Human Factors and Simulation, Vol. 83, 2023, 201–211
https://doi.org/10.54941/ahfe1003578

Can Machine Learning be a Good Teammate?

Leslie M. Blaha and Megan B. Morris

Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433, USA

ABSTRACT

We hypothesize that successful human-machine learning teaming requires machine learning to be a good teammate. However, little is understood about what the important design factors are for creating technology that people perceive to be good teammates. In a recent survey study, data from over 1,100 users of commercially available smart technology rated characteristics of teammates. Results indicate that across several categories of technology, a good teammate must (1) be reliable, competent and communicative, (2) build human-like relationships with the user, (3) perform their own tasks, pick up the slack, and help when someone is overloaded, (4) learn to aid and support a user's cognitive abilities, (5) offer polite explanations and be transparent in their behaviors, (6) have common, helpful goals, and (7) act in a predictable manner. Interestingly, but not surprisingly, the degree of importance given to these various characteristics varies by several individual differences in the participants, including their agreeableness, propensity to trust technology, and tendency to be an early technology adopter. In this paper, we explore the implications of these good teammate characteristics and individual differences in the design of machine learning algorithms and their user interfaces. Machine learners, particularly if coupled with interactive learning or adaptive interface design, may be able to tailor themselves or their interactions to align with what individual users perceive to be important characteristics. This has the potential to promote more reliance and common ground. While this sounds promising, it may also risk overreliance or misunderstanding between a system's actual capabilities and the user's perceived capabilities. We begin to lay out the possible design space considerations for building good machine learning teammates.

Keywords: Human-machine teaming, Teammate-likeness, Good teammates, Machine learning, Explainable artificial intelligence, Human-autonomy teaming, Individual differences

INTRODUCTION

Artificial Intelligence (AI) has rapidly advanced in recent decades (Jordan and Mitchell, 2015), permeating our daily lives through work (e.g., smart assistants), leisure (e.g., recommending content on streaming devices), social activities (e.g., taking and tagging selfies with a group of friends), routine tasks (e.g., navigating to a destination), healthcare (e.g., diagnosing issues), and important societal processes (e.g., offender management; Mehrabi et al., 2021), among others. Machine Learning (ML) is one of the most common types of AI that uses existing data, typically large datasets, to teach itself by gaining experience in specific input-output behavior (Jordan and Mitchell, 2015; Silver et al, 2018). As ML continues to become more intertwined in humans' daily lives, human-machine teaming (HMT) becomes a critical factor for the success of these operations. But what makes ML a good teammate?

Currently, little is understood about the important design factors that create technology that people perceive to be good teammates. Several lines of research have proposed design challenges toward establishing effective human-autonomy teams (e.g., Klein et al., 2004) and teammate-likeness models (e.g., Wynne and Lyons, 2018). Recently, Blaha and colleagues asked which of these elements people thought were important, not to be effective teams, but to be good teammates. Blaha et al. (2023) collected ratings from over 1,100 users of commercially available smart technology on the importance of 116 teammate characteristics, identified from the existing HMT literature, for smart technology with which they were familiar to be a good teammate. In an exploratory factor analysis, they identified seven factors. *Reliable, Competent, and Communicative* characteristics were rated the most important for being a good teammate (*Mdn* = 5.90 on a Likert scale from 1 to 7). This was followed by teammates having *Common Goals*, providing *Cognitive Aiding*, behaving with *Predictability*, balancing workloads by *Divide & Conquer*, and exhibiting *Transparency*. *Human-like Relationship Building* was rated the least important, though at the median rating of 4.47 it was still well above the minimum rating of 1 (anchored at "Not at all important"). Notably, individuals who gave higher importance ratings to the *Human-like Relationship Building* factor also tended to rate smart technology more as a teammate than as a tool. Additionally, individual differences in traits such as tendency to be an early technology adopter, propensity to anthropomorphize, personality, propensity to trust smart technology, and schemas and attitudes toward technology were associated with factor importance ratings as well as perceptions of whether smart technologies are teammates or tools (Morris et al., under review). These individual differences may have important implications for ML system design.

In the remainder of this paper, we explore the implications of these good teammate characteristics and associated individual differences in the design of ML algorithms and their user interfaces. Note that when we use the term ML, we are not referring to a single, specific ML algorithm; rather we consider the space of technologies that leverage computational algorithms and optimization to train representations that map given inputs to a desired type of output, inclusive of hierarchical clustering, support vector machines, decision trees, neural networks, expert logic systems, and more. We use "ML" to refer to algorithms themselves and "ML systems" to refer to the combinations of algorithms and user interfaces that bring ML into HMT settings.

RELIABLE, COMPETENT, AND COMMUNICATIVE

In the aforementioned study, the factor *Reliable, Competent, and Communicative* was perceived as the most important for smart technology to be a good teammate. Characteristic descriptors loading strongly on this factor included "My teammate understands the tasks", "My teammate responds the same way under the same conditions at different times", "My teammate performs their tasks reliably", and "My teammate communicates in a way that is familiar to me". Blaha and colleagues interpreted this factor to indicate that a good teammate understands the team and teammates' tasks, executes their tasks reliably, consistently, and accurately, and communicates bidirectionally with their teammates to promote mutual understanding between them.

ML algorithms can generally be trained to be highly accurate. Accuracy in algorithms like neural network classifiers and decision trees is a function of data volume and quality, distinguishability of classes/decision paths, and the methods for training. Once trained, the level of accuracy is known and quantifiable (e.g., F_1 score on test data). This can support being a good teammate as human teammate expectations can be calibrated to the performance parameters.

Adversarial ML examples (Szegedy et al., 2014), however, and other tests of generalizability repeatedly demonstrate the fragility of ML. A trained algorithm working on data well aligned to its training set will perform reliably (i.e., "My teammate responds the same way under the same conditions at different times"). Given an out-of-sample data point, there is a good chance the ML will not correctly classify/decide the output a human would intend for the input. Adversarial noise can cause errors that may or may not also be made by a human (Elsayed et al., 2018). Consequently, ML capabilities today do not meet good teammate criteria for "My teammate never does anything unexpected" or "can perform under a variety of circumstances". While advances in adversarial training or one-shot learning are helping to make ML more adaptable to new data, ML is not yet able to recover from errors, be corrected by the human, or be "adaptive when they try a new task" without additional training.

ML algorithms are not intrinsically communicators. Therefore, on its own, ML cannot achieve a number of highly important good teammate characteristics such as "My teammate is able to talk to me in plain language", "My teammate communicates in a way that is familiar to me", or "I can direct my teammate to important signals or information". Even as we write this, the nascent ChatGPT is making headlines for its advances in communicative capabilities. Leveraging a combination of transformer ML architecture, training on massive data sets (GPT3 has 175 billion tunable parameters trained on 300 billion language tokens; Brown et al., 2020), and clever, adaptive interface designs, ChatGPT demonstrates the potential for ML systems to possess characteristics that will promote good teammate perceptions. More work will need to be done, however, to demonstrate that ML has any true depth of understanding in line with characteristics like "My teammate has an understanding of my goals" (i.e., mutual understanding or theory of mind) or "My teammate pushes information at the right time when needed" (i.e., proactive communication).

Related to this factor is the idea that ML should be free from bias and not discriminate. This is reflected in the descriptor "My teammate should not act against (or undermine) the team's best interests," which is a characteristic associated with concepts of benevolence, or that a teammate should be generally oriented toward being helpful (see also Wynne and Lyons, 2018).

Although ML is used in many low-stakes commercial applications, such as music recommendation, it is increasingly used in high-stakes applications, such as hiring, which have critical impacts on individuals' lives. There are two main sources of bias, the data, such as measurement bias, representation bias, sampling bias, etc., and the algorithm, such as popularity bias, emergent bias, evaluation bias, etc. (Mehrabi et al., 2021; see also Friedman and Nissenbaum, 1996). Bias in either of these sources can perpetuate bias in the data-algorithm-users loop. Given the pervasiveness of bias in human-machine interaction, this is a challenging area to address. To be a good teammate, ML bias must be addressed in the algorithm development stage and mitigated throughout system design.

ADAPTIVE COGNITIVE AIDING

ML will need to aid and improve the user's abilities to be a good teammate. While abilities include both cognitive and physical, only factors associated with *Adaptive Cognitive Aiding* were rated as important for smart devices to be teammates. The six descriptors loading onto this factor were that my teammate "improves my thinking", "improves my abilities", "can learn from my behavior", "can correct my behavior if I run into difficulty", "adjusts their behavior based on my behavior", and notices "when something important happens".

ML has traditionally been developed in isolation of direct user input (supervision is through the data curation and design process), but this is not optimal in HMT where strengths of both ML and human expertise can be complementarily exploited, such as leveraging ML to triage large volumes of cybersecurity data to provide targeted information to human decision makers (Franklin et al., 2017). Specifically, these systems must have the objective of optimizing team performance instead of solely its own performance (Wilder et al., 2020) and be designed to adapt to direct end-user inputs.

Algorithms supporting interactive machine learning (IML) are strong candidates for being good teammates because they learn incrementally through direct feedback with a user (Amershi et al., 2014). IML systems leverage user's direct interaction with ML outputs to adjust their decisions to match the user inputs, usually made through some sort of visual analytics interface (Arendt, Grace, and Volkova, 2018; Boukhelifa et al., 2020) or preference voting options (e.g., thumbs up/down in recommender systems)[1].

PREDICTABLE

Not only does ML need to be reliable to be a good teammate, but it also must be *Predictable*. This factor had two strongly loaded descriptors: "It is clear what my teammate intends to do before they take action" and "I can predict what my teammate does". The first statement reflects a desire for a teammate to signal their intentions, which could be done either through dynamic information sharing (status update messages/cues) or through interface design (the back and forth structure of chat boxes). ML has the potential to

[1] Formally known as active preference learning, see Eric et al. (2007).

be predictable; after training, many algorithms will produce the same output for a given input. Re-training may change the input-output mappings, but at any point in time, ML that is not actively training should be predictable. ML systems can be designed to support signals conveying changes in the ML over training iterations.

The second *Predictable* descriptor reflects a desire for an understanding of the system's processes, which can evolve in a person through experience or explicit training with the system that is *Reliable, Competent, and Communicative*. This may be challenging to develop in non-experts in ML technologies, because they do not tend to have mental models of ML algorithms; indeed, Bos et al. (2019) demonstrated that after a session of interacting with the high-performing image classifier in the lab, people were only able to predict its behavior 73% of the time, on average. The challenge with low predictability is that people are left guessing when the ML will fail or succeed, putting overall team performance at risk, as well as risking human rejection of the algorithm altogether.

DIVIDE & CONQUER

A basic reason humans operate in teams is to accomplish more complex goals than a single person can accomplish alone. We seek to develop machine teammates to operate in much the same way: support execution of complex tasks that would benefit from skills that augment or supplement humans (Seeber et al., 2020). A hallmark of autonomous agents, especially the increasingly intelligent systems enabled by AI/ML that the community envisions for human-agent teams, is "agentic capability" (Wynne and Lyons, 2018) or some degree of autonomous (without human oversight) decision making ability and authority (Chen and Barnes, 2014) that can be implemented in different degrees (Parasuraman, Sheridan, and Wickens, 2000) in support of executing complex missions (Barnes, Chen, and Jentsch, 2015). The *Divide & Conquer* factor captured a number of descriptors reflecting that good teammates are agents who can execute tasks on their own, including descriptors such as "My teammate can take initiative to start or finish tasks", "can work on separate parts of a task than me", and divide work to "play to our individual strengths." Additionally, a good teammate "can share my workload" as needed and "can pick up the slack when I am overloaded".

ML is poised to be a good teammate that can play to its unique strengths in support of the human teammates executing complex missions, particularly the machine's ability to tackle the V's of big data: variety, velocity, volume, volatility (Dasgupta et al., 2018). Machines can store and process more data than humans can. ML, particularly modern deep neural networks, are capable of finding and exploiting patterns in big data that humans have not found. ML can do this on streams of data at a speed and scale that humans cannot. Humans, on the other hand, have unique reasoning and sensemaking skills that machines cannot yet match. If systems are able to integrate the strengths of human reasoning with the strengths of ML data processing, the human-ML team will be able to do things with data analytics that neither can do alone (Baber et al., 2018).

It is also possible to automate many repetitive, predictable, or low level tasks in ways that machines could take over task execution from human operators. Spam filters on email, for example, use ML-based pattern recognition to flag and sort email for humans. However, more research is needed to develop ML with the range of cognitive skills and flexibilities that would enable broad interoperability, so ML-enabled agents can take over from humans on many tasks, particularly tasks in unknown or uncertain conditions. They need the ability to execute the same reasoning, inference, and decision making to be interoperable, as well as the ability to infer the goals, actions, and current state of the human teammate. Advances in integrations of computational cognitive models with ML shows promise toward developing ML that can leverage and learn from formal representations of human intelligence (e.g., Trafton et al., 2020).

TRANSPARENCY

Related to the need for a good ML teammate to be communicative and predictable, is the idea that ML should be transparent. Indeed, the *Transparency* factor captured characteristics like "My teammate provides their reasoning to support recommendations they make", "My teammate can provide an explanation for their behavior", and "My teammate is open about their decisions and actions".

ML is often characterized as a "black box", partially because the logic of many algorithms is unknown to non-experts and partially because, even to expert designers, the training process may leverage parameter numbers and statistical patterns that are beyond human reasoning to shape the ML's internal representation. The latter is particularly true for techniques like deep reinforcement learning and deep neural networks, probabilistic graphical models, or support vector machines, which are all techniques that construct their own internal representation of the data against which the algorithm makes its decision assignments. As a result, the subfield of Explainable AI (XAI) has emerged, pushing the development of methods that take into account the target user's needs and seeks to provide transparency about the salient data and reasoning processes underlying the ML's decisions, recommendations, or actions (Arrieta et al., 2020; Biran and Cotton, 2017; Gunning et al., 2019). Such XAI capabilities directly address the descriptors of good teammate transparency, and likely support the communication characteristics of the *Reliable, Competent, and Communicative* factor. But given the scale and complexity of some models (Brown et al, 2020), it will be challenging for many ML algorithms to achieve full *Transparency.*

COMMON GOALS

Two descriptors strongly loaded onto the *Common Goals* factor: "My teammate is motivated to help me" and "My teammate and I have common goals". A third descriptor on this factor, dually loaded onto the *Reliable, Competent, and Communicative* factor, is: "My teammate is clear about their goals". Thus a good teammate has common goals, supportive of the human teammate, and communicates about them. It is hard to argue that ML has any goals other than those of the designer or end-user. Today, even though there are algorithms being used to design or optimize algorithms, they do not have independent goals.

RELATIONSHIP BUILDING

The *Human-like Relationship Building* factor accounted for the second largest proportion of variance in Blaha and colleague's factor analysis but had the lowest overall importance rating (Blaha et al., 2023). Thus, humans have consistent views that relationship building is characteristic of good teammates but it is not as important as those characteristics emphasizing performance and communication. A range of descriptors fell into this common factor, such as explicit statements about relationship building and emotions (e.g., "My teammate makes me feel appreciated"), shared success ("My success is dependent on the my teammate's success"), mutual understanding ("My teammate helps me understand how they behave", "My teammate can understand my non-verbal cues"), and anthropomorphism ("My teammate looks and feels like a human"). Empirical evidence is mixed about whether anthropomorphism positively influences team performance. Many examples can be found where people prefer agents or robots that are personable and expressive (e.g., Hamacher et al., 2016) while other studies point to human-like embodiment of AI as having no impact on team performance (e.g., Haring et al., 2021).

There are few if any *Human-like Relationship Building* characteristics that ML can achieve alone; most characteristics will be a product of the ML systems designed to facilitate human interaction, such as visual analytic interfaces or embodied robots. The one exception may be the concept of mutual failure ("If my teammate fails, the team will fail"), as ML that performs poorly or is adversarially attacked is likely to cause cascading degradations in the team.

INDIVIDUAL DIFFERENCES CONSIDERATIONS

Morris et al. (2023) delved into several individual differences associated with perceptions of factors' importance for being a good teammate and the perception of whether smart technology is a teammate or a tool. Generally, individuals' propensity to trust in smart technology, early technology adoption, agreeableness, and propensity to anthropomorphize were positively associated with importance ratings and teammate ratings. Other individual differences such as schema about automation behavior and attitudes toward technology had significant positive and negative relationships with some factors' importance ratings and teammate ratings (Morris et al., 2023). This suggests it may be worthwhile to design ML systems to adapt to users' individual differences to enhance teaming and performance. For example, individual differences data on one's propensity to anthropomorphize could be fed to an ML system via questionnaires to adapt ML performance to the user. A ML visual analytics interface could adapt to be more personable or expressive in how it communicates information to that user. Given that schemas and attitudes toward smart technology affect views of whether the technology is a teammate, this has implications for potential training with ML systems to enhance teaming. For example, although perfect automation schema is positively associated with teammate perceptions, users should be educated that ML systems are not infallible, much like a human teammate, to deter overconfidence in the system. Similar educational efforts could target negative attitudes toward technology based on fears of AI threats as capabilities advance.

PITFALLS

Do we want ML algorithms or systems to possess all the characteristics of a good teammate? Researchers have demonstrated that humans can form teams with machines since at least Nass, Fogg, and Moon's (1995) experiment illustrating that people who think their performance is interdependent with that of a computer show increases in several social psychology measures of team membership. These included increased behavioral conformity as well as increased perceptions of being a team, of self-similarity to the computer, of the quality of computer-provided information, of their own level of cooperativeness, and their own openness to influence. Depending on the complexity of the missions and the consequences of error, these may not always be desirable.

If we design systems to high levels of desirable characteristics, we run the risk of creating gaps between perceived and actual ML capabilities. Such gaps can result in humans misunderstanding the machine's intelligence resulting either in overconfidence in the system, overreliance, compliance errors, or in under-reliance on the system and disuse errors (Parasuraman and Riley, 1997). High levels of anthropomorphism can give a false impression of systems possessing human-level reasoning or skills. Additionally, these capabilities might increase negative attitudes toward ML systems as users might see them as a threat due to increased human-likeness.

We may also desire ML teammates that can offer checks and balances to human teammates; human teammates can move between leader/follower roles, question each other's goals, actions, and decisions, and suggest alternative courses of action to each other. High performing teams regularly review lessons learned and seek ways to improve as a team (Company et al., 2007). It may be possible to build ML systems where the interfaces adapt and become too tailored to a user, such that they do not offer different/conflicting suggestions or alternative inferences. This could produce gaps in human and ML knowledge or capabilities, promote biased echo chambers, or perpetuate poor decisions.

ML can continue to learn and adapt overtime (usually through retraining, but also reweighting, one-shot learning, and other emerging techniques). Anecdotally, in free responses, participants in the Blaha et al. (2023) study mentioned that machines also needed to be secure and protect privacy to be good teammates. If ML systems are gathering information about their human teammates to be reactive, responsive, adaptive, communicative, etc., attention will need to be paid to how that data is protected if integrated into re-learning processes, particularly if that re-learning requires integration or reaching out to a larger ML knowledge base. This will also be a consideration for even a local system that simultaneously supports multiple humans on a single team. A good ML teammate should probably not tell all your secrets.

CONCLUSION

At present, ML only has some intrinsic characteristics that support ML being a good teammate; other characteristics can be derived from capabilities of ML systems. As we have discussed, current ML technologies are demonstrating many elements of each factor reviewed. There are promising ways that ML research can achieve capabilities supporting all seven of the good teammate characteristics reviewed herein. With good design and careful consideration of pitfalls, an ML system can be a good teammate.

ACKNOWLEDGMENT

The authors would like to thank C. Fallon, B. M. Hartzler, R. Wong, J. King, and T. Curley for their insights and discussions about the characteristics of good teammates. This research was funded by the Department of Defense. The views expressed are those of the authors and do not reflect the official guidance or position of the United States Government, the Department of Defense, or of the United States Air Force. Distribution A: Approved for public release; distribution unlimited. AFRL-2023-0833.

REFERENCES

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4), pp. 105–120.

Arendt, D., Grace, E. and Volkova, S. (2018). Interactive machine learning at scale with CHISSL. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), pp. 8194–8195.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, pp. 82–115.

Baber, C., Cook, K., Attfield, S., Blaha, L. M., Endert, A., and Franklin, L. (2018). A conceptual model for mixed-initiative sensemaking. *Proceedings of ACM SIGCHI Sensemaking Workshop*.

Barnes, M. J., Chen, J. Y., and Jentsch, F. (2015). Designing for mixed-initiative interactions between human and autonomous systems in complex environments. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1386–1390.

Biran, O. and Cotton, C. (2017). Explanation and justification in machine learning: A survey. *IJCAI-17 Workshop on Explainable AI (XAI)*, 8(1), pp. 8–13.

Blaha, L. M., Fallon, C., Morris, M. B., Hartzler, B. M., and Wong, R. (2023, under review). Characteristics of a good teammate. Retrieved from psyarxiv.com/gn93z.

Bos, N., Glasgow, K., Gersh, J., Harbison, I. and Paul, C. L. (2019). Mental models of AI-based systems: User predictions and explanations of image classification results. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1), pp. 184–188.

Boukhelifa, N., Bezerianos, A., Chang, R., Collins, C., Drucker, S., Endert, A., Hullman, J., North, C. and Sedlmair, M. (2020). Challenges in evaluating interactive visual machine learning systems. *IEEE Computer Graphics and Applications*, 40(6), pp. 88–96.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, pp. 1877–1901.

Chen, J. Y., and Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, 44(1), pp. 13–29.

Company, N., Dubinsky, R., Druskat, V. U., Mangino, M. and Flynn, E. (2007). What makes good teams work better: Research-based strategies that distinguish top-performing cross-functional drug development teams. *Organization Development Journal*, 25(2), pp. P179–P186.

Dasgupta, A., Arendt, D. L., Franklin, L. R., Wong, P. C., and Cook, K. A. (2018). Human factors in streaming data analysis: Challenges and opportunities for information visualization, *Computer Graphics Forum*, 37(1), pp. 254–272.

Elsayed, G., Shankar, S., Cheung, B., Papernot, N., Kurakin, A., Goodfellow, I. and Sohl-Dickstein, J. (2018). Adversarial examples that fool both computer vision and time-limited humans. *Advances in Neural Information Processing Systems,* 31.

Eric, B., Freitas, N. and Ghosh, A., (2007). Active preference learning with discrete choice data. *Advances in Neural Information Processing Systems*, 20.

Franklin, L., Pirrung, M., Blaha, L., Dowling, M. and Feng, M. (2017). Toward a visualization-supported workflow for cyber alert management using threat models and human-centered design. *2017 IEEE Symposium on Visualization for Cyber Security (VizSec)* (pp. 1–8).

Friedman, B. and Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3), pp. 330–347.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. and Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, 4(37), eaay7120.

Hamacher, A., Bianchi-Berthouze, N., Pipe, A. G., and Eder, K. (2016). Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction, *Proceedings of 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 493–500).

Haring, K. S., Satterfield, K. M., Tossell, C. C., De Visser, E. J., Lyons, J. R., Mancuso, V. F., Finomore, V. S., and Funke, G. J. (2021). Robot authority in human-robot teaming: Effects of human-likeness and physical embodiment on compliance. *Frontiers in Psychology*, 12, pp. 625713.

Jordan, M. I. and Michell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), pp. 255–260.

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R. and Feltovich, P. J. (2004). Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), pp. 91–95.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), pp. 1–35.

Morris, M. B., Hartzler, B., M., Blaha, L. M., Fallon, C., and Wong, R. (2023, under review). Individual differences in perceptions of technology as a good teammate.

Nass, C., Fogg, B. J., and Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6), pp. 669–678.

Parasuraman, R. and Riley, V., 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), pp. 230–253.

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans* 30(3), pp. 286–297.

Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G. J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T. and Lillicrap, T. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419), pp. 1140–1144.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing properties of neural networks. *Proceedings of the 2014 International Conference on Learning Representations*.

Trafton, J. G., Hiatt, L. M., Brumback, B., and McCurry, J. M. (2020). Using cognitive models to train big data models with small data. *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 1413–1421).

Wilder, B., Horvitz, E., and Kamar, E. (2020). Learning to complement humans. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*.

Wynne, K. T., and Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, 19(3), pp. 353–374.