Hybrid AI for Collaborative Data Science

Sophy Chhong and Lina Witzel

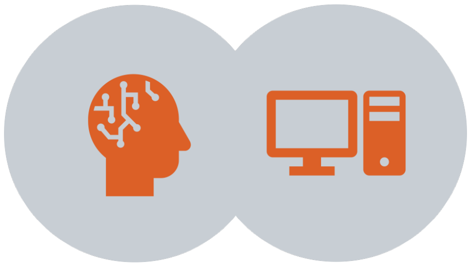
University of Zurich
Seminar: Current Trends in Distributed and Dynamic Information Systems
Fall Semester 2020

1 Introduction

Artificial Intelligence (AI) and Machine Learning (ML) are prevalent terms in recent years, most notably due to major advances in performance and diverse applications [12]. Intelligent assistants, outcome predictions and search engines are just a few examples that make up every day life with many more to come. Although this technological superiority is impressive, automation brought by AI gave birth to distrust and fear, the worst one being the replacement of human intelligence and work [13]. To address the fear and also for its beneficial nature, researchers are invested in augmenting humans with AI instead, creating Hybrid AI or Hybrid Intelligence (HI) [1]. Hybrid AI focuses on the complementary skill-set of both human and machine agents who together can reach a goal neither can reach on its own [1]. Although, research on Hybrid AI is rather young and many challenges and obstacles need to be overcome, areas such as collaborative HI, adaptive HI, responsible HI and explainable HI show immense potential to humans [1].

One potential area for which HI can be useful is data science, where demands of services exceed qualified practitioners [13]. In their study, Wang et al. [13] identified mixed feelings by the data scientists, some are afraid of automation as it could negatively impact their work, others see AI as a beneficial collaborator. Nonetheless, all of them think that AI is an unavoidable future since the biggest contribution is reducing the time in which data scientist have to spend with the data instead of its insight. Transparency as the main component, could reduce initial fear and mistrust in the adoption of AI, with accuracy and quality of results also playing an important role. The researchers believe that AI can take up different roles: as a teacher, a collaborator and as a data scientist. Without eliminating the data science practitioners, their work instead shifts from preparing data to guiding AI with domain expertise and decisions.

In this work, we want to focus on collaborative HI where the interest is in the mutual benefiting work synergy of AI and humans. We want to explore what the specific challenges and advantages are in the collaboration between human and machine when they complement their skills and capabilities [1]. First, we define the terms AI, Hybrid AI and collaborative Hybrid AI and name example applications for each of these. Afterwards, we introduce characteristic machine and human factors before we look at their collaboration. Lastly, we discuss limitations, future research and end with a summary of related works and some lessons learned.



2 Definitions and Example Applications

2.1 Artificial Intelligence

A variety of definitions for Artificial Intelligence (AI) exist but the most generic one would be: systems that have "the ability to accomplish complex goals, learn, reason, and adaptively perform effective actions within an environment." [9, p.638]. The underlying Machine Learning (ML) techniques give it the capability to interact and react to its environment [9]. Furthermore, imitation or replication of human intelligence can be defined as the overall goal of AI which can be, for example, tested by the Turing test [9]. AI has gained a lot of interest in recent years, most due to the advances in ML which allowed technology such as speech interaction, intelligent assistants, autonomous cars and many more to rise [12]. In research circles, the possibilities and the technical feasibility AI and ML are shown off in conferences, presenting better and better performance [12].

2.2 Hybrid AI

Exceeding the definition of AI, Hybrid Intelligence (HI) or Hybrid AI explicitly includes the involvement of humans. Akata et al. [1, p. 19] therefore define HI: "as the combination of human and machine intelligence, augmenting human intellect and capabilities instead of replacing them, to make meaningful decisions, perform appropriate actions, and achieve goals that were unreachable by either humans or machines alone."

The amount of the human and machine involvement can vary depending on the application (Figure 1). While conversational bot assistants (e.g. Alexa[1]) can be seen as purely automated applications, enterprise data enrichment include active learning of the machine based on the input of human agents. E.g., the platform Appen (formerly CrowdFlower)[2] combines human-labelled training data, easy to deploy machine learning, and human-in-the-loop capabilities to help companies create value from their unstructured data. For artificial assistants like Clara[3], a virtual employee helping you organize your schedule, the

[1] https://developer.amazon.com/en-US/alexa
[2] https://appen.com
[3] https://claralabs.com

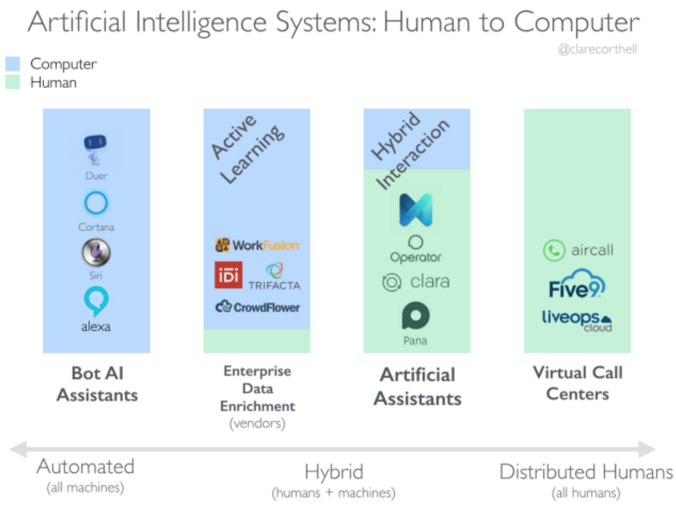


Fig. 1: Artificial Intelligence systems and the proportional human and machine involvement [8]

human involvement is even higher, resulting in a greater amount of hybrid interaction. On the end of distributed humans, there are virtual call centers (e.g. aircall[4]), purely supporting human interaction.

2.3 Collaborative Hybrid AI

As seen in the previous chapter, the amount of involvement of the human and machine agents can highly vary. With their involvement, the amount and types of interactions change as well. We focus on collaborative Hybrid AI which according to Akata et al. [1, p. 20]: "goes beyond the established notions of human-in-the-loop machine learning or interactive AI by aiming for reciprocity between human and computer agents." In the context of data science this reciprocity could stand for the machine's ability to guide the human (e.g. by visualising predictions for decision-making) and at the same time the human being able to guide the machine (e.g. by adjusting models or predictions according to expert knowledge). With regard to the reciprocal interaction, there are several machine and human factors that may influence collaborative Hybrid AI, further discussed in the following chapters.

[4] https://aircall.io

3 Machine Factors

3.1 Automation

Due to the rising demand of data science services and not enough qualified practitioners, much of the current work on AI is focused on automating different steps of the workflow [13]. The most challenging and time-consuming step is the pre-processing of data which includes the finding, cleaning, feature engineering, etc., of data [13]. Although humans are capable of doing this task, AI can be more time-efficient and consistent in its approach, being especially good in computing large data for repetitive tasks [9], allowing humans to do more productive work which is unsuited for machines [13]. With the fast technological advances, even faster and larger data can be handled which is extremely essential as the amount of data continues to increase [1]. Even though automation is inevitable and brings many positive aspects, data science practitioners have mixed feelings as they fear becoming obsolete [13].

3.2 Complex Computation

In complex situation such as decision making, AI can be of great use due to its "superior quantitative, computational, and analytical capabilities" [11, p. 581]. Its major advantage is the brute force in systematically processing and analyzing large data sets [11]. Another strong point is the ability of AI to recognize complex underlying patterns in data which a human would not be able to detect [9]. Not surprising, a long history of research has been invested in this area [9] resulting in advances in technical capabilities such as performance, speed and accuracy [12]. With deep learning methods, even better model and hyperparameter combinations can be selected to create the best performing ML model [13].

3.3 Customizability

AI can be viewed as an additional tool in decision making [1]. As such it can be customized to the needs of data scientists - the main users - who demand the feasibility to explore and experiment with data [3]. The possibilities vary from level of summarization, visualization, to boundary decisions and more [3].

Machines' unique point to be customizable allows not only professionals but novice users to efficiently work with AI [6], opening the door to a large user base. This is needed as AI shall not only be used by experts, but by anyone. As such large technological companies have launched products and services [13], such as Google's AutoML [5], allowing non-experts to build customized ML models [6].

[5] https://cloud.google.com/automl

4 Human Factors

4.1 Trust

A key factor for relations and collaboration between humans is trust, as stated by Corrigan [7, p. 1287]: "[...] trust grows organically only if there is trustworthiness present, and trustworthiness requires transparency, frequent and honest communication, mutual respect, and inclusivity." Usually a human's trust grows with time and transparent communication based on the ability to holistically perceive and interpret social and affective behavior (i.e. facial expressions and gestures) [1]. As for machines, this means it would need to be able to imitate this behavior in order to exhibit trust. On the other hand, it is yet unclear how the exhibition of trust for machine agents would affect the human attitude towards the machine.

4.2 Domain Knowledge

One of the outstanding factors that humans bring to the collaboration with AI is their domain knowledge and ability to intuitively decide based on a holistic view [11, 13]. Human experiences are greatly dimensional while the AI may act more variable [6]. Even in the collaboration between humans, the knowledge of subject matter experts is seen as indispensable [13]. It is not yet clear, how machines can understand and adapt human intuition. On the other hand, knowledge graphs are an option to represent domain knowledge [15], but the question remains, whether this knowledge compares to a lifetime of working experience. In case of health care, even if a patient's documentation can be fully automated, would we want the machine to be able to prescribe any medication without a doctor overseeing and/or adjusting the process?

4.3 Social Norms and Beliefs

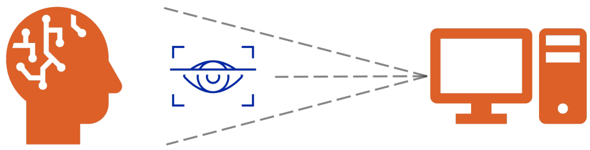
As stated by Akata et al.[1, p. 21]: "People can quickly and efficiently interpret social situations along various parameters (for example, mutual dependence, power, and conflict), and this can shape their willingness to cooperate." While humans might have the intuitive awareness of social biases and cultural contexts, the machine needs to be designed with regard to such [2]. It requires ethical considerations about if the machine should anticipate the same behavior and if so, how machines should exhibit social norms and beliefs. AI is just about to learn those skills as to incorporate social norms [1] and one has to be careful, as this bears the danger of AI exploiting and manipulating the human's beliefs or reinforcing undesirable social norms and biases.

4.4 Morality

Morality and the decision of what is right (moral) and wrong (immoral) should be familiar to any human, although, there is not always a clear or conscious answer. If it is a challenge for humans, how should machines exhibit ethics? Blasi [4, p. 205] elaborates on the relation of moral cognition and moral action and emphasizes the distinction of both: "While it may be possible to imagine a logic that exclusively applies to the domain of thinking, morality requires by definition the investment of knowledge in action." The difference of knowing and using knowledge with regard to ethics largely distinguishes humans and machines. Machines might be able to develop the logics, but not necessarily how these should be applied appropriately. Furthermore, Blasi [4] uses the self to explain moral cognition and action and it is questionable whether machines can achieve this notion of a self or if this is even desirable.

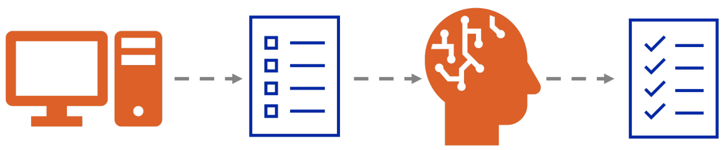
5 Collaborative Hybrid AI

5.1 Multimodal Interaction



Trust seems to be a key factor for collaborative Hybrid AI but also one of the most challenging factors [7]. Multimodal interaction, which refers to the reliable interpretation of signals that the human reveals to the machine [1], might enable trustworthiness. Progress in computations of social signals resulted in the machine's possibilitiy to perceive and interpret the social and affective behavior, model and synthesize social cues, constructs and emotions [1] and act as a trustworthy team partner. While multimodal interaction enables trustworthiness, the challenge lies in accurately perceiving and interpreting social and affective behaviour as well as modeling and synthesizing desirable social cues, constructs and emotions.

5.2 Cognitive Extender

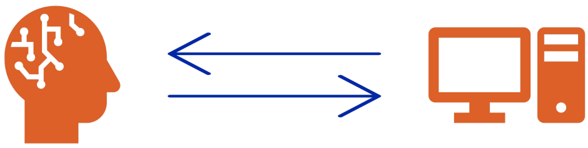


As an essential part in Hybrid AI collaboration, Chander et al. [6] and Wang et al. [13], suggest that not only should the machine agent be able to guide the human but also vice versa. Wang et al. [13] furthermore emphasize the importance of including human domain experts in the collaboration with AI, as to make sure that the machine produces correct output of high quality. Overall, concerning the humans' unique attribute of domain knowledge, AI should rather be seen as a cognitive extender, enhancing human processes such as memory processes, visual and auditory processing, communication, navigation, conceptualisation and more [10]. Automation and the capabilities in data processing allows AI to support the human agent [13], although the amount of automation needs to be appropriate. Data science workflows still require human-in-the-loop as they bring problem specific domain knowledge to understand the data [13]. Furthermore, research in accelerating human-in-the-loop is done to shorten the cycle between input and output, allowing the human more immediate feedback and ability to change parameters [16]. The notion of the machine as a cognitive extender gives collaborative Hybrid AI the advantage of enhancing human processes and augmentation instead of only automation. Nonetheless, it creates the challenges of allocating the work efficiently and appropriately while at the same time benefiting as much as possible from domain knowledge and automation.

5.3 Theory of Mind

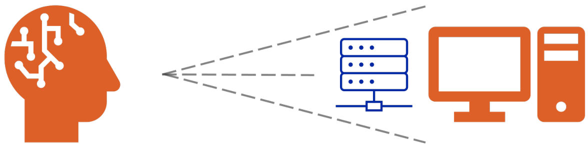
Akata et al. [1] call the process of attributing beliefs, goals and other mental attitudes on others the Theory of Mind (ToM). Making use of the ToM is an attitude that is of great advantage for humans and simulations with computer agents have shown that machines as well benefit from having a ToM [14]. In order to ensure an effective collaboration, the machine agent's ToM may even exceed the human's ToM [14] so that Hybrid AI can produce output for the greater good [5]. Some preliminary work in developing social agents has been initiated based on the computational abilities of AI [1]. As part of their guideline, Amershi et al. [2] propose that the system should learn from the users' behaviour, as well as mitigate the social biases while matching the relevant social norms. While this can become an advantage for collaborative Hybrid AI systems (e.g. in the context of a company wanting to ensure the right norms are matched [6]), there is also the danger of reinforcing undesired social norms and biases. Furthermore, it poses the question of how much deception and manipulation should be allowed [5].

5.4 Interactivity



Concerning ethics, the importance of distinguishing between moral cognition and action [4] requires the interactivity of humans and machines. Even if AI can develop ethical logics, ethical decision making should be based on the interactive collaboration of Hybrid AI [1]. Research is pushing the agenda of interactive AI, e.g. in data science where customizability can improve exploration of input and output [3], leading to power in the hands of the human [12]. While interactivity offers improved exploration of input and output, it is challenged by ethical dilemmas and the amount of customizability.

5.5 Transparency



As complex computation reduces workload for humans, transparency is needed to promote trust in successful collaboration [1, 7]. The "black box" model of AI computation hides decisions made, which users do not appreciate [3]. While data scientists have a broader understanding of technology, they may exhibit trust easily compared to stakeholders or novice users who lack the knowledge about machines and AI [13]. Minimizing the black box may enable trustworthiness but at the same time competes with the challenge to make complexity understandable and to serve different needs of transparency, which may vary for the individual users and stakeholders. Therefore, it should also be carefully evaluated whether collaborative Hybrid AI is appropriate in certain use cases where users lack the knowledge about AI or its application.

6 Discussion and Future Research

In the combination of machine and human factors, efficient collaboration depends on several inter-plays and trade-offs. Some of them we presented and summarize in an influence graph for Hybrid AI collaboration (Figure 2). The graph could serve as a basis to discuss the different influences, advantages and challenges of collaborative Hybrid AI and inform the design and development of such a system.

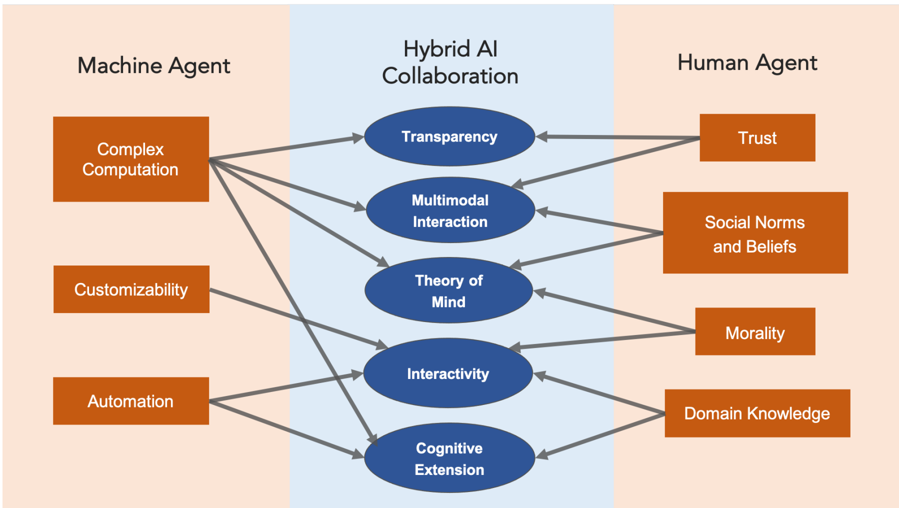


Fig. 2: Model of factors of the machine and human agent that influence Hybrid AI collaboration

Although not holistic, the graph illustrates the high number of influences of the collaboration and the value of reciprocal interaction and augmentation that Wang et al. [13] emphasize. The amount of transparency, one of the characteristics for collaborative Hybrid AI, can be adjusted with regard to complex computation and trust. Multimodal interaction is enabled by complex computation and influenced by the trust, social norms and beliefs of the human agent. The social norms and beliefs furthermore affect the Theory of Mind, which is also enabled by the complex computation of the machine agent and influenced by human morality. Interactivity and the notion of a cognitive extension are both influenced by automation and the domain knowledge of the human. While the morality of the human agent can be best incorporated through interactivity, customizability may govern interactivity. Complex computation furthermore allows for the role of Hybrid AI as a cognitive extension, both agents complementing each other symbiotically.

6.1 Future Research

Guidelines (e.g. [2]) address some of these factors and characteristics and inform the design of Hybrid AI collaborations. Still, the implementation and testing of real world Hybrid AI applications is missing. Future research may furthermore focus on unanswered questions like "How can the machine and human factors be incorporated and balanced in order to achieve efficient and effective Hybrid AI collaboration?" or "What kind of attributes should the ToM of the machine have?".

7 Related Works

Although, the research field of Hybrid AI is relatively new, there is some noteworthy literature to be mentioned. Wang et al. [13] conducted a study with data science practitioners, introducing them AutoAI. Results show mixed feelings of doubts, fear and even hope. The researchers point out that AI can take up different roles, shifting the work of data science practitioners to guiding AI in desired output creation. Multiple works [1, 3, 6, 9, 11, 12] study the complementary skill-set of human and AI and how their combination can overcome some of their individual shortcomings. Although terminology for this collaboration differs, some calling it interactive ML/AI [3, 12] or Transparent AI [6], with others favouring the term Hybrid Intelligence (HI) [1, 9]. The works of Chander et al. [6] and Jarrahi et al. [11] studied the collaboration in the organization context, putting focus on the transparency, interactivity and (non-expert) user, as the final decision maker. Others highlight the importance of creating human-centered AI systems [3, 12], giving them the control of the overall system to give a sense of safety and self-determination. In ML, accelerating human-in-the-loop is of interest as the time between pipeline input and output is often not desirable [16]. With appropriate optimization, the feedback cycle could be shortened and give the human the possibility for more immediate input and output update [16]. Akata et al. [1] and Dellermann et al. [9] perceive HI as the next step in technological progress and propose an agenda for further research. State-of-the-art reports in collaborative HI, adaptive Hi, responsible HI and explainable HI show the immense potential but also the challenges that need to be overcome.

With AI also being a cognitive extender of humans, they bring new opportunities, but also ethical and technical challenges [10]. Researchers saw the need to study AI extenders, placing them on a cognitive spectrum and investigate the ethical consequences and implications [10]. Another work [14] studied the performance of different level ToM of agents in simulations, finding that second-order ToM agents may outperform agents with limited capabilities. Chakraborti and Kambhampati [5] explore how AI can use manipulation and deception to achieve greater good. Not all actions or outcomes are desired, thus they studied the moral and ethical dilemma which is posed on the design of autonomy. Amershi et al. [2] propose a design framework for human-AI interactions. They gathered guides from industrial and academic recommendations, consolidating them into 18 general guidelines.

Other related works focus only on the human agents, proposing contemplative practices, i.e. Yoga Nidra, to improve their capabilities in problem solving [7], or study the connection between moral cognition and moral action and what the consequences are [4].

8 Lessons Learned

8.1 Human-in-the-loop vs. machine-in-the-loop

For collaborative Hybrid AI, both agents - machines and humans - bring different capabilities to the collaboration. As we have seen in the definition of Hybrid AI (Chapter 2.2), the proportion of involvement can vary for each application. Therefore, more discussions about the amount of participation of each agent with regard to the use case should be brought up. With the reciprocity of collaborative Hybrid AI, the allocation of work and responsibility between the machine and the human should be carefully assessed. Furthermore, if both agents have the ability to guide each other, new challenges arise if the work gets more and more inseparable.

8.2 Remaining challenges of collaborative Hybrid AI

As explored in Chapter 5, the human and machine factors introduce several trade-offs to collaborative Hybrid AI. The inter-plays between both agents' capabilities are use case specific and need to be evaluated. Design guidelines like the one from Amershi et al. [2] try to ensure the appropriate design and development of Hybrid AI implementations addressing several characteristics of the collaboration. Nonetheless, during our research we discovered that there are many factors that may influence the interaction between humans and AI, beyond the factors we included in our graph (Chap. 5, Fig. 2). For example, for data science there are various stakeholders that need to be taken into account [13]. Overall, guidelines may serve as a starting point, but there is more research needed on the different factors and their influence in real world applications.

8.3 Giving answers to questions of a new research field

As the research field of Hybrid AI is just on the rise, there is not yet a satisfying amount of literature. Therefore, in order to search for solutions to problems and challenges that Hybrid AI poses, one should take into account (older) literature on similar research, e.g. the paper of De Weerd et al. on the Theory of Mind [14]. As Hybrid AI is influenced by a diversity of research fields, the inclusion of different disciplines is required. For example, Psychology can be taken into account for human factors, Computer Science for machine factors and Human-Computer Interaction (HCI) for the collaboration itself.

References

[1] Zeynep Akata et al. "A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence". In: *Computer* 53.8 (2020), pp. 18–28.

[2] Saleema Amershi et al. "Guidelines for human-AI interaction". In: *Proceedings of the 2019 chi conference on human factors in computing systems*. 2019, pp. 1–13.

[3] Saleema Amershi et al. "Power to the people: The role of humans in interactive machine learning". In: *Ai Magazine* 35.4 (2014), pp. 105–120.

[4] Augusto Blasi. "Moral cognition and moral action: A theoretical perspective". In: *Developmental review* 3.2 (1983), pp. 178–210.

[5] Tathagata Chakraborti and Subbarao Kambhampati. "(When) can AI bots lie?" In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 2019, pp. 53–59.

[6] Ajay Chander et al. "Working with Beliefs: AI Transparency in the Enterprise." In: *IUI Workshops*. 2018.

[7] James M Corrigan. "Augmented intelligence—the new AI—unleashing human capabilities in knowledge work". In: *2012 34th International Conference on Software Engineering (ICSE)*. IEEE. 2012, pp. 1285–1288.

[8] Claire Corthell. *Hybrid Intelligence: How Artificial Assistants Work*. `https://medium.com/@clarecorthell/hybrid-artificial-intelligence-how-artificial-assistants-work-eefbafbd5334`. Accessed: 2020-11-19. 2016.

[9] Dominik Dellermann et al. "Hybrid intelligence". In: *Business & Information Systems Engineering* 61.5 (2019), pp. 637–643.

[10] José Hernández-Orallo and Karina Vold. "Ai extenders: The ethical and societal implications of humans cognitively extended by Ai". In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 2019, pp. 507–513.

[11] Mohammad Hossein Jarrahi. "Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making". In: *Business Horizons* 61.4 (2018), pp. 577–586.

[12] Albrecht Schmidt. "Interactive Human Centered Artificial Intelligence: A Definition and Research Challenges". In: *Proceedings of the International Conference on Advanced Visual Interfaces*. 2020, pp. 1–4.

[13] Dakuo Wang et al. "Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–24.

[14] Harmen de Weerd, Rineke Verbrugge, and Bart Verheij. "How much does it help to know what she knows you know? An agent-based simulation study". In: *Artificial Intelligence* 199-200 (2013), pp. 67–92.

[15] Ledell Wu et al. "Starspace: Embed all the things!" In: *arXiv preprint arXiv:1709.03856* (2017).

[16] Doris Xin et al. "Accelerating human-in-the-loop machine learning: Challenges and opportunities". In: *Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning*. 2018, pp. 1–4.