overview of AI and then discusses the HAI initiative. The paper discusses some key challenges HAI researchers face, including AI ethics, explainable AI (XAI – providing automated rationale for for AI algorithms and algorithmic decisions), providing fairness in AI, avoiding automation bias (when humans naively accept computer-based advice), visualizing AI uncertainty, mitigating and detecting algorithmic bias (e.g. when machine learning can be racist or exhibit gender bias), ensuring the responsible democratization of AI, the idea of ethically aligned chatbots and a brief discussion on how AI can be used to enhance the UX which is different to looking at how UX can enhance AI.

ARTIFICIAL INTELLIGENCE
Intelligence can be difficult to define. According to the Cambridge dictionary, intelligence is "the ability to learn, understand, and make judgments or have opinions that are

based on reason" [4]. Humans seem to define intelligence in alignment with their own intelligence, which is arguably biased and may not be agreeable to an extraterrestrial higher intelligence. For example, if spatial intelligence were a priority then an impartial judge might rank Google Maps and bees superior to humans. The same dictionary defines AI as "the study of how to produce machines that have some of the qualities that the human mind has…" [5]. The term 'artificial' is unfortunate and might portray that the intelligence is not real, which is an argument formulated by Searle's Chinese room argument [6]. Some researchers use alternative similar terms such as augmented intelligence, computational intelligence and intelligent systems but the term AI has prevailed. Interestingly, AI and human intelligence are arguably complementary hence it may not always be appropriate to seek to replicate all parts of human intelligence in a machine. For example, humans have excellent visual systems for rapid effortless object recognition and complex natural language understanding which are difficult challenges in AI. Humans can also generalize knowledge and form heuristics from one or two experiences (single shot learning) whereas AI often requires large datasets to induct knowledge (in the case of machine learning). However, humans have a poor working memory as their attentional capacity can only hold 7±2 items [7] whereas the working memory (RAM) of machines is significant. Humans are also unimpressive when it comes to performing arithmetic whereas performing millions of calculations is an obvious strength of AI and computing. Kahneman [8] refers to two kinds of thinking in the dual process theory of cognition: system 1 (thinking fast) and system 2 (thinking slow). System 1 is associated with the limbic system and denotes how humans effortlessly and rapidly classify objects and sounds. This is also known as intuition, which is an automatic mind that we use for unconscious reasoning. System 2 is what we use for conscious reasoning and arithmetic which takes effort. Whilst system 1 in human cognition arguably uses a large number of information bits (e.g. visual processing), it is effortless, whereas whilst system 2 may only involve a small number of bits (e.g. arithmetic), it seems challenging and requires effort. Conversely, system 2 thinking is not challenging for machines given that they can reason with millions of rules and perform calculations easily, however machines traditionally find system 1 thinking more challenging (at least before deep learning). In this sense, human and machine intelligence are complementary and provide a basis for human machine symbiosis and human-machine integration or teaming [9]. Arguably, humans are already becoming 'cyborg like' given smart phones travel with us and augment our intelligence as we navigate our way through work and recreation.

AI is a discipline covering topics such as knowledge engineering, robotics, machine learning, natural language understanding and computer vision. However, a key focus in AI research has been machine learning even though the press often conflates AI with robotics, but thankfully this is changing. Most AI applications today are known as narrow AIs given algorithms are usually trained for one specific 'narrow' task or purpose and have no autonomy or artificial general intelligence. Deep learning has become synonymous with AI, however deep learning is not only a sub-topic of AI, it is a sub-topic of machine learning. Deep learning is not new, it is a well established neuro-inspired technique that uses many layers of artificial neurons, and its recent fame and use is due to the availability of big data and CPUs to train a multi-layer algorithm to optimize the weights in neurons [10].

Historically AI-based decision support was dominated by knowledge engineering, expert systems and symbolic reasoning given machines are gifted in reasoning using a large set of facts and logical IF-THEN-ELSE statements. However, developing rules for rule-based systems required excessive computer programming and conversations with end-users and human experts to develop the rules. In a sense, this form of AI was human centered as the computer program was co-engineered with the end user. However, often the human centeredness disappeared when it came to develop the user interface of these expert systems [11]. Nevertheless, AI engineers found this approach to be expensive, time-consuming and cumbersome, which led to the widespread use of machine learning given this technique learns using data alone and that these algorithms can be built in minutes as opposed to months. With machine learning, the computer does most of the work whereas the engineer did most of the work in knowledge engineering. Most machine learning algorithms are perhaps built without involving the end user and without thinking about how it would be used (intended but specifically unintended) by the end user. Interestingly, there is an argument that a unified theory of AI may involve resurrecting knowledge engineering and combining it with machine learning [12]. In some way this might align well with the dual process theory of human cognition where system 1 thinking involves fast pattern recognition using ML and system 2 thinking involves slower reasoning using knowledge engineering. This is an idea worth supporting given it is inspired by human cognition theory as opposed to human neurology or biology which is modern day focus for evolutionary and connectionist computing researchers. Nevertheless, since AI has become prevalent and accepted in many industries, it has become a necessity to involve the human at the center of AI development which has led to HAI initiatives (or human-centered AI). And as a result, many leading universities including Stanford (https://hai.stanford.edu) and MIT (https://hcai.mit.edu) have launched HAI research institutes.

HUMAN CENTERED AI

HAI is an attempt to bridge HCI research and AI research for the common good of humans. Specifically, it is to position the human at the center of the AI development lifecycle. According to a recent HAI paper [3], AI should not replace humans but should instead augment human capabilities. This is a rerouting for AI since a lot of AI systems are projected to replace human jobs as opposed to assisting them (the fourth industrial revolution). This statement of not replacing humans is a key attempt to put humans at the center but is certainly not the only objective of HAI. The same article suggests that HAI solutions should be ethical, explainable, comprehensible, useful and usable [3]. They also provide a working HAI framework which includes three components: 1) ethically aligned design, 2) human factors design and 3) technology enhancement [3].

More than ever, HCI and UX researchers need to address challenges related to human-AI interactions. Most HCI and UX methods as well as usability engineering methods and heuristics were never developed with AI systems in mind. However, a recent article presented at CHI 2019 provided a set of Human-AI interaction design guidelines [13]. Given humans are increasingly engaging with AI systems and making decisions based on algorithms, it is imperative that UX researchers are involved in order to include end-user values throughout the AI development lifecycle. This idea of human-centeredness is not new to UX research, given UX researchers pioneered user-centered design practices, living labs and co-creation methodologies [14]. Integrating HCI with AI development, ensures a multidisciplinary approach since HCI already involves the sub-disciplines of design and human factors psychology.

A HAI example

As an example, let's suppose we take a HAI approach in developing an AI based clinical decision support system. This might involve the following process. Firstly, HCI ethnographic methods such as contextual enquiry can be used to assess current decision-making processes in a human only system. Of course this could be supported by interviews and focus groups. Subsequently, co-creation workshops could be used to co-prototype an AI-based decision support system [15] which involves co-designing the user interfaces and information design of the algorithmic output. This would allow end users to design the visual hierarchy and nomenclature of the AI-based information making it understandable and meaningful. Multiple prototypes should be co-designed allowing for A/B or multivariate testing in a simulation. Simulation in user testing of AI prototypes is one method that is not new to AI with reference to chatbot design, i.e. wizard of Oz testing (i.e. where the user believes they are conversing with an AI but they are deceptively conversing with a human who is remotely controlling the user interface). A usability test aided by a high-fidelity simulation can involve the use of drama and artefacts [16] to emulate real world decision making scenarios [17] aided by the different AI prototypes that were previously co-designed. Using counterbalancing, a HAI team can compare the prototypes in terms of their usability, UX, trust, effectiveness and acceptance. They could also use eye tracking in the simulation to objectively assess the visual attention of the human-AI interaction and the information hierarchy. As a result, a chosen prototype can be refined and only then should AI engineering begin. After and during the AI engineering phase, the end user can iteratively retest the developed system. Moreover, other HAI methods need to be considered at various stages. This includes assessing the data provenance in any machine learning components since data can entail biases and have unintended harmful consequences such as racial discrimination (algorithmic bias is discussed later in the paper). The system would also need to be audited by a HAI team for its explainability, comprehensibility, traceability and transparency, which are concepts discussed later in this paper.

ETHICAL AI

A key aspect of a HAI solution is ethically aligned design which is being largely pioneered by the IEEE [18]. Ethics (aka moral philosophy) deals with how agents should morally act and discern between right and wrong. Ethics is a difficult subject that has been debated by many philosophers from Aristotle to Kant. Understanding human ethics and altruism without considering AI or machine ethics is challenging alone. There are many frameworks and theories in ethics research. Two common frameworks include consequentialism (e.g. utilitarianism) and deontology. The former is an ethical framework where any action is good if the outcome is the greatest good (aka 'the means justify the end'), and the latter is in contrast where the action must be good and virtuous regardless of the outcome. Today, utilitarian ethics is likely the leading school of thought given the emergence of trolley problems that are being considered in the design of autonomous vehicles [19]. A typical trolley scenario is whether a self-driving car should risk driving into a child who runs onto the road or swerve and likely encounter elderly pedestrians. Researchers at MIT [19] have attempted to crowdsource the ethics of autonomous vehicles by allowing a crowd of humans to decide what a self-driving car should do in dilemmic scenarios. Crowdsourcing ethics is one approach and aligns well with democracy and supports existing evidence that there is wisdom in crowds - even in machine learning a crowd of algorithms (known as an ensemble) will commonly outperform a single agent.

Nevertheless, a myriad of principles for AI ethics and computer ethics have been developed over the years by many organizations. Moreover, best practices and even standards by the IEEE are being proposed [18]. A key challenge will be integrating AI ethics into AI development lifecycles, nonetheless ethical guidelines are essential to guide HAI solutions. A key observation is that decision aids and ethical AI audit tools will need to be developed for ethical principles to be effectively adopted and operationalized. AI engineers are not likely to bear in mind a long list of principles as they exist today in large PDFs, but given people remember stories - each principle should be augmented with digital stories and cases studies [20]. Perhaps ethical AI principles can even be gamified to increase their adoption and dissemination or perhaps data scientists could be regularly self-assessed using a system similar to hazard perception testing in UK driving tests

Ethical design is not a new challenge for UX researchers and as such, they can bring their knowledge to the HAI development lifestyle [21]. UX researchers have been exposed to the ethics of persuasive design (which can be used for a lot of good) but can give rise to dark patterns and evil design scenarios where end-users are manipulated into making decisions as persuaded by the user interface design and UX. For example, user interfaces can exploit known human cognitive biases such as the scarcity effect and loss aversion to encourage users to impulsively spend their money online. It is also well known that the HCI of a system can be designed in such a way as to encourage and promote technology addiction [22]. Nevertheless, there are many new conundrums for AI, e.g. should systems reflect the real world or promote a utopia (for example searching CEO on Google images – should it present images where 50% are female?) and avoiding algorithmic bias in machine learning. Amongst the myriad of AI ethical principles that exist, there are two key challenges for HAI solutions that involve explainable AI (or XAI) and algorithmic bias which are discussed in the following sections.

EXPLAINABLE AI

AI and algorithmic decision support systems provide recommendations or decisions that the end user should consider. Historical rule-based decision support algorithms could easily provide a recommendation alongside its decision logic which is important for building human-machine trust and for liability management. However, many machine learning based algorithms are considered opaque or black-box which is undesirable for end-users. After all, would you trust advice from another person if they could not provide rationale? Humans seek rationale from doctors and other advisors, perhaps they can go beyond rationale and seek levels of uncertainty, reassurance and additional opinions. Conversely, it must be said that humans often make decisions based on emotions and intuition which can involve 'tacit' knowledge which can not be easily explained. As a consequence, one might argue that requiring AI systems to always provide decision logic is unreasonable.

DARPA are leading XAI research and have published a program update [23]. There are different levels of an explainable AI. A key principle is transparency. However, from one perspective, an AI algorithm can be transparent without being explainable, e.g. one could be exposed to the weights in a deep neural network and whilst this arguably opens the box, it does not provide any explainability. According to a recent paper [3], "The ultimate goal of XAI should be to ensure that target users can understand the outputs, thus helping them improve their decision making efficiency". There is a need for AI to explain AI. The use of saliency maps have been used to improve the explainability of deep learning algorithms. In AI applications on images, these maps would show the end-user which features or pixels the algorithm used in making its automated decision. Saliency maps can perhaps be more appropriately called attention maps. Other machine learning techniques such as decision trees can easily provide decision explanations as they uses binary rules in a hierarchy to make the classification, however these rules can often be counter-intuitive. Techniques such as k-nearest neighbor can also provide rationale since they classify based on similarity to neighbors (prior classified cases). Other methods to support XAI might be to reveal similar cases and the number of similar cases used in the machine learning training dataset which would certainly make the AI system more transparent highlighting data shifts and any lack of training sample cases.

HAI solutions should involve HCI researchers who are responsible for designing explanation user interfaces [3]. These explanation interfaces should be co-designed with end-users. Not only should the interface provide explanations for any algorithmic decisions, but they should provide multiple levels of explanations allowing the end-user to interrogate the AI decision making process, perhaps right to the level of any datasets that were used in the machine learning development in order to reveal the full data provenance and its limitations. Giving the user the option to interrogate likely increases trust and integrity in human-machine collaborations. In addition, the explanation user interface should include metadata related to its algorithmic decision, akin to a mini-fact sheet [24]. Often algorithms present a decision to an end-user without informing the user of the typical accuracy achieved by the algorithm for that class/decision. This is important, for example if the algorithm is poor at detecting heart attacks then this should be presented to the user otherwise the user may assume that the 'computer is always right'. Also if there is a poor signal to noise ratio in the case being considered, this too should be presented as noise will affect the AI's decision making. Research also suggests that the presentation of an uncertainty or confidence index alongside the algorithm decision could calibrate the trust between the user and the AI (see section on automation bias). In addition, it might even be reasonable to present typical confounding decisions or likely alternative decisions to encourage the end-user to reason and contrast decisions [25]. This helps transfer the responsibility and accountability to the end-user. Presenting lists and multiple options is not a new challenge to HCI researchers who already use best practices derived from research (e.g. Hicks law [26] which tells us that the decision time is a log function of the number of options presented). These aforementioned methods also mitigate what is known as 'automation bias' which is discussed in the following section.

AUTOMATION BIAS AND AI UNCERTAINTY

Automation bias is when end-users of AI systems over trust or complacently accept almost every decision recommended by the AI algorithm [27]. It is akin to cognitive biases such as anchoring and confirmation bias. An obvious example of automation bias is when a car driver over trusts a satnav or a self-driving car to the point where they end up in a ditch. If an end-user feels incompetent then they will almost certainly depend on and comply with any AI decision, even when the AI system is categorically wrong or misleading. Automation bias is something that must be mitigated in HAI solutions, especially if we are to augment human abilities and not replace them.

Researchers have documented that clinicians can be easily misled by incorrect AI diagnostic statements when reading an electrocardiogram (ECG) [27]. Computerised algorithms that auto-interpret ECGs often provide simple diagnostic statements with little metadata and explainability. Research has shown that when an algorithmic decision is presented with a low certainty index, the end-user performs better and their trust in the machine is calibrated [28]. This is not a new concept, since aviation research invented an automated trust index to calibrate trust between pilots and automated cockpit features [29]. The UX community have already begun to visualize uncertainty [30] [31] [32] using different metaphors as simple indices may not always be appropriate or optimal for conveying AI uncertainty. A particular AI system in medicine known as progressive based ECG interpretation [33], guides the user through the decision making process and allows the human to make their own micro-decisions whilst the AI decisions are only presented at the end of the process to mitigate automation bias and anchoring at the start. The system also presents multiple competing diagnostic statements alongside explanations to facilitate transparency and enforce the user to reason and make the final decision (known as a differential diagnosis).

ALGORITHMIC BIAS

A sense of fairness is key to human justice and ethics. Obviously, terms like bias and discrimination can have different definitions in machine learning and are normally objectives since the engineer aims to develop an algorithm that 'discriminates' for example between those with and without cancer. However, algorithmic bias is when an algorithm exhibits socio-demographic discrimination that is unintended or intended. It is morally wrong for humans to exhibit this type of discrimination and so it is morally wrong for AI algorithms. It is even possible for physical products to be designed with an unintended or unconscious bias, for example a product could possibly be designed with unintended gender bias if the product is designed when only considering the size of typical male hands or other gender specific characteristics. Similarly, when building a machine learning classifier, one could naively use a male only dataset in training the algorithm and later find that the algorithm does not perform as well in a female population. There have been many examples of unintended algorithmic bias perhaps partly due to a lack of human centeredness at the start of the AI development lifecycle. The COMPAS algorithm is well documented to have been biased, resulting in greater judicial sentences for black males and females [34]. This was likely caused by a dataset that exhibited bias from the human annotation process. And there are many reports of other algorithms used for job recruitment, face detection and object detection that did not perform appropriately or equitably amongst certain groups. These biased cases are perhaps mainly due to using biased datasets in training the machine learning algorithms. Human bias 'in' unsurprisingly results in computer biased outputs, e.g. if you use white males in training an algorithm to detect faces then the algorithm should not be assumed to be generalizable. Having humans at the center at the outset of AI development can help pre-identify these risks and identify potentially unintended harmful consequences. To mitigate bias, both gender and racial variables can be protected variables that are unused in the machine learning phase. However, algorithms could be tested and compared amongst these different socio-demographic groups to ensure fairness. The outcome of mitigating algorithmic bias might mean that fairness decreases accuracy which can be supported since fairness should arguably trump accuracy. Both IBM [35] and Aequitas [36] have developed interactive toolkits to aid data scientists in detecting potential bias in their datasets prior to machine learning.

CONVERSATIONAL USER INTERFACES

This paper has thus far mainly discussed HAI solutions in the context of algorithms and AI based decision support. However, recent advances in AI and particular natural language processing have allowed humans to interact with computers using speech or via messaging akin to SMS. Natural user interfaces have been a topic for many years and included gesture based interfaces using depth cameras etc. but speech and chatbot based interfaces have become prevalent due to interaction channels like Facebook messenger and smart speakers like Alexa. Moreover, chatbots are now being proposed as a medium to support healthcare including symptom checking and support for mental wellbeing [37]. HAI is particularly important when developing conversational user interfaces (aka chatbots). This is mainly due to the fact that the UX design of a human-AI dialogue is a new challenge that requires a new form of interaction design. New UX and HAI methods are now needed to design chatbots. Nevertheless, there are other HAI challenges when deploying chatbots. This includes ethical challenges. The Computers As Social Actors (CASA) theory is a well-known theory that purports that humans treat computers as social actors analogous to how they treat other sentient beings [38]. Moreover, given chatbots go beyond traditional user interfaces and portray a higher degree of personhood and anthropomorphism, the CASA phenomenon is perhaps further reinforced. As a result, the interaction itself is persuading the end user that they are interacting with a human when in fact they are not. This is made worse if the chatbot has a convincing persona and does not disclose that it is a bot and not a human. Is this ethical? It is foreseeable that many users can befriend an intelligent chatbot and even develop a relationship with a bot. The chatbot can also simulate empathy, emotions and sentiment but any such emotional portrayal is not real with respect to the Chinese room argument in AI [6]. Is it ethical for a chatbot to pretend to care about the user? With these challenges in mind, it is important for HAI researchers to investigate what is acceptable in chatbot interaction and deployment with the priority of human good. A key concern with conversational smart speakers is of course unwarranted surveillance and data privacy. And more generally, researchers and users need to discuss in great detail the tradeoff between utility and privacy. For example, Mulvenna et al. consider eldercare using a home-video surveillance system which has important merits including peace of mind for next of kin whilst also potentially being life-saving (e.g. in the case of a fall) but at the expense of giving up privacy and perhaps autonomy [39].

DEMOCRATISIING AI

AI and particularly machine learning are becoming more democratized [40], although even chatbot design is arguably democratized using tools that conceal AI processes and provide a novice-friendly interface for dialogue design (referring to tools such as Bot Society [41] and Chat-fuel [42]). Historically, in the same way that the Internet was democratized using graphical user interfaces (in the form of Internet browsing software), so is machine learning democratized. Interactive machine learning or end-user machine learning systems allow novice users to build algorithms without writing code and by simply uploading their dataset in a CSV file [40]. With the emergence of usable Cloud ML based tools, the democratization of machine learning is inevitable. However, allowing any domain expert without any appreciation of statistical and machine learning principles is a likely risk for harmful unintended consequences. Possibly a general AI literacy could be embedded in school curriculums akin to numeracy or English literacy. There is also an argument for AI literacy to be integrated into specific disciplines and institutions such as medical schools and perhaps even law. HAI researchers have the opportunity to improve the UX of interactive machine learning user interfaces and guide the 'responsible' democratization of AI. HAI researchers should place the human at the center when considering the ramifications of such end-user led AI systems.

WHAT ABOUT USING AI TO ENHANCE THE UX

In alignment with HAI, this paper has discussed how human-centered design can improve AI, however it is known that AI can bilaterally be used to enhance human-centered designs/UX. AI and affective computing technologies such as computer vision for facial expression analysis [43], electroencephalography [44] and wearable sensors [45] can provide automatic insight into the user's emotions and psychophysiology providing important predictions about the UX in real-time. In addition, even eye tracking and AI can be combined to understand human attention and determine cognitive states such as confusion [46]. Eye tracking with AI could also be used to determine user uncertainty and even user competence when interpreting a data visualization [47]. This in turn could be used in a decision support system and be exploited by the explanation user interface in an adaptive XAI framework [48]. AI techniques can also be used to mine user interaction log data revealing key insights into user interaction patterns, habits and even reveal user archetypes using k-means clustering [49] [50] [51] [52].

CONCLUSION

HAI research is essential to ensure that AI solutions responsibly prioritize the end user and human values. HAI is a critical initiative and has challenges that require a multi-disciplinary effort involving experts from the social sciences, moral philosophy, law, cognitive science, decision science, psychology, anthropology and of course the HCI and AI disciplines. HAI research will address key human-AI challenges that will enhance the integrity and trust between humans and machines allowing for appropriate human-machine integration (also called human machine teaming). Now is the time to put human values at the center of AI systems.

REFERENCES

1. Lynch, S. Andrew Ng: Why AI Is the New Electricity. https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity.
2. Floridi, L. The Fourth Revolution: How the Infosphere is Reshaping Human Reality. OUP Oxford, 2014.
3. Xu, W. Toward Human-Centered AI: A Perspective from Human-Computer Interaction. Interactions 26.4 (2019): 42-6.
4. Cambridge Dictionary. Intelligence. https://dictionary.cambridge.org/dictionary/english/intelligence
5. Cambridge Dictionary. Artificial intelligence. https://dictionary.cambridge.org/dictionary/english/artificial-intelligence
6. Searle, J. The Chinese Room Argument. Encyclopedia of cognitive science (2006)
7. Miller, G. A. The Magical Number Seven, Plus Or Minus Two: Some Limits on our Capacity for Processing Information. Psychological review 63.2 (1956): 81

8. Kahneman, D. Thinking, Fast and Slow. Macmillan, 2011.
9. Farooq, U. and Grudin, J. Human-Computer Integration. *Interactions* 23.6 (2016): 27-32.
10. McAllister, P. et al. Combining Deep Residual Neural Network Features with Supervised Machine Learning Algorithms to Classify Diverse Food Image Datasets. *Computers in biology and medicine* 95 (2018): 217-33.
11. Sittig, D. F. et al. Grand Challenges in Clinical Decision Support. *Journal of Biomedical Informatics* 41.2 (2008): 387-92.
12. Domingos, P. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake our World. Basic Books, 2015.
13. Amershi, S. et al. Guidelines for Human-AI Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, 2019. 3.
14. Mulvenna, M. et al. Participatory Design-Based Requirements Elicitation Involving People Living with Dementia Towards a Home-Based Platform to Monitor Emotional Wellbeing. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE , 2017. 2026-2030.
15. Cleland, B. et al. Usability Evaluation of a Co-Created Big Data Analytics Platform for Health Policy-Making. International Conference on Human-Computer Interaction. Springer, 2019. 194-207.
16. Torney, H. et al. A Usability Study of a Critical man–machine Interface: Can Layperson Responders Perform Optimal Compression Rates when using a Public Access Defibrillator with Automated Real-Time Feedback during Cardiopulmonary Resuscitation? IEEE Transactions on Human-Machine Systems 46.5 (2016): 749-54.
17. Currie, J. et al. Eye Tracking the Visual Attention of Nurses Interpreting Simulated Vital Signs Scenarios: Mining Metrics to Discriminate between Performance Level. IEEE Transactions on Human-Machine Systems 48.2 (2017): 113-24.
18. Chatila, R. et al. The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems [Standards]. IEEE Robotics & Automation Magazine 24.1 (2017): 110
19. Awad, E. et al. The Moral Machine Experiment. Nature 563.7729 (2018): 59.
20. McNamara, A., Smith, J. and Murphy-Hill, E. Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?. Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, 2018. 729-733.
21. Mulvenna, M., Boger, J and Bond, R. Ethical by Design: A Manifesto. Proceedings of the European Conference on Cognitive Ergonomics 2017. ACM, 2017. 51-54.
22. Eyal, N. Hooked: How to Build Habit-Forming Products. Penguin UK, 2014.
23. DARPA. Explainable Artificial Intelligence (XAI). 2017. www.darpa.mil/attachments/XAIProgramUpdate.pdf
24. Mojsilovic, A. Factsheets for AI Services. 2018. www.ibm.com/blogs/research/2018/08/factsheets-ai/
25. Novotny, T. et al. The Role of Computerized Diagnostic Proposals in the Interpretation of the 12-Lead Electrocardiogram by Cardiology and Non-Cardiology Fellows. International journal of medical informatics 101 (2017): 85-92.
26. Proctor, R. W., and Schneider D.W. Hick's Law for Choice Reaction Time: A Review. *The Quarterly journal of experimental psychology*. (2017): 1-56.
27. Bond, R. R., et al. Automation Bias in Medicine: The Influence of Automated Diagnoses on Interpreter Accuracy and Uncertainty when Reading Electrocardiograms. Journal of electrocardiology 51.6 (2018): S6-S11.
28. Knoery, C.R. et al. SPICED-ACS: Study of the Potential Impact of a Computer-Generated ECG Diagnostic Algorithmic Certainty Index in STEMI Diagnosis: Towards Transparent AI. *Journal of electrocardiology* (2019)
29. McGuirl, J. M. and Sarter, N. B. Supporting Trust Calibration and the Effective use of Decision Aids by Presenting Dynamic System Confidence Information. Human factors 48.4 (2006): 656-65.
30. Brown, R. Animated Visual Vibrations as an Uncertainty Visualisation Technique. Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia. ACM , 2004. 84-89.
31. Guo, S. et al. Visualizing Uncertainty and Alternatives in Event Sequence Predictions. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM , 2019. 573.
32. McCormack, J. et al. In a Silent Way: Communication between AI and Improvising Musicians Beyond Sound. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM , 2019. 38.
33. Cairns, A. W. et al. A Computer-Human Interaction Model to Improve the Diagnostic Accuracy and Clinical Decision-Making during 12-Lead Electrocardiogram Interpretation. Journal of Biomedical Informatics 64 (2016): 93-107.

34. Angwin, J. et al. Machine Bias. 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

35. Bellamy, R. et al. AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. *arXiv preprint arXiv:1810.01943* (2018).

36. Saleiro, P. et al. Aequitas: A Bias and Fairness Audit Toolkit. *arXiv preprint arXiv:1811.05577* (2018)

37. Cameron, G. et al. Towards a Chatbot for Digital Counselling. Proceedings of the 31st British Computer Society Human Computer Interaction Conference. BCS Learning & Development Ltd., 2017. 24.

38. Reeves, B, and Nass, C.I. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge university press, 1996.

39. Mulvenna, M., Hutton, A., Coates, V., Martin, S., Todd, S., Bond, R., and Moorhead, A., Views of caregivers on the ethics of assistive technology used for home surveillance of people living with dementia. Neuroethics 10, no. 2 (2017): 255-266.

40. Bond, R. et al. Democratisation of Usable Machine Learning in Computer Vision. *arXiv preprint arXiv:1902.06804* (2019).

41. Botsociety. Design your next chatbot or voice experience. 2019. https://botsociety.io.

42. Chatfuel. Increase sales, reduce costs, and automate support on Facebook: Chatfuel is the world's leading chatbot platform for Facebook Messenger. 2019. https://chatfuel.com.

43. Samara, A. et al. Sensing Affective States using Facial Expression Analysis. International Conference on Ubiquitous Computing and Ambient Intelligence. Springer , 2016. 341-352.

44. Menezes, M. L. R. et al. Towards Emotion Recognition for Virtual Environments: An Evaluation of EEG Features on Benchmark Dataset. Personal and Ubiquitous Computing 21.6 (2017): 1003-13.

45. Currie, J. et al. Wearable Technology-Based Metrics for Predicting Operator Performance during Cardiac Catheterisation. International journal of computer assisted radiology and surgery 14.4 (2019): 645-57.

46. Samara, A. et al. Adaptive User Experience Based on Detecting User Perplexity. British HCI Conference 2018. 2018.

47. McLaughlin, L. et al. Computing Eye Gaze Metrics for the Automatic Assessment of Radiographer Performance during X-Ray Image Interpretation. International journal of medical informatics 105 (2017): 11-21.

48. McLaughlin, L. et al. The Potential of Hybridising Interactive Eye Tracking Technology with Decision Support in Medical Image Interpretation. British HCI Conference. 2018.

49. McCauley, C. O. et al. Evaluating User Engagement with a Reminiscence App using Cross-Comparative Analysis of User Event Logs and Qualitative Data. Cyberpsychology, Behavior, and Social Networking 22.8 (2019): 543-51.

50. Bond, R. et al. Exploring Temporal Behaviour of App Users Completing Ecological Momentary Assessments using Mental Health Scales and Mood Logs. Behaviour & Information Technology (2019): 1-12.

51. O'Neill, S. et al. Data Analytics of Call Log Data to Identify Caller Behaviour Patterns from a Mental Health and Well-being Helpline. Health informatics journal (2018)

52. Mulvenna, M. et al. HILDA-A Health Interaction Log Data Analysis Workflow to Aid Understanding of Usage Patterns and Behaviours. The 2nd Symposium on Social Interactions in Complex Intelligent Systems (SICIS) at Artificial Intelligence and Simulation of Behaviour Convention (AISB-2018). 2018.