What Can Crowd Computing Do for the Next Generation of AI Systems?

Ujwal Gadiraju and Jie Yang
Web Information Systems
Delft University of Technology
The Netherlands
{*u.k.gadiraju, j.yang-3*}*@tudelft.nl*

Abstract

The unprecedented rise in the adoption of artificial intelligence techniques and automation in many contexts is concomitant with shortcomings of such technology with respect to robustness, interpretability, usability, and trustworthiness. *Crowd computing* offers a viable means to leverage human intelligence at scale for data creation, enrichment, and interpretation, demonstrating a great potential to improve the performance of AI systems and increase the adoption of AI in general. Existing research and practice has mainly focused on leveraging crowd computing for training data creation. However, this perspective is rather limiting in terms of how AI can fully benefit from crowd computing. In this vision paper, we identify opportunities in crowd computing to propel better AI technology, and argue that to make such progress, fundamental problems need to be tackled from both computation and interaction standpoints. We discuss important research questions in both these themes, with an aim to shed light on the research needed to pave a future where humans and AI can work together seamlessly, while benefiting from each other.

1 Introduction

Artificial intelligence techniques and machine learning in particular, are drastically changing our lives through technological revolutions across several domains such as transportation, health, finance, education, and manufacturing. AI systems at the forefront of such innovations have garnered a growing barrage of concerns, not only due to issues pertaining to performance – such systems have been observed to easily fail in situations slightly different from those encountered in the training instances [1] – but also due to the ethical and societal implications that arise as a result of using these systems [7, 6, 3, 9, 35].

Problems exist and manifest both in AI systems and in the interaction between end users with such systems. On the one hand, machine learning models have been criticized for the lack of robustness, fairness, and transparency [26, 14, 20]. Such model-related problems can be attributed to data problems to a large extent: for models to learn comprehensive, fine-grained, and unbiased patterns, they have to be trained on a large number of high-quality data instances with the right distribution that is representative of real application scenarios. Creating such data is not only a long, laborious, and expensive process, but sometimes even impossible when the data is extremely imbalanced or the distribution constantly evolves over time. On the other hand, AI systems often demonstrate inconsistent and unpredictable behavior that can confuse users, erode their confidence, and may eventually lead to the abandonment of the systems [10, 2]. Systems with such behavior violate established usability guidelines of traditional user interface design (e.g., minimizing the unexpected changes), posing an ever bigger challenge for the design of intuitive and effective user interfaces. The problem is further complicated by the variability of interfaces for AI systems, ranging

from the conventional Web-based interfaces to the emerging Voice-based ones. There is a limited understanding of how users perceive automated decisions and how their behavior is mediated or influenced by the interfaces.

The two schools of challenges pertaining to AI systems, characterised as computational and inter-actional ones, are in fact highly related to each other. From the computation perspective, a better understanding of user interactions can help identify the focal point of system development and potentially spark new research directions. A prominent example is machine learning interpretability, inspired by the observation that explainable results are more in demand by users than highly accurate ones. From the interaction perspective, more robust and interpretable systems can help build trust and increase system uptake [19, 40]. As AI systems become more commonplace, people must be able to make sense of their encounters and interpret their interactions with such systems.

A promising approach to address both computational and interactional challenges while building AI systems, is the use of *crowd computing*, which offers a viable means to engage a large number of human participants in data related tasks and in user studies.

Crowd computing has been conceptualised in various ways – as being related to crowdsourcing, human computation, social computing, cloud computing and mobile computing [31]. Over the last decade there has been a steady rise in the adoption of crowd computing solutions across a variety of domains [12]. In the context of overcoming the computational and interactional challenges facing the current generation of AI systems, recent work has shown how crowd computing can be leveraged to either debug noisy training data in machine learning systems [46], understand which machine learning models are more congruent to human understanding in particular tasks [22, 47], or to advance our understanding of how AI systems can influence human behavior [15].

Based on the existing evidence of how crowd computing can play an important role in tackling computational and interactional challenges in developing new-age AI systems, in this vision paper, we highlight research themes that need to be pursued to ensure that AI systems can create a future where we are better off than we currently are – both as individuals and as a society.

2 Crowd Computing and Human-Centered AI

In this section, we discuss important challenges that need to be addressed to make advances in the next generation of AI systems from two main standpoints – (1) Human-in-the-loop AI, and (2) Human-AI interaction. The former concerns the computational role of humans for AI, i.e., AI *by* humans, while the latter concerns the interactional role of humans with AI systems, i.e., AI *for* humans.

2.1 Human-in-the-Loop AI

In what follows, we analyze the fundamental computational challenges in the quest for robust, interpretable, and hence trustworthy AI systems. We argue that to tackle such fundamental challenges, research should explore a novel crowd computing paradigm, which we refer to as "crowd conceptual computing". In such form of crowd computing, crowd workers can contribute knowledge at the conceptual level; this comes in contrast to the current paradigm where crowd intelligence is utilised on a per-datum basis, e.g., labelling and debugging individual data instances.

Robust AI by Crowds. Machine (deep) learning models have proven to be "shallow" – they often learn spurious correlations in the data – and "brittle" – they are unable to make sense of situations slightly different from the training data. Consequently, current AI systems often fail when required to make predictions on data beyond the training distribution, which is of crucial need in practice. Those issues constitute what is now referred to as the robustness or reliability issue, generally viewed as a main obstacle for wide deployment of AI systems [13, 26].

Robust AI requires models to be encapsulated with causality and better generalisation ability, which are the main advantageous characteristics of conventional symbolic AI methods focusing on knowl-edge representation and reasoning. Recent discussions in the AI community has therefore converged to the idea of developing neurosymbolic methods that benefit from both the robustness of symbolic methods and the flexibility of deep learning. Few discussions have, however, touched upon the questions of what knowledge is required, and where and how to obtain such knowledge. Historical

research in expert systems has shown that the amount of knowledge for a specific task can be very large that can easily go beyond readily available knowledge bases and what individuals can provide.

Building on top of the Web, crowd computing systems can reach an unprecedented number of people, thus offering a feasible approach to leveraging human intelligence at scale for knowledge creation. Classical per-datum based crowd computing techniques, however, are ill suited for the problem, when the outcome contributions are data instances as opposed to the knowledge in need. Take for example, unknown unknowns of machine learning, which is a major class of errors produced by unreliable AI systems. Such errors are caused by missing or underrepresented concepts in the model. Each of those concepts can be instantiated as various data instances. Crowd computing has been used to detect unknown unknowns and fix them by contributing instances for training data augmentation. Such an approach however, is limited not only in terms of efficiency but also effectiveness, due to the intrinsic shallowness and brittleness of machine learning models.

Interpretable AI by Crowds. Interpretability in AI refers to "the ability to explain or to present in understandable terms to a *human*" [14] how the system makes predictions for individual instances (i.e., local interpretability) or how the system works with respect to a specific class of instances (i.e., global interpretability). The problem is closely related to the robustness problem: being able to inspect what an AI system has learned is useful to identify what it has not.

Humans as the object in the definition of AI interpretability implies the following key requirements for the design of interpretability methods: i) presentation of interpretations need to match humans' mental representations of *concepts* as humans understand the world through concepts that are associated with observable properties; ii) interpretability methods also need to take the flexible needs of humans as explanation consumers into account, allowing humans to gain insights about system behavior with multi-concept queries that involve the (non-)presence of multiple concepts flexibly named by humans. Existing interpretability methods, however, fail to meet those requirements. Existing local methods generally generate explanations by highlighting relevant input units – e.g., words in a sentence or pixels in an image [38, 39], which require efforts from human users to make sense of; global methods generate interpretations representing relevant concepts with a set of examples – e.g., pieces of text or image patches [23, 18], which do not support multi-concept questions for in-depth model understanding.

A natural approach to fill the semantic gap is involving humans in the interpretation process. Similar to crowd computing for robust AI, where the goal is to characterise what a model has not learned, crowd computing for interpretable AI seeks to explain what a model has learned. The latter again requires crowd computing on the conceptual level for human interpretability and query flexibility.

2.2 Human-AI Interaction

Principles for human-AI interaction have been discussed in the HCI community for several years [2]. However, in the light of recent advances in AI and the growing role of AI technologies in human-centered applications, a deeper exploration is the need of the hour. As different research communities aim to progress in this direction, we need to explore and develop fundamental methods and techniques to harness the virtues of AI in a manner that is beneficial and useful to the society at large. Crowd computing methods can allow us to carry out large-scale behavioral experiments and randomized controlled trials [16, 5], that are necessary to representatively study, and make advances in our understanding of Human-AI interaction. We foresee crowd computing to play a pivotal role in addressing important challenges in the following themes.

Congruence of machine learning models with human understanding. Complex machine learning models are deployed in several critical domains including healthcare and autonomous vehicles nowadays, albeit as functional blackboxes. Models which correspond to human interpretation of a task are more desirable in certain contexts and can help attribute liability, build trust, expose biases and in turn build better models [41]. It is therefore of paramount importance to understand how and which models conform to human understanding of various tasks. What is the relationship between expectations and trust when humans interact with AI systems? How can effective machine learning models be built, while conforming to human expectations?

Explaining AI systems to humans and supporting decision-making. AI systems offer computational powers that vastly transcend human capabilities. In conjunction with the ability to autonomously detect data patterns and derive superior predictions, AI systems are projected to complement, trans-

form and in several cases even substitute human decision-makers. This process broadly revolutionizes all the relevant stages of economical, political and societal decision-making. Despite these dynamics, the impact of AI systems on human behavior remains largely unexplored. We need to address this crucial gap by carrying out interdisciplinary research to advance the current understanding of impact of AI systems on human decision-making. Despite the recent surge in interpreting decisions of complex machine learning models to explain their actions to humans [28], little is known about what constitutes a sufficient explanation from a user's vantage point and the contextual settings. Moreover, how such criteria varies across the landscape of different stakeholders interacting with AI systems needs to be better understood. Different individuals and user groups alike, can have varying attitudes towards the same technology due to a range of factors including their familiarity with the technology [11], individual traits [37], cultural differences [21], or contexts [42]. How can explanations be adapted and personalized across diverse stakeholders with an aim to improve the effectiveness of their interaction with AI systems?

3 A Vision for the Future

Open-ended Crowd Knowledge Creation. For the purpose of both, robust or interpretable AI, knowledge creation in real-world machine learning tasks is a complex, open-ended task. Research on this problem needs to investigate not only the extraction of knowledge from the training data and model, but also the creation of any knowledge crowds deem as relevant, which can easily go beyond knowledge encoded in existing knowledge sources. Such a problem is related to multiple ongoing research lines, such as crowd knowledge creation [43], complex task design [45, 17], open-ended crowdsourcing [4], machine intelligence for human work [44, 30], etc. In the crowd computing community specifically, it has been widely recognized that the future of research in this field should enable crowd work that is complex, collaborative, and sustainable, such that human workers can both earn and learn from their work in an enjoyable manner [24]. Aligned with this goal, we advocate a novel crowd computation paradigm aiming at bringing human computation to the conceptual level for knowledge creation. The open-endedness of this new kind of knowledge creation tasks further calls for research on leveraging the cognitive ability, and creativity in particular, of human workers.

Crowd knowledge creation for tackling problems in AI systems further contributes to the vision of a human-AI collaborative future: by acquainting human workers with the strengths and weaknesses of AI algorithms through knowledge creation tasks, we envision a future where human workers and AI can work together seamlessly while benefiting from each other.

Conversational Human-AI Interaction. Conversational interfaces have been argued to have advantages over traditional GUIs due to having a more human-like interaction [29]. Recent work in crowd computing has shown that conversational interfaces can lead to an increased satisfaction and engagement in online work settings when compared to conventional web interfaces [27, 32, 33]. Conversational interfaces have also been found to be conducive for memorable interactions with information retrieval systems [34]. Messaging applications such as Telegram, Facebook Messenger, and Whatsapp, are regularly used by an increasing number of people mainly for interpersonal communication and coordination purposes [25]. Users across cultures, demographics, and technological platforms are now familiar with their minimalist interfaces and functionality. By building conversational interfaces that people may generally be more familiar with, for their interaction with AI systems, we can potentially lower the barrier for the adoption of such systems.

Trust plays a central role in Human-AI interaction – the adoption and successful utilization of AI systems is mediated by trust. Therefore, it is important to investigate whether novel conversational interfaces can be built to facilitate trust in AI systems. Several factors have been identified to be capable of increasing trust toward conversational agents including appearances, voice features, and communication styles [36]. These findings suggest that human interaction with AI systems can potentially be enhanced by leveraging conversational interfaces to improve engagement, and build trust. By facilitating a more natural type of interaction, conversational interfaces can also lower the barrier for crowd computing to address the robustness and interpretability issues of AI systems, in particular conversational systems themselves as a representative type of AI-complete systems [8].

Crowd computing offers promising means to overcome fundamental challenges in computation and interaction, and herald a new generation of human-centered AI systems.

References

[1] Death of Elaine Herzberg. Wikipedia [Accessed: 2020-04-12].

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. Guidelines for human-ai interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–13, 2019.

[3] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Propublica [Online; posted: 23-May-2016].

[4] Ines Arous, Jie Yang, Mourad Khayati, and Philippe Cudré-Mauroux. Opencrowd: A human-ai collaborative approach for finding social influencers via open-ended answers aggregation. In *Proceedings of The Web Conference 2020*, pages 1851–1862, 2020.

[5] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. The moral machine experiment. *Nature*, 563(7729):59–64, 2018.

[6] Thomas Beardsworth and Nishant Kumar. Who to sue when a robot loses your fortune. https://www.bloomberg.com/news/articles/2019-05-06/who-to-sue-when-a-robot-loses-your-fortune. Bloomberg [Online; posted: 05-May-2019].

[7] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 'it's reducing a human being to a percentage' perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems*, pages 1–14, 2018.

[8] Maxime Chevalier-Boisvert, Dzmitry Bahdanau, Salem Lahlou, Lucas Willems, Chitwan Saharia, Thien Huu Nguyen, and Yoshua Bengio. Babyai: First steps towards grounded language learning with a human in the loop. *arXiv preprint arXiv:1810.08272*, 2018.

[9] Jeffrey Dastin. Amazon scraps secret ai recruiting tool that showed bias against women. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. Reuters [Online; posted: 09-October-2018].

[10] Maartje De Graaf, Somaya Ben Allouch, and Jan Van Diik. Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 224–233. IEEE, 2017.

[11] Ewart J De Visser, Samuel S Monfort, Ryan McKendrick, Melissa AB Smith, Patrick E McKnight, Frank Krueger, and Raja Parasuraman. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3):331, 2016.

[12] Gianluca Demartini, Djellel Eddine Difallah, Ujwal Gadiraju, and Michele Catasta. An introduction to hybrid human-machine information systems. *Foundations and Trends in Web Science*, 7(1):1–87, 2017.

[13] Thomas G Dietterich. Steps toward robust artificial intelligence. *AI Magazine*, 38(3):3–24, 2017.

[14] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

[15] Alexander Erlei, Franck Awounang Nekdem, Lukas Meub, Avishek Anand, and Ujwal Gadiraju. Impact of algorithmic decision making on human behavior: Evidence from ultimatum bargaining. In *Proceedings of the 8th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2020)*, 2020.

[16] Ujwal Gadiraju, Sebastian Möller, Martin Nöllenburg, Dietmar Saupe, Sebastian Egger-Lampl, Daniel Archambault, and Brian Fisher. Crowdsourcing versus the laboratory: towards human-centered experiments using the crowd. In *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*, pages 6–26. Springer, 2017.

[17] Ujwal Gadiraju, Jie Yang, and Alessandro Bozzon. Clarity is a worthwhile quality: On the role of task clarity in microtask crowdsourcing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pages 5–14, 2017.

[18] Amirata Ghorbani, James Wexler, James Y Zou, and Been Kim. Towards automatic concept-based explanations. In *Advances in Neural Information Processing Systems*, pages 9277–9286, 2019.

[19] Ella Glikson and Anita Williams Woolley. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, (ja), 2020.

[20] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*, pages 3315–3323, 2016.

[21] Kerstin Sophie Haring, David Silvera-Tawil, Yoshio Matsumoto, Mari Velonaki, and Katsumi Watanabe. Perception of an android robot in japan and australia: A cross-cultural comparison. In *International conference on social robotics*, pages 166–175. Springer, 2014.

[22] Sungsoo Ray Hong, Jessica Hullman, and Enrico Bertini. Human factors in model interpretability: Industry practices, challenges, and needs. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–26, 2020.

[23] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, et al. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In *International conference on machine learning*, pages 2668–2677, 2018.

[24] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1301–1318, 2013.

[25] Rich Ling and Chih-Hui Lai. Microcoordination 2.0: Social coordination in the age of smartphones and messaging apps. *Journal of Communication*, 66(5):834–856, 2016.

[26] Gary Marcus. The next decade in ai: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*, 2020.

[27] Panagiotis Mavridis, Owen Huang, Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. Chatterbox: Conversational interfaces for microtask crowdsourcing. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, pages 243–251, 2019.

[28] Brent Mittelstadt, Chris Russell, and Sandra Wachter. Explaining explanations in ai. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 279–288, 2019.

[29] Robert J Moore, Raphael Arar, Guang-Jie Ren, and Margaret H Szymanski. Conversational ux design. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 492–497, 2017.

[30] Natalia Ostapuk, Jie Yang, and Philippe Cudré-Mauroux. Activelink: deep active learning for link prediction in knowledge graphs. In *The World Wide Web Conference*, pages 1398–1408, 2019.

[31] Kalpana Parshotam. Crowd computing: a literature review and definition. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, pages 121–130, 2013.

[32] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. Estimating conversational styles in conversational microtask crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–23, 2020.

[33] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. Improving worker engagement through conversational microtask crowdsourcing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[34] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. Towards memorable information retrieval. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, pages 69–76, 2020.

[35] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. Machine behaviour. *Nature*, 568(7753):477–486, 2019.

[36] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction*, pages 1–16, 2020.

[37] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1–8. IEEE, 2015.

[38] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Workshop Track Proceedings*, 2014.

[39] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda B. Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *CoRR*, abs/1706.03825, 2017.

[40] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, and Aad van Moorsel. The relationship between trust in ai and trustworthy machine learning technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 272–283, 2020.

[41] Jiaxuan Wang, Jeeheh Oh, Haozhu Wang, and Jenna Wiens. Learning credible models. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2417–2426, 2018.

[42] Lin Wang, Pei-Luen Patrick Rau, Vanessa Evers, Benjamin Krisper Robinson, and Pamela Hinds. When in rome: the role of culture & context in adherence to robot recommendations. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 359–366. IEEE, 2010.

[43] Jie Yang, Alessandro Bozzon, and Geert-Jan Houben. Knowledge crowdsourcing acceleration. In *International Conference on Web Engineering*, pages 639–643. Springer, 2015.

[44] Jie Yang, Thomas Drake, Andreas Damianou, and Yoelle Maarek. Leveraging crowdsourcing data for deep active learning an application: Learning intents in alexa. In *Proceedings of the 2018 World Wide Web Conference*, pages 23–32, 2018.

[45] Jie Yang, Judith Redi, Gianluca Demartini, and Alessandro Bozzon. Modeling task complexity in crowdsourcing. In *Proceedings of the 4th AAAI Conference on Human Computation and Crowdsourcing*, 2016.

[46] Jie Yang, Alisa Smirnova, Dingqi Yang, Gianluca Demartini, Yuan Lu, and Philippe Cudré-Mauroux. Scalpel-cd: leveraging crowdsourcing and deep probabilistic modeling for debugging noisy training data. In *The World Wide Web Conference*, pages 2158–2168, 2019.

[47] Zijian Zhang, Jaspreet Singh, Ujwal Gadiraju, and Avishek Anand. Dissonance between human and machine understanding. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23, 2019.