Adaptive Autonomy as a Means for Implementing Shared Ethics in Human-AI Teams

Allyson I. Hauptman, Beau G. Schelble, Nathan J. McNeese
Clemson University
105 Sikes Hall
Clemson, SC 29634
ahauptm@clemson.edu, bschelb@clemson.edu, mcneese@clemson.edu

Abstract

Rapid increases in artificial intelligence technologies are resulting in closer, more symbiotic interactions between artificially intelligent agents and human beings in a variety of contexts. One of those contexts that has only recently been receiving much attention is human-AI teaming, where an AI agent operates an interdependent teammate, rather than a tool. These teams present unique challenges in creating and maintaining shared team understanding, specifically when it comes to a shared team ethical code. Because teams change in composition, goals, and environments it is imperative that AI teammates be capable of updating their ethical codes in concert with their human teammates. This paper proposes a two-part model in order to implement a dynamic ethical code for AI teammates in human-AI teams. The first part of the model proposes that the ethical code, in addition to its team role, be used to inform an adaptive AI-agent of when and how to adapt its level of autonomy. The second part of the model explains how that ethical code is consistently updated based upon the AI agent's iterative observations of team interactions. This model makes multiple contributions to the community of human-centered computing, because teams with higher levels of team cognition exhibit higher levels of performance and longevity. More importantly, it proposes a model for more ethical use of AI teammates on human-AI teams that is applicable to a variety of human-AI teaming contexts and permits room for future innovation.

Introduction

The rapid advancement of technology has led to increased close interaction between humans and artificially intelligent (AI) agents (O'Neill et al. 2020). These interactions take place everywhere from the workplace (Nyre-Yu, Gutzwiller, and Caldwell 2019) to virtual gaming platforms (Zhang et al. 2021). Yet, despite decades of advancement, these AI agents will never reach their true potential until they can operate as effective teammates. This is well supported by Johnson and Vera's research into the autonomy paradox and AI teaming, which explains that AI technologies are most successful and accepted when they are interwoven into human work practices as part of a team (Johnson and Vera 2019). A vital part of incorporating AI agents into these practices is that the agent is operating with enough autonomy to fulfill its team role independently. The right amount of autonomy the agent possesses should be dictated by the team's goals and environment (Suzanne Barber, Goel, and Martin 2000), which is particularly important to team ethics, because ethical codes are highly dependent on contextual factors (Bagdasarov et al. 2013). Since a team's goals and environment change, the agent's level of autonomy may need to change, which is why we chose to build a model based on an AI teammate's decision adapt its level of autonomy.

The ability for a team to come together and perform is directly linked to its ability to coalesce around an ethical ideology (Petersen, Pavlidis, and Semendeferi 2014). Studies show that a shared understanding of team ethics underpins organizational success (Haskins, Liedtka, and Rosenblum 1998). Yet, as teams get larger and more complex, it becomes increasingly difficult for teams to implement and maintain a shared ethical ideology (Petersen, Pavlidis, and Semendeferi 2014). AI teammates certainly add to the complexity of a team, and thus human-AI teams need to possess a foundational way to implement and maintain a shared ethical ideology.

Ethics are based in human reasoning, and humans are extremely sensitive to just how much of that reasoning needs to be involved in decisions and actions performed by artificial intelligence (Jarrahi 2018). Because of this, the level of autonomy (LOA) that an AI teammate is operating from is not only a question of performance, but also a question of ethics. Research shows that human beings tend to equate the ethicality of an AI agent's actions to the level of human input and oversight that should have been involved in its decision to act (Aitken et al. 2020). Hence, the agent's level of autonomy, which equates to the level of human control over its decisions and actions, has an important ethical component that this paper's model seeks to address.

The level of control an AI teammate should have is then partially reliant on the team's shared understanding of what is ethical in the current moment. In team situations, the ethical code by which the team operates is largely based upon the norms that develop through team interactions (Flathmann et al. 2021). This means that the ethical code is dynamic, and all teammates, including artificial ones, need to be capable of updating their shared understanding of the team's ethical code. In human-AI teams, it is even more important for AI agents to quickly incorporate changes to the team's ethical code, because human teammates are more likely to point out violations of team values and norms by an AI teammate than they would another human teammate (Bansal et al. 2019). The recognition of such violations is harmful to team dynamics and team performance, and so the more fluidly the agent can perceive and update the code, the better.

This paper presents a two-part model for implementing and maintaining a dynamic team ethical ideology for human-AI teams that leverages adaptive autonomy, a process by which the agent changes at what level of autonomy it is operating (Martin and Barber 1996). The first part of the model explains that the state logic behind adaptation that allows AI agents to continuously sense its environment and adapt to changing conditions could be used to consistently sense for changing team ethical interactions (Fereidunian et al. 2007). The second part of the model details how the agent's ethical code would be updated based upon repeated interactions that reflect true changing ethical norms.

The use of adaptive autonomy as a means for implementing shared ethics in human-AI teams is an important contribution to the field of human-agent teaming, as a team member's ability to predict his or her teammates actions is a vital component of team trust and performance (Demir et al. 2021). AI agents capable of updating their ethical code in this manner would result in more cohesive, higher performing human-AI teams. More pivotal, AI teammates capable of maintaining and acting upon a shared team ethical code are vital to the ethical use of AI agents in modern society. This paper will build upon a previously presented model for ethical human-AI teaming (Flathmann et al. 2021) and propose a model for ethical adaptive autonomous teammates in order to enable more ethical and successful human-AI teams.

Background

Human-AI Teaming

A team composed of humans and AI agents is fundamentally different than an all-human team. HATs are teams composed of at least one human and one AI agent that operate independently towards shared goals (O'Neill et al. 2020; McNeese et al. 2018). McNeese and colleagues found that in order for an AI agent to be considered a full-fledged teammate, it should possess at least partial agent autonomy, meaning that the agent is capable of executing a decision once that decision is approved (O'Neill et al. 2020). This level of autonomy enables the agent to fulfill a unique team role, without which it would merely be a tool used by a human filling that role; however, this heightened level of autonomy also contributes to a couple major challenges in building effective, cohesive HATs, including teammate predictability and common ground (Klien et al. 2004). Because AI agent decision logic is based upon a static set of pre-programmed rules that are independent from the environment, it is difficult for human teammates to predict how it will operate in new environments (Shilton 2018). This lack of insight into the AI agent's decisions hinders human teammate trust in the AI agent, and research shows that this is why humans are sensitive to the careful balance between AI autonomy and human control (Jarrahi 2018). Because team composition, goals, and environments change, the right balance is not static, which is why we identified the need to incorporate dynamic decision logic into AI teammates.

Adaptive Autonomy

Autonomous agents can make and execute decisions with varying levels of human input, which Parasurman, Sheridan, and Wickens categorized into levels of autonomy (LOA). These levels range from 0 (the human decides everything) to 10 (the computer decides everything) (Parasuraman, Sheridan, and Wickens 2000). Further research by Wickens and colleagues posited that AI agents really possess degrees of autonomy, because there is a defined level of not only what actions they can perform, but also to what extent they can perform them (Wickens et al. 2010). In other words, an agent may be able to throw a ball, but how far and at what target need to also be defined. This is an important concept to our study, because it hints at the idea that the rules guiding an agent's level of autonomy extend beyond just its capabilities, which is supported by Wickens and colleagues additional assertion that we may need AI agents who can change their degree of autonomy based upon the situation (Wickens et al. 2010).

Martin and Barber's research addressed this idea with the concept of adaptive autonomy, in which the agent could essentially be programmed with state machine logic that would allow it to change at which LOA it is operating (Martin and Barber 1996). Barber, Goel, and Martin further proposed that this adaptive capability could also permit AI agents to adapt their decision making framework, as opposed to just the actions they are allowed to execute (Suzanne Barber, Goel, and Martin 2000). This research sparked the idea that an agent's ethical code could be part of the logic that causes an agent to adapt. In a 2009 study of human-automation teams, Langan-Fox, Canty, and Sankey found that time-sensitive team environments also require that such transfers of control be implicit, as opposed to manually driven (Langan-Fox, Canty, and Sankey 2009). A way in which this could be addressed in human-AI teaming is to have AI agents use environmental inputs as factors in their decision making logic.

Ethics in Human-AI Teams

Just as in all-human teams, human-AI teams need to possess a shared understanding of professional ethics for the field in which they work (Smith 2019). This is because a team's shared ethical understanding directly correlates to the team's degree of intra-team trust (Hall 2005). The strength and longevity of a team is highly reliant on intra-team trust, and it serves as main indicator of the team's shared understanding of goals and values (DeRosa et al. 2004). In HATs trust is impacted by the human team members' ability to predict the decisions that the AI is going to make based on an implicit ethical contract (Jacovi et al. 2021). The rules of this contract are those that dictate what behaviors are and are not ethical. Research into building ethically bound AI agents found that these rules include how much freedom and autonomy the agents posses (Rossi and Mattei 2019) and the need

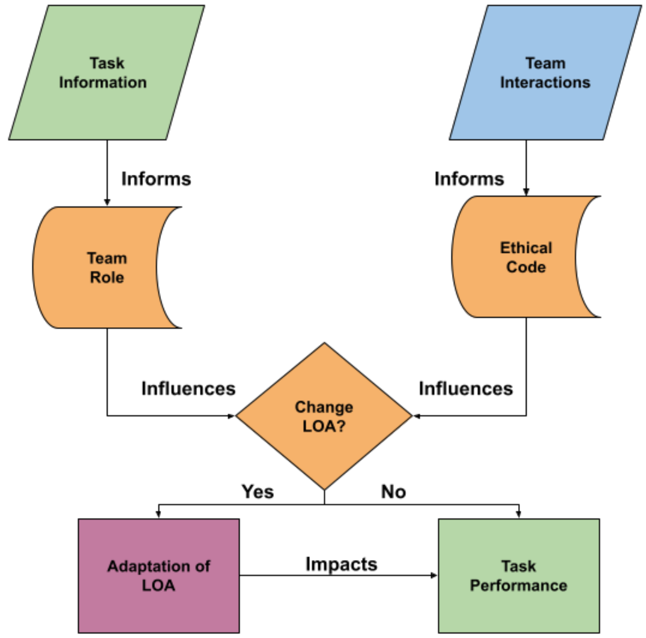

Figure 1: Model of Adaptive Autonomy Informed by Team Ethics

periodic alignment between the AI and socio-technical system within which it operates (Osoba, Boudreaux, and Yeung 2020). Smith's ethical framework for human-machine teams builds on this concept, stating that a AI agents should be restricted from acting in situations where its behavior would be unpredictable or require the insight of its human teammates (Smith 2019).This directly aligns with this paper's choice of adaptive autonomy as the means of implementing changing ethical rules for AI teammates.

The discussion surrounding ethical AI agents is not new to the HCI community. Previous research has highlighted the need for and suggested general ethical principles for the use of autonomous AI agents (Weyns 2020), but what has not been properly addressed is how to technically implement these principles in AI agents, particularly in teaming contexts. A model created by Flathmann and colleagues proposed that in order for this alignment to occur within the context of teaming, the AI agent needs to incorporate team interactions into its conception of the team's ethical ideology (Flathmann et al. 2021). Team interactions serve as a practical embodiment of ethical principles and provide a method for translating abstract concepts to the AI agent (Mittelstadt 2019). Thus, the agents can start with a baseline ethical code and use those interactions to recognize their human teammates' personal ethical codes in order to create a shared team ethical ideology (Flathmann et al. 2021). Flathmann and colleagues' model provides a solid starting point for what a shared ethical code looks like for HATs (Flathmann et al. 2021); however, it does not address a method for technically implementing that ethical code into AI teammates, which is what our model provides.

Model for Adaptive Ethical Autonomy

Flathmann and colleagues devised a model for ethical human-AI teaming that proposed a few key concepts from which our model is derived: that all AI teammates need to start with a baseline ethical ideology, that the AI's ethical ideology is informed and reinforced by team interactions, and that the AI teammate should be capable of continuously learning about the team's ethical ideology (Flathmann et al. 2021). From these core concepts, we devised and propose a two-part model for implementing a shared ethical ideology in human-AI teams through adaptive autonomy. Part I of the model explains how the AI's perception of the team's shared ethical ideology is used as a decision factor in its adaptation logic. Part II of the model explains how the AI teammate's ethical ideology is continuously informed and reinforced through team interactions.

Part I: Informing Adaptation with Ethics

This paper suggests that one way to implement a shared ethical code in AI teammates is through the use of adaptive autonomy. This concept stems from Maes' early research into adaptive autonomy which emphasized that the world itself is best model for AI learning (Maes 1993). Adaptive AI agents possess machine learning modules that allow them to assess the environment and make decisions using those assessments (Martin and Barber 1996). These modules could, as described by Flathmann and colleagues, receive and learn from interactions with the AI's human teammates in order to adapt their ethical constructs (Flathmann et al. 2021).

Recall that Barber and colleagues suggested adaptation as a way to alter not only an agent's LOA, but also its decision making framework. Why not, then, implement a means for incorporating ethics into that framework, such that an AI teammate's decision to adapt would be guided not only by the team's task, but its ethical code as well. Figure 1 models this decision making process. For both this portion of the model, as well as the portion described in Figure 2, the color scheme is as follows: green refers to team tasks and roles, blue refers to team interactions, orange refers to the AI agent's stored data and processes, purple refers to AI agent actions.

The model shows two main environmental inputs, task information and team interactions. The task information informs the agent on its team role in accomplishing the task. The team interactions inform modifications to, or reinforcements of, components to the ethical code. It is important to note that these team interactions include those that are not directed at the agent itself. Studies show that humans act more intentionally and selfishly when interacting with AI systems (March 2021). The degree to which human teammates treat the AI agent as they do other human teammates varies greatly based on how big of an ontological gap they perceive between them and the AI teammates (Guzman 2020). Thus, if the agent were to make assessments on the ethical code of the team based only on one-on-one encounters, its code would be flawed; rather, it needs to also consider how its teammates interact with one another, as well. These aggregate assessments would update the machine's ethical code as needed.

When the AI agent is required to perform a task, it then uses its conception of its team role and its current ethical code to dictate whether or not it should change its level of autonomy in order to perform the task at hand. If it decides its LOA needs to change, it executes the adaptation prior to attempting the team task, which, hopefully, has a positive impact on the task performance. If it decides its LOA does not need to change, it can immediately begin to tackle the team task. This process describes how the ethical code is, overall, incorporated into the AI's state machine logic, but it is also necessary to define how changes are made to the agent's ethical code in terms of how it interprets its team's interactions.

Part II: Ethical Decision Making Process

The entirety of the model presented in Part I is predicated on the AI agent's ability to recognize and evaluate team interactions, compare them to its existing ethical code, and update its ethical rules as necessary. The success of the model presented in Figure 1 is thus dependent on how well the AI agent can interpret and evaluate its team's interactions and translate them into updated or new ethical rules.

Pimental, Kuntz, and Erkov's model for ethical ethical decision making in organizations provides a good starting point for this process. Pimental and colleagues divided ethical decision making into four main stages: recognizing the ethical dilemma, identifying the impacts of the decisions, identifying relevant organizational norms, and identifying other legal regulations relevant to the issue (Pimentel, Kuntz, and Elenkov 2010). Taking these four stages into account, we adapted them for use by an AI teammate in evaluating the team interactions it perceives:

1. Recognizes and defines the ethical dilemma observed.

2. Identifies impact of the ethical decision on relevant stakeholders.

3. identifies team norms associated with the relevant issues.

4. Identifies relevant existing ethical rules in its code and increments a counter if the interaction opposes an existing rule.

5. If necessary, make modification to its ethical code.

This decision making process guides the AI agent's evaluation of if and how to modify its existing ethical code based upon the team interactions that it perceives. Social information processing theory tells us that work group members interpret and take cues from each other's behaviors in order to develop norms and produce a homogeneous manner of behavior (Mayer et al. 2009). Thus, the agent's rule making process needs to be designed in a way that requires the agent to make changes based upon consensus, as opposed to after every new interaction. This is why there needs to be some sort of a counter and minimum threshold, N, that dictates how many of a type of interaction an agent needs to perceive in order to modify or build an ethical rule. Otherwise, its code could be improperly altered by deviant behavior. For instance, if a stressed team member makes a decision that is unethical and in-congruent with the team's values, it should be treated as an outlier by the agent, and not the norm. Figure 2 further explains the process.

The model shows the AI agent observing a team interaction and defining the ethical dilemma it perceives. It then identifies the impact of its teammates decisions and the correlating team norms. The agent then queries its existing ethical code to identify existing ethical rules based on those norms. Next, it compares the interactions to the identified rules and determines if the interactions align with the rules. If they do match, there is no reason to make any alterations to its existing ethical code, and the norm has been successfully reinforced. If they do not match, it increments the counter. This value represents the number of times an interaction needs to be repeated in order to be considered a "team norm." Once the counter is incremented, the agent checks if the counter = N, the preset value for a norm. If it does not, the agent makes no modification to its ethical code. If the counter does equal N, then it modifies its ethical code to include the new or altered rule. In this way, the model provides a mechanism for ensuring that changes to the ethical code are not based upon deviant behavior.

An example of this process is if AI teammate on a computer forensics team were to view a team member read the personal data files of a user during a forensics investigation without asking the team lead for permission. It identifies the stakeholders and impacts of this, such as user privacy and data access, and queries its ethical code for existing relevant rules. It locates one that states personal user files require approval due to privacy concerns. It increments its counter for altering this rule, which brings it to C = 2. It compares it against its set threshold, N = 5, determines it should not yet change the ethical rule, and makes no more changes. Its existing ethical code is then used in the decision making pro-

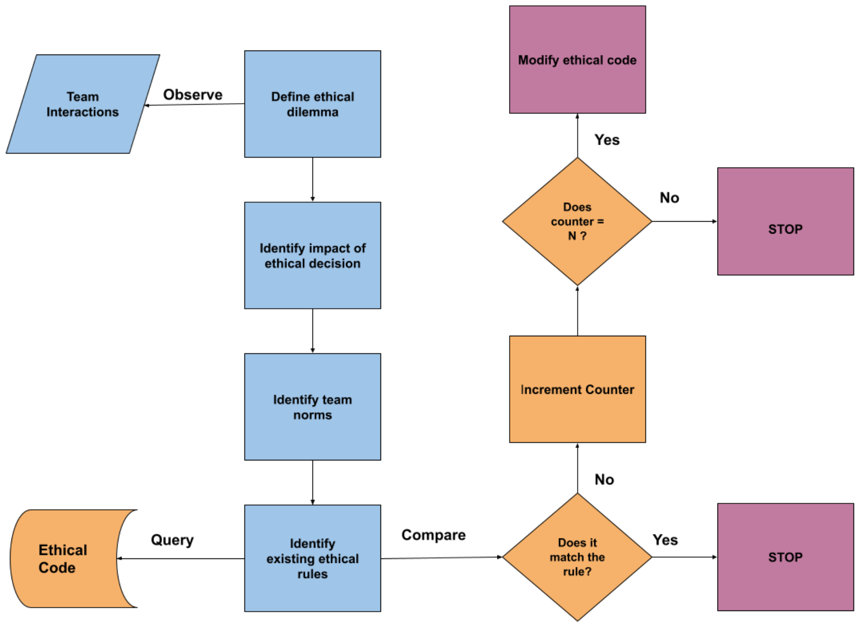

Figure 2: Decision Process for Modifying the Ethical Code

cess described in the first part of the 1, in which it determines it is operating already at the right LOA for its current task of analyzing the computer file systems. If the agent views the same interaction three more times, and the counter reaches 5, it would redefine the ethical rule such that it did not require permission, and elevate its LOA to allow it to act without human approval over its decision. This is a hypothetical example of how the parts of the model could work and interact.

Implications and Future Work

Implications of the Model

The proposed model offers a number of practical contributions for for current and future instances of human-AI teaming. One of the major advantages of this model is that, within this concept, the agent consults both its team role and current ethical code every time it attempts to perform a task, enabling the conditions of its adaptation to be somewhat dynamic. This ensures that the agents decisions to adapt are based on natural interactions with the team, as opposed to the timely interventions against which Langan and colleagues warned (Langan-Fox, Canty, and Sankey 2009). Such natural interactions enable a HAT to both build a cohesive team and accomplish its team goals, because human teammates can trust their AI teammates to continuously make decisions in line with their understanding of good ethical conduct. This trust breeds cooperation, which results in high-performing teams (Hakanen, Häkkinen, and Soudunsaari 2015).

This model can be translated across a variety of human-AI teaming contexts. The team interactions that the AI agent perceives and utilizes in updating its ethical code should be context dependent. For an emergency care team, which is often forced to make in-the-moment ethical decisions, this could be observations of patient triage (Sevimli, Dursun, and Karadas 2015). For a network penetration team, which is repeatedly tasked to determine its rules of engagement, it could be observations what teammates decide are off-limits for the penetration test (Faily, McAlaney, and Iacob 2015). By leaving these team interactions undefined, the model can be easily tailored to work for a specific human-AI teaming context. This is important not only for its utility, but also for its longevity. The use of human-AI teams in new areas of industry, academia, and entertainment is only beginning, and it would be short-sighted to build an ethical model that does not allow for this growth (Mittelstadt 2019).

Limitations and Future Work

Both a limitation and site for future work is the need for empirical validation and testing of the model for continued refinement. An important next step for this research would be the observation of human interactions with AI teammates to see which interactions in what contexts should be used for ethical code adaptation, followed by experimenting with those interactions to see which are most effective. It would also be beneficial to study if human teammates try to influence the AI agent's ethical coding through purposeful interactions, and if it is more beneficial or detrimental for human teammates to know how their actions influence their AI teammate's ethical understanding. A limitation of this paper's model is that it does not consider interactions between AI teammates. An important future consideration would be how AI to AI teammate interactions may affect the team's shared ethical code, as well as if these interactions should be permitted to alter the AI agents' ethical code.

Conclusion

As human-AI teams increase in size and practice, there is an increased need to ensure that these teams are capable of maintaining a shared ethical ideology. This paper presents a model for incorporating a dynamic shared ethical understanding between human and adaptive AI teammates. This is done by utilizing team interactions as ethical signals for the AI agent to consider and use to modify its ethical code. This updated code is then used as a decision factor in the agent's determination of whether or not to adapt its level of autonomy. In this way the agent consistently updates its ethical code based on team interactions and adjusts its decision making framework to reflect changes in ethical expectations and constraints. Figure 2 describes how team interactions inform the agent's ethical code, and Figure 1 shows how this in turn informs its adaptation decisions, thus laying the basis for implementing and maintaining a shared ethical understanding within human-AI teams.

Acknowledgments. This work was supported by AFOSR Award FA9550-20-1-0342 (Program Manager: Laura Steckman).

References

Aitken, M.; Ng, M.; Toreini, E.; van Moorsel, A.; Coopamootoo, K. P.; and Elliott, K. 2020. Keeping it Human: A Focus Group Study of Public Attitudes Towards AI in Banking. In *European Symposium on Research in Computer Security*, 21–38. Springer.

Bagdasarov, Z.; Thiel, C. E.; Johnson, J. F.; Connelly, S.; Harkrider, L. N.; Devenport, L. D.; and Mumford, M. D. 2013. Case-based ethics instruction: The influence of contextual and individual factors in case content on ethical decision-making. *Science and Engineering Ethics* 19(3): 1305–1322.

Bansal, G.; Nushi, B.; Kamar, E.; Weld, D. S.; Lasecki, W. S.; and Horvitz, E. 2019. Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2429–2437.

Demir, M.; McNeese, N. J.; Gorman, J. C.; Cooke, N. J.; Myers, C. W.; and Grimm, D. A. 2021. Exploration of teammate trust and interaction dynamics in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems* .

DeRosa, D. M.; Hantula, D. A.; Kock, N.; and D'Arcy, J. 2004. Trust and leadership in virtual teamwork: A media naturalness perspective. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management* 43(2-3): 219–232.

Faily, S.; McAlaney, J.; and Iacob, C. 2015. Ethical Dilemmas and Dimensions in Penetration Testing. In *HAISA*, 233–242.

Fereidunian, A.; Lehtonen, M.; Lesani, H.; Lucas, C.; and Nordman, M. 2007. Adaptive autonomy: smart cooperative cybernetic systems for more humane automation solutions. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, 202–207. IEEE.

Flathmann, C.; Schelble, B. G.; Zhang, R.; and McNeese, N. J. 2021. Modeling and Guiding the Creation of Ethical Human-AI Teams. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 469–479.

Guzman, A. L. 2020. Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication* 1(1): 3.

Hakanen, M.; Häkkinen, M.; and Soudunsaari, A. 2015. Trust in building high-performing teams: conceptual approach. *Electronic Journal of Business Ethics and Organization Studies* 20.

Hall, M. A. 2005. The importance of trust for ethics, law, and public policy. *Cambridge quarterly of healthcare ethics* 14(2): 156–167.

Haskins, M. E.; Liedtka, J.; and Rosenblum, J. 1998. Beyond teams: Toward an ethic of collaboration. *Organizational Dynamics* 26(4): 34–50.

Jacovi, A.; Marasović, A.; Miller, T.; and Goldberg, Y. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635.

Jarrahi, M. H. 2018. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons* 61(4): 577–586.

Johnson, M.; and Vera, A. 2019. No AI is an island: the case for teaming intelligence. *AI Magazine* 40(1): 16–28.

Klien, 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): 91–95.

Langan-Fox, J.; Canty, J. M.; and Sankey, M. J. 2009. Human–automation teams and adaptable control for future air traffic management. *International Journal of Industrial Ergonomics* 39(5): 894–903.

Maes, P. 1993. Modeling adaptive autonomous agents. *Artificial life* 1(1_2): 135–162.

March, C. 2021. Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players. *Journal of Economic Psychology* 102426.

Martin, C. E.; and Barber, K. S. 1996. Multiple, simultaneous autonomy levels for agent-based systems. In *Proceedings of the fourth international conference on control, automation, robotics, and vision*, 1318–1322. Citeseer.

Mayer, D. M.; Kuenzi, M.; Greenbaum, R.; Bardes, M.; and Salvador, R. B. 2009. How low does ethical leadership flow? Test of a trickle-down model. *Organizational behavior and human decision processes* 108(1): 1–13.

McNeese, N. J.; Demir, M.; Cooke, N. J.; and Myers, C. 2018. Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors* 60(2): 262–273.

Mittelstadt, B. 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1(11): 501–507.

Nyre-Yu, M.; Gutzwiller, R. S.; and Caldwell, B. S. 2019. Observing cyber security incident response: qualitative themes from field research. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, 437–441. SAGE Publications Sage CA: Los Angeles, CA.

Osoba, O. A.; Boudreaux, B.; and Yeung, D. 2020. Steps Towards Value-Aligned Systems. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 332–336.

O'Neill, T.; McNeese, N.; Barron, A.; and Schelble, B. 2020. Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors* 0018720820960865.

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): 286–297.

Petersen, A. M.; Pavlidis, I.; and Semendeferi, I. 2014. A quantitative perspective on ethics in large team science. *Science and engineering ethics* 20(4): 923–945.

Pimentel, J.; Kuntz, J.; and Elenkov, D. S. 2010. Ethical decision-making: an integrative model for business practice. *European Business Review* .

Rossi, F.; and Mattei, N. 2019. Building ethically bounded AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 9785–9789.

Sevimli, S.; Dursun, R.; and Karadas, S. 2015. Health professionals of emergency service: evaluation of triage and ethical values in disaster medicine. *Eurasian Journal of Emergency Medicine* 14(3): 107.

Shilton, K. 2018. Values and ethics in human-computer interaction. *Foundations and Trends® in Human–Computer Interaction* 12(2).

Smith, C. J. 2019. Designing trustworthy AI: A human-machine teaming framework to guide development. *arXiv preprint arXiv:1910.03515* .

Suzanne Barber, K.; Goel, A.; and Martin, C. E. 2000. Dynamic adaptive autonomy in multi-agent systems. *Journal of Experimental & Theoretical Artificial Intelligence* 12(2): 129–147.

Weyns, D. 2020. Towards a code of ethics for autonomous and self-adaptive systems. In *Proceedings of the IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, 163–165.

Wickens, C. D.; Li, H.; Santamaria, A.; Sebok, A.; and Sarter, N. B. 2010. Stages and levels of automation: An integrated meta-analysis. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 54, 389–393. Sage Publications Sage CA: Los Angeles, CA.

Zhang, R.; McNeese, N. J.; Freeman, G.; and Musick, G. 2021. " An Ideal Human" Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW3): 1–25.