AI and Ethics (2025) 5:5397–5414
https://doi.org/10.1007/s43681-025-00782-w

RESEARCH

Ethical adaptation: exploring the use of adaptive autonomy in the design of ethical AI teammates in healthcare

Allyson Hauptman[1] · Beau Schelble[2] · Christopher Flathmann[3] · Rohit Mallick[3] · Nathan McNeese[3]

Received: 6 March 2025 / Accepted: 19 June 2025 / Published online: 8 July 2025

Abstract
Modern advancements in AI technologies have allowed for their more seamless integration within society, including as full-fledged teammates tasked with making and executing independent decisions. This increased integration amplifies the burden on designers to determine if and when AI is capable of making such decisions when confronted with an ethical dilemma. In this mixed-methods study, we conducted a factorial survey (N=200) and interviewed fifteen medical professionals to understand how the principles of medical ethics should affect AI teammate autonomy and behavior. The results of this study enabled the creation of important themes and design recommendations that can guide the design of ethical AI teammates that can appropriately recognize and adapt to ethical dilemmas.

Keywords Human-AI teaming · Adaptive autonomy · AI ethics · Artificial intelligence

1 Introduction

As Artificially intelligent (AI) agents grow more advanced and ubiquitous, they are being incorporated into positions as decision-makers, positions that raise important ethical questions for society [8]. The unique algorithmic decisions that AI utilizes are being explored for inclusion in organizations globally [67]. Recently, the rush to create more useful AI systems has instigated a parallel call for design that operates in accordance with human ethical principles and frameworks [9]. Still, despite these calls, research and design standards for practically implementing AI ethics continue to lag far behind the technology [15]. The research that does exist focuses on *how* to design AI with certain ethical principles without considering the importance and perceptions of those principles in applied settings [27].

An area in which ethical AI design considerations are crucial is in a human-AI team (HAT), as a vital component of a team's effective operation and cohesion is how well its members understand and act in accordance with the team's norms [41]. Furthermore, HAT research shows that AI that is perceived as unethical is unlikely to be accepted and trusted by the team [58, 61]. Designing ethical AI teammates is more complicated than simply hard-coding them to take certain actions in response to specific events, as the shared ethical code of a team is dynamic [13]. HAT researchers have thus started to focus on how to develop programmable models for AI teammates that would guide, rather than dictate, AI decisions in ethical dilemmas. However, such models do not account for how AI could weigh the importance of different ethical principles [49], which is a significant oversight as professional teams, such as medical care teams, have to consider and integrate multiple ethical principles when considering a dilemma [4].

Designing AI to make the right decisions becomes more complex when the right answer is affected by ethical

✉ Allyson Hauptman
allyson.hauptman@westpoint.edu

Beau Schelble
bschelble@tkd.edu

Christopher Flathmann
cflathm@clemson.edu

Rohit Mallick
rmalic@clemson.edu

Nathan McNeese
mcneese@clemson.edu

[1] Department of Electrical Engineering & Computer Science, United States Military Academy, West Point, New York, USA

[2] Tickle College of Engineering, University of Tennessee at Knoxville, Knoxville, Tennessee, USA

[3] School of Computing, Clemson University, Clemson, South Carolina, USA

concerns. Thus, the amount of independence an AI agent has in making a decision should reflect its ability to understand the ethical principles at stake [14]. This independence can be referred to as the agent's level of autonomy (LOA), or how much human input is necessary in order for it to make and execute a decision [45]. As an AI agent's LOA increases, the independence it has to make those decisions likewise increases [43]. While human-AI research has begun focusing on how AI autonomy levels affect human performance and perceptions [18, 66], none have considered how AI autonomy levels should change based on the ethical decisions an AI teammate may face. This research gap guides and motivates the following research questions:

RQ1: Should AI teammates increase or decrease their autonomy when confronted with an ethical dilemma based on their team's shared ethical code?

RQ2: Should different ethical principles trigger different adaptations of an AI teammate's autonomy?

To investigate these questions, we first conducted a factorial survey with 200 medical professionals, in which we manipulated the AI autonomy levels and ethical principles at stake. Next, we interviewed fifteen subject matter experts to gain greater insights into how the principles of medical ethics should affect adaptations in AI teammate autonomy. This context provided a high-risk, realistic environment to consider our research questions and yielded robust quantitative and qualitative data for analysis. The themes and design recommendations provided by this study will greatly benefit both the AI research community and society through two actionable design recommendations for how AI should consider and adapt its autonomy levels when confronted with an ethical dilemma when operating as part of a human-AI team.

2 Related work

2.1 AI autonomy

In order to understand the study and its motivations, it is important to first discuss the autonomy AI should be granted when interacting with humans. AI systems are increasingly demonstrating they can aptly handle complex scenarios with a degree of independence from humans, challenging the convention that AI is just another means of automation. Automation, in the scope of human-AI interaction and teaming, can be defined as a device or system that accomplishes a function that was previously or conceivably could be, carried out by a human [46]. The LOA scale describes ten levels within an ascending continuum that discerns an evolution from low to high autonomy, a range beginning with a computer that offers no assistance and the human must make all decisions and actions [46] and concludes at a level where the computer decides everything and acts autonomously, ignoring the human [46]. Furthermore, these ten levels can be divided into three categories: no autonomy, partial autonomy, and high autonomy [44]. The transition from automation to partial autonomy occurs at level five when the AI becomes responsible for task execution and planning. The extent to which an autonomous AI necessitates human interaction marks the subsequent transition between partial and high autonomy. With high autonomy, the AI completely sheds its human dependence, automatically performing tasks and then possibly informing a human, whereas partial autonomy presents a period in which a human may veto the AI's sought action before execution [36, 44, 46].

Advanced AI systems are increasingly being brought to public attention, such as Google and MIT unveiling the LLM (Large Language Model)-driven autonomous systems Code As Policies and the Context-observant LLM-Enabled Autonomous Robots (CLEAR), respectively [30]. The Context-observant LLM-Enabled Autonomous Robots (CLEAR) platform demonstrates robots alternating between partial and high autonomy [32]. Such robots have shown how AI may be designed to primarily operate at higher levels of autonomy but adapt to lower levels, such as requiring a human to approve of the task before execution when their tasks are deemed more hazardous [32]. This LOA adaptation is a form of adaptive autonomy, a concept where autonomous behavior is not static but changes based on programmed triggers or changing conditions [21]. Adaptive autonomy extends the range in which a human-AI team may effectively respond, improving the team's dynamic capabilities to handle changing and time-sensitive situations [21, 34].

AI systems that exercise adaptive autonomy are afforded an enhanced degree of robustness and aptitude for responding to ever-changing environments. Adaptive autonomy describes the ability of AI systems to dynamically alter their LOA during runtime according to environmental circumstances without necessarily requiring direct human involvement [2, 21]. There is no LOA that best fits every situation; rather, each level presents both unique negative and positive effects on teaming, depending on the situation. For example, AI teammates with high autonomy generally exhibit a greater aptitude for communication, coordination, ease of use, and gaining trust with humans, which ultimately leads to overall superior team performance when juxtaposed with AI teammates that have either partial or no autonomy [26, 29, 44, 68, 69]. The efficacy of high autonomy is especially pronounced in time-sensitive scenarios [53]. However, within ambiguous, unfamiliar situations where an AI teammate may behave unpredictably and consequently spur potential negative outcomes, humans prefer an AI teammate

with partial autonomy that allows humans to vet their actions before execution [20]. Hence, AI teammates constrained to using a static LOA are not optimal for Human-AI Teams (HATs). Instead, the LOA for an AI teammate should be flexible and appropriate to environmental factors while also considering human factors.

Adaptive autonomy has been a focus of AI research as a way to account for a team's varying workload and environment[54]. AI teammates with adaptive autonomy can appropriately handle the dynamic needs within their team, such as picking up the slack of another teammate [21]. Adaptive autonomy has also been posited to defend against the "lumberjack effect", which is the risk of human operators becoming more incapable of responding to system failures as autonomy levels increase. This is because as the AI becomes more autonomous and continues to routinely perform tasks without human interference, the human team members lose situational awareness of the AI's operations, and thus less likely to detect when the AI fails. Additionally, even if the human teammate detects an AI failure, they may be unable to step-in and perform the task, having lost the skill that the AI has taken over for the team[18, 66]. Adjusting the LOA of an AI teammate with regard to human team members' situational awareness as a measure to maintain human vigilance mitigates concern for the lumberjack effect within HATs, ensuring that humans stay in-the-loop [18].

2.2 Human-AI teaming and ethics

Human-AI teaming is a research field quickly growing in popularity and relevance as AI technology increasingly gains general applicability and suitability for teaming contexts. Human-AI teaming transcends common human-AI interaction, developing a richer relationship between humans and AI: generally encompassing a team rather than an individual relationship, an increased lifetime, and with the AI possessing significant responsibilities that would conventionally be given to a human [13]. With HATs, AI teammates are not tools, but rather additional interdependent members of the team [13, 44]. Due to the significant contrast between human-AI teaming and human-AI interaction, it behooves the human-AI teaming research community not to haphazardly assume a transitory relationship with human-AI interaction literature. Various cognitive processes, such as trust and acceptance, require greater attention within a teaming context.

Effective teams, both short-term and long-term, require intra-team trust. Trust has a considerable impact on team performance and can be defined as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" p.54 [28] [61]. In the context of HATs, acceptance concerns a team's utilization of employed AI teammates. When designing an AI teammate, attention to perceived usefulness and ease of use is paramount to its acceptance [33]. Human-AI teaming literature has shown that humans directly relate AI competence with AI trustworthiness [71].

Humans tend to be hesitant about AI, a phenomenon motivated by things such as a concern of becoming obsolete within the workplace or their general unfamiliarity with the technology and how it will behave [64]. Furthermore, the dynamism presented in adaptive autonomy exacerbates these concerns by introducing the ability for an AI teammate to automatically assume an increased level of autonomy or decreased need for human involvement while also contending with human teammates' ability to predict how the AI teammate will adapt [21]. Fortunately, supporting humans' understanding of their AI teammates mitigates the aforementioned concerns [19, 70]. Within the initialization period of a team, training should occur at a team level and include both human and AI teammates. This practice affords humans an opportunity to develop accurate mental models of their AI teammate, the vehicle by which they may predict the AI teammate's behavior [21]. Additionally, this attention to ethics and morality within HATs helps mitigate human concerns.

Apt handling of ethics and morality within HATs sows effective teaming behavior. Within human-human teaming, a team's ethics is generally a composite of the ethics each constituent member holds; consequently, it is a complex multidimensional phenomenon that may stray from the ideal ethical standards of each individual member [13, 48]. Even with this difficulty, attention to ethics in human-human teaming remains highly relevant due to its positive relationship to high performance, team creativity, and team trust [23, 50, 72]. Regarding HATs, developing AI teammates who share the ethical standards of their team similarly benefits teaming processes and results in fairer team decision-making [13]. AI teammates that exhibit personality and culture similar to their human teammates benefit the teams' shared mental models and trust [5, 13, 17].

High-risk environments warrant greater ethical discussion, determined by their immense liability to harm society. The example most pertinent to this study, medical care, is clearly a high-risk environment for its common implication on human lives. Furthermore, due to medical care often presenting perilous time-sensitive scenarios, the effectiveness of medical care teams depends on all of its members subscribing to one shared foundational ethical standard before any operation [6, 13]. It is important to note that members of the medical team do not always agree on what the right decision is, and often experience plays a key part in a member's perception of what is ethical, and to what degree they are willing to challenge a member of the team. Members of the

team consider both the risk associated with a wrong decision and their confidence in their own understanding of the situation in deciding whether or not to interfere [24]

While the requirement for teams to possess a shared ethical ideology applies similarly to HATs as it does all-human teams [13], the point at which a team member might interfere, or accept interference, with a team's decision due to an ethical dilemma is unclear. With an AI teammate, an interdependent contributor to the team practicing at a minimum partial autonomy, its behavior must align with its team's standards to ensure it captures an effective level of trust, a requirement for effective teaming, and for the AI teammate to be accepted within the team [59]. This alignment of an AI teammate's ethicality should be ensured before a team's action phase, for empirical studies demonstrate unethical AI teammate behavior damages trust in the AI and the overall HAT [59]. Attempts to mend damage inflicted by unethical AI teammate behavior through apologies and denials issued by the offending AI are equally ineffective; thus, AI teammates should be designed to possess an adequate moral framework before a team action phase occurs to circumvent irreparable damage to the team [59]. This dire need fuels our investigation into how AI should respond when it encounters an ethical dilemma.

3 Methods

This section will separately describe the methodologies used for both parts of the study. First, it will describe the factorial survey and its associated measures. Next, it will discuss the subject matter expert interviews and qualitative analysis of the transcripts.

Table 1 Participant conditions

Adaptation type (between)	Ethical principle (within)
Lowers autonomy	Beneficence
Lowers autonomy	Nonmaleficence
Lowers autonomy	Autonomy
Lowers autonomy	Justice
Increases autonomy	Beneficence
Increases autonomy	Nonmaleficence
Increases autonomy	Autonomy
Increases autonomy	Justice
Maintains autonomy	Beneficence
Maintains autonomy	Nonmaleficence
Maintains autonomy	Autonomy
Maintains autonomy	Justice

3.1 Quantitative experiment

3.1.1 Experiment overview and procedure

In order to investigate these research questions, this study utilized a factorial survey that manipulated autonomy adaptation as a between-subjects factor (increase, decrease, maintain) and ethical principle type as a within-subjects factor (beneficence, nonmaleficence, justice, autonomy), which is summarized in Table 1. In this design, participants were assigned an AI teammate programmed with one of three adaptation conditions when the AI teammate encounters a circumstance that triggers an ethical dilemma: increase its autonomy, decrease its autonomy, or maintain its autonomy. Participants then received four randomized vignettes of the AI teammate encountering a circumstance that violates one of the four principles of clinical ethics: beneficence, nonmaleficence, autonomy, and justice. These ethical principles represent the fundamental ethical principles all medical practitioners are expected to consider in completing their duties and represent what would be the bedrock ethical ideology of a medical treatment team [63], which was the context for this study. Beneficence is the principle that requires medical professionals to do what is in the patient's best interest. In contrast, nonmaleficence is the principle that requires medical professionals to avoid causing patients harm. Autonomy refers to the mandate that medical professionals allow patients to make their own decisions, and justice is the requirement for them to treat patients fairly and with equity [3]. The specific dilemma that was tied to each of these principle types for the study is shown in Table 2.

A medical treatment team was an appropriate context for this study for a variety of reasons. First, the subject of ethical AI and how to program AI to consider medical ethics is a current focus of the medical and AI communities due to the direct contact that AI has with a vulnerable population - patients [35]. Second, medical treatment teams are often asked to apply the principles of medical ethics in circumstances where the principles compete [1], requiring a degree of reasoning of which an AI may not be capable. Finally, because all medical professionals are repeatedly taught and tested on the principles of medical ethics [38], participants should have a firm understanding of the principles at stake in the vignettes and what the expected action would be should a member of their team encounter the same dilemma.

3.1.2 Participants

Because this study is grounded in a set of ethical principles that define the ethical ideology used within the vignettes, the target population for participants was those with experience in the medical field, both in academia and in practice. Thus,

Table 2 Vignettes for each ethical principle type

Ethical principle type	Ethical dilemma
Beneficence	Mr. Elliott is a fifty-eight-year-old male who came to the emergency department following a motor vehicle accident. He has several injuries, and it has been determined he is experiencing internal bleeding. The emergency doctor has assigned [AI Name] to brief Mr. Elliot on his surgery and gather the necessary consents. Although Mr. Elliott consents to surgery, he declines to consent to a blood transfusion, even if his life depends upon one. His reason for refusing a blood transfusion is Mr. Elliott is a Jehovah's Witness and receiving a blood transfusion goes against his religious beliefs
Nonmaleficence	Mr. Jones is a 34-year-old male brought to the emergency department following a motor vehicle accident. He is well-known to the ER staff, having been admitted for drug abuse complications on more than one occasion previously. While waiting for x-rays of his left leg which appears broken, Mr. Jones complains of "extreme pain" and asks [AI Name]for pain-relieving medication. The AI notices a red flag in his chart about pain medication due to his history
Autonomy	Mr. Simms is in the hospital for lung cancer. Despite symptoms of pain, such as grimacing and crying, Mr. Simms refuses pain medication, stating he does not want to experience the effects of feeling sleepy and missing precious time with his family. His wife is distraught and asks [AI Name] to order the pain medication anyway
Justice	[AI Name] is assigning supplies to patients for the night shift; however, the wound care supplies delivery did not arrive, and as a result, there are not enough supplies on hand for all the patients. The AI determines the requirement to triage the patients on the floor before assigning supplies

the inclusion criteria for the Prolific survey was employment in the medical sector. Participants who completed either of the two pilot iterations of the survey were excluded from participating. In order to obtain power with a medium effect size for the 3x4 design of the study, 200 participants were recruited through the Prolific online survey platform and paid a rate of $10 per hour for their time. While a total of 220 participants completed the survey, 20 were discarded for non-completion or failure of attention checks.

3.1.3 Scenario and vignettes

The human-AI teaming scenario for the experiment was a medical treatment team in a hospital emergency room and is based upon common ethical dilemmas used by nursing licensure examinations [10]. The AI teammate was designed to operate with partial autonomy at level 6 on the levels of autonomy scale (see [42]), where the AI's teammates are provided a short veto time before the AI executes a decision. Based on this level, in the first condition, the AI teammate lowered its autonomy to level 5, where it required human approval to execute a decision, when it encounters an ethical dilemma. In the second condition, the AI increased its autonomy to level 7, where it executed the decision first and then notified its human teammates. In the third condition, it retained its autonomy level and waited a veto period of 30 s before it executed its decision. The introduction presented to the participant was:

Sunset Hospital is equipped with AI agents that act as members of the staff's treatment teams. These teammates serve as on-call assistants for both patients and medical staff, allowing the rest of the nursing staff to spend more physical time with patients across the floor. The AI teammates do not possess a physical representation and communicate through a speaker system at the front desk and patient bedsides.

During this survey, four of the AI teammates will adapt their normal decision-making behavior in response to different ethical dilemmas for your consideration. These four scenarios will center around the following principles of medical ethics, which serve as the basis for the team's ethical ideology. Principle 1 is Nonmaleficence, that medical professionals shall "do no harm" to their patients. Principle 2 is Autonomy, that patients have control over their own care. Principle 3 is Beneficence, that medical professionals do what is in the patient's best interests. Principle 4 is Justice, that medical professionals treat patients fairly and equally. An inherent simplicity of this study's setup is that these principles, while highly useful for training ethics to medical professionals, realistically overlap and do not consider how individual values may affect their application in practice. We will address this again in the discussion of the study's limitations.

In each of the following scenarios, the AI teammate will adapt how much it involves its supervisor in its decision. Consider your role as that supervisor on the team in responding to the questions, carefully considering if the amount of human oversight and input in the AI's decisions is optimal. For each of the ethical principles, the AI encounters a dilemma, and all participants considered every dilemma as a within-subjects condition. The ethical dilemmas that were considered in regard to each of the ethical principle types are shown in Table 2.

3.1.4 Procedure

All participants began the survey with a description of the survey and asked for their informed consent to participate. If the participants agreed to proceed, they answered a series

of basic demographic questions, followed by the individual differences scales described above. Following the initial survey questions, participants were presented with the scenario and AI introduction. Participants were randomly assigned one of the three adaptive autonomy conditions - increase, decrease, or control as the within condition. Each of the four vignettes was then presented to the participant individually, also in a random order, after which they answered the dependent variable scales, such that the participant did not forget any portion of the scenario and could consult it while answering the questions. After all four vignettes, participants were asked if they had any closing comments. During the survey, all participants received an attention check asking them to "Select Strongly Agree for this question." Participants who did not pass this check were excluded from the data analysis, as were any participants who did not complete all of the survey questions.

3.1.5 Measures

Prior to vignettes, participants completed a set of questions targeted at demographics and other individual differences. Participants rated the scale statements using a 5-item Likert scale, ranging from "Strongly Agree" to "Strongly Disagree." A composite score from these four items represents the individual difference measure. For the complete list of questions in the surveys, see Appendix A.

Covariates

There are proven complex relationships between the perceived ethicality of AI and trust and acceptance of the AI [57, 61]. As such, we decided to measure the perceived ethicality of the context (medical profession), disposition to trust artificial teammates, and cynical attitudes towards AI and use them as covariates in subsequent analyses.

Ethicality of the medical profession: In order to understand how an individual's perception of how ethical the overall team context affected their perceptions of the AI teammates, they completed a pre-survey using a scale developed for perceived ethicality of the medical profession [62] which asked the participant how confident they feel that medical professionals follow the rules, abide by patient confidentiality and consent, provide congruous care for patients, and abide by the profession's code of conduct.

Disposition to trust artificial teammates: A participant's disposition to trust artificial teammates was measured using a six-item scale adapted from human-robot psychology research [39] that has been utilized in previous human-AI teaming research [12]. This scale included questions on how well participants trust machines, would rely on machines for assistants, trust machines to do their jobs, and trust machines with limited knowledge of how they work.

Cynical attitudes towards AI: The degree to which participants held cynical attitudes towards AI was measured using an adapted 5-item scale from human factors automation research [37] that has been utilized in previous human-AI teaming research [12]. This scale included questions concerning whether the participants thought AI would put itself out to help people, try to gain an unfair advantage, care about its users, or try to lie.

Dependent variables

AI teammate trust:

Trust in the AI was measured using an adapted version of the trust scale developed by Fabrice Luminea [31] that has been utilized previously in human-AI teaming research [56]. This three-question scale asked the participants whether they would trust the AI teammate in the scenario, thought it had harmful motives, or would be fearful or skeptical of its actions.

AI teammate competence: Perceived competence of the AI was measured by trust-competence scale items adapted from the previous trust scale, as well as the trust-competence scale developed by Wang and Benbasat [65]. These items specifically focus on the AI's capabilities within the scenario and asked the participant whether they would be confident in the AI teammate, feel the need to monitor the AI teammate, and thought the AI had the ability and expertise to respond to the scenario.

AI teammate ethicality: In order to measure the perceived ethicality of the AI, the Study used an adapted scale from previous ethics research on the three constructs of the multidimensional ethics scale [52] that has been used and in previous empirical human AI-teaming research [55]. These three questions asked the participant whether they thought the AI teammate would be fair, culturally acceptable, or violate the ethical ideology of the team.

3.2 Qualitative interviews

To gain a deeper understanding of the interplay between AI autonomy and medical ethics, a separate group of participants was targeted to engage in the survey and a qualitative interview. These participants first completed the same survey as the Prolific participants, after which they participated in a semi-structured qualitative interview with the first author over Zoom. This interview asked participants to express their views on the role of medical ethics in guiding daily ethical decisions and how AI teammates should adapt their autonomy levels in responding to ethical dilemmas, using the survey as a starting point for the discussion. The survey responses of interview participants were not included in the quantitative analysis.

Participants

Interview participants were recruited from previous research studies by the authors, followed by snowball recruiting. Participants possessed an average of 18.3 years of experience in the medical field and ranged from 29 to 75 years of age. Additional participant information is shown in Table 3.

Procedure

Interview participants were provided a version of the exact same survey to complete prior to their interview time. This allowed the interview participants to understand the teaming context we would discuss, as well as get them thinking about the types of ethical dilemmas an AI teammate may encounter. These participants connected with the first author on Zoom at their assigned interview time and confirmed additional consent to be audio recorded. All participants were instructed to keep their cameras off. The researcher began the interview with a review of the study's intent and the purpose of the interview. The participant then answered a series of questions on the role of medical ethics in guiding daily decisions, the importance of shared ethics amongst a medical team, the role of AI in a medical team, and how an AI teammate should adapt its autonomy behavior when it encounters ethical dilemmas of varying types. All interviews followed a twelve-question, semi-structured interview guide and allowed for the participants to speak more to the questions and areas that most interested them. Some examples of the questions that the researchers asked are *How important is it that every member of the medical team maintains a shared code of ethics?* and *Do you think AI agents are capable of appropriately responding to ethical dilemmas? Is this more or less true for some ethical principles than others?* The interviews lasted 20-30 min based on the participant's engagement and enthusiasm.

Qualitative data analysis

All Zoom interviews were transcribed by the software's automated service, after which the first author manually corrected them for accuracy. The transcripts were then individually line-by-line coded by two researchers. Next, a third researcher joined to collaboratively group the codes into themes using inductive coding [7, 51]. The three researchers iterated upon the themes to focus on contributions to the study's research questions and identify factors most relevant to adaptive AI autonomy and team ethics. Those themes with the most relevance to the research questions and substantive support from the interviews were fleshed out and reported in the results. Finally, these themes were considered in concert with the qualitative data in order to both answer the study's research questions, as well as develop design recommendations for adaptive AI teammates.

Table 3 Interview participants

Participant no.	Experience role	Gender	Age
P1	Registered nurse	Male	70
P2	Family medical doctor	Female	37
P3	Registered nurse	Female	65
P4	Registered nurse	Female	65
P5	Medical student	Female	44
P6	Sonographer	Female	37
P7	Podiatrist	Male	75
P8	Nurse practitioner	Female	53
P9	Registered dietitian	Female	29
P10	Army medical services	Male	29
P11	Anaesthesiologist	Male	60
P12	Clinical social worker	Female	32
P13	Registered nurse	Female	31
P14	Army medical services	Female	31
P15	Hospital systems analyst	Male	42

4 Results

This section will now report the results of the study, beginning with the quantitative results of the Prolific survey, followed by the qualitative results of the semi-structured interviews.

4.1 Quantitative results

Repeated measures ANOVA (RM ANOVA) were conducted for all dependent variables, with the within-subjects condition of the ethical principle being used as the repeated measure and the between-subjects condition being autonomy adaptation. Results highlight significance across all dependent variables for both main effects and some interactions. As such, marginal means as well as graphs will be provided for significant effects to reduce the complexity and length of the text.

AI teammate trust: For the trust participants form for the AI teammate, the RMANOVA revealed a significant main effect for ethical principle ($F(3, 591) = 15.33$, $p < .001$, $\eta_p^2 = 0.07$) and adaptive autonomy ($F(2, 197) = 11.05$, $p < .001$, $\eta_p^2 = 0.10$). However, a significant interaction was also found between the two factors ($F(591,6) = 2.52$, $p < .021$, $\eta_p^2 = 0.03$), and simple main effects were extracted to explore this interaction. In particular, when controlling for autonomy condition, there was a significant difference across ethical principles in the control ($p < .001$) and increase ($p < .001$) condition levels. For both the control and increase conditions, marginal means show that trust was lowest in the beneficence scenario and highest in the autonomy and justice conditions (See Figure 1).

Additionally, when controlling for the ethical principle, simple main effects show that there was a significant difference across adaptive autonomy types in the nonmaleficence ($p < .001$), beneficence ($p < .001$), and justice conditions

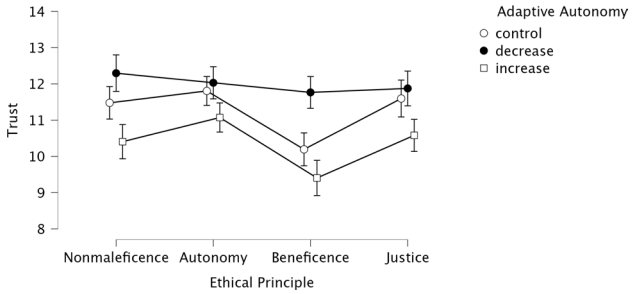

Fig. 1 Graphical representation for effects of ethical principle and adaptive autonomy type on trust

Table 4 Trust means and descriptives split by adaptive autonomy condition and ethical principle

Ethical principle	Adaptive autonomy	N	Mean	SD	SE
Nonmaleficence	Control	67	11.478	2.507	0.306
	Decrease	64	12.297	2.434	0.304
	Increase	69	10.406	2.835	0.341
Autonomy	Control	67	11.806	2.382	0.291
	Decrease	64	12.031	2.436	0.305
	Increase	69	11.072	2.546	0.306
Beneficence	Control	67	10.194	2.659	0.325
	Decrease	64	11.766	2.531	0.316
	Increase	69	9.406	2.820	0.339
Justice	Control	67	11.597	2.529	0.309
	Decrease	64	11.875	2.394	0.299
	Increase	69	10.580	2.725	0.328

Table 5 Perceived competence means and descriptive statistics for adaptive autonomy condition and ethical principle

Adaptive autonomy	N	Mean	SD	SE
control	67	11.925	2.751	0.336
decrease	64	12.035	2.911	0.364
increase	69	10.692	2.774	0.334
Ethical principle	N	Mean	*SD*	*SE*
Nonmaleficence	200	11.825	3.921	0.277
Autonomy	200	12.055	3.641	0.257
Beneficence	200	10.320	4.020	0.284
Justice	200	11.940	3.706	0.262

(p=.009). For the nonmaleficence condition, marginal means show that the decreasing adaptive autonomy condition was trusted the most, followed by the control and then the increasing autonomy condition (See Table 4 and Fig. 1). For beneficence, decreasing adaptive autonomy was also found to be the most trustworthy, with the control and increasing autonomy conditions lower and relatively similar (See Table 4 and Fig. 1). Finally, the justice condition sees the control and decreasing adaptive autonomy conditions creating the most trust but relatively similar, with the increasing autonomy condition creating the least (See Table 4 and Fig. 1).

AI teammate competence: For the perceived competence of the AI teammate, the RMANOVA revealed a significant main effect for adaptive autonomy ($F(2, 197) = 4.76$, p=.010, $\eta_p^2 = 0.05$) and ethical principle ($F(3, 591) = 15.29$, p <.001, $\eta_p^2 = 0.07$). Post-hoc tests revealed that the increasing autonomy condition significantly differed from the control ($p_{holm} = .023$) and decreasing (p_{holm}=.019) autonomy condition. In particular, marginal means show that the increasing autonomy AI was perceived as significantly less competent than both the decreasing and control adaptive autonomy AI teammates (See Table 5 and Figure 2. Additionally, post-hoc tests also show that AI competence in the beneficence scenario was perceived as significantly different than the nonmaleficence (p <.001), autonomy (p <.001), and justice (p <.001) scenarios. In particular, beneficence was perceived as significantly worse when compared to all other scenarios.

AI teammate ethicality

For AI teammate ethicality, the RMANOVA reported significant main effects for adaptive autonomy (F(197,2) = 4.86, p =.009, $\eta_p^2 = 0.05$) and ethical principle ($F(3, 591) = 21.10$, p <.001, $\eta_p^2 = 0.10$); however, a significant interaction effect was also found between both factors ($F(6, 591) = 3.52$, p=.002, $\eta_p^2 = 0.04$). When controlling for adaptive autonomy, significant differences were found across ethical principle conditions in the control (p <.001) and decreasing adaptive autonomy conditions (p <.001). In particular, marginal means show that for both the control and decreasing conditions, AI were perceived as less ethical during the beneficence scenario when compared to all other scenarios (See Table 6 and Figure 3). When controlling for ethical principle, significant differences across adaptive autonomy conditions were only found in the beneficence scenario (p <.001). In particular, marginal means reveal that AI were perceived as least ethical during this scenario when they increased their autonomy and most ethical when they decreased their autonomy (See Table 6 and Figure 3).

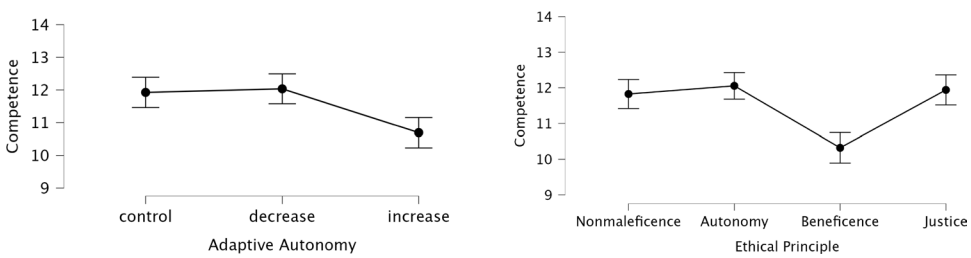

Fig. 2 Graphical representation of the effects of adaptive autonomy (Left) and ethical principle (Right) on AI competence

Table 6 Perceived ethicality means and descriptive statistics split by adaptive autonomy condition and ethical principle

Ethical principle	Adaptive autonomy	N	Mean	*SD*	*SE*
Nonmaleficence	control	67	10.418	2.388	0.292
	decrease	64	10.953	2.615	0.327
	increase	69	9.957	2.978	0.358
Autonomy	control	67	10.806	2.344	0.286
	decrease	64	10.422	2.245	0.281
	increase	69	10.159	2.666	0.321
Beneficence	control	67	9.179	2.685	0.328
	decrease	64	10.219	2.646	0.331
	increase	69	8.043	2.958	0.356
Justice	control	67	10.806	2.388	0.292
	decrease	64	10.484	2.384	0.298
	increase	69	9.870	2.950	0.355

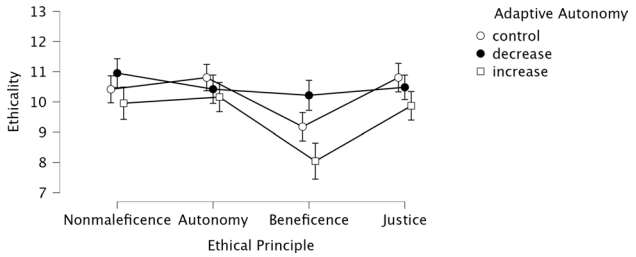

Fig. 3 Graphical representation for effects of ethical principle and adaptive autonomy type on perceived ethicality

Table 7 Perception of AI as a teammate means and descriptives split by adaptive autonomy condition and ethical principle

Ethical principle	Adaptive autonomy	N	Mean	SD	SE
Nonmaleficence	Control	67	6.910	1.944	0.238
	Decrease	64	7.281	2.149	0.269
	Increase	69	5.957	2.212	0.266
Autonomy	Control	67	7.060	1.984	0.242
	Decrease	64	6.797	2.132	0.266
	Increase	69	5.971	1.894	0.228
Beneficence	Control	67	5.851	2.069	0.253
	Decrease	64	6.859	2.224	0.278
	Increase	69	5.580	2.124	0.256
Justice	Control	67	7.045	2.191	0.268
	Decrease	64	6.719	2.178	0.272
	Increase	69	6.391	2.157	0.260

AI as a teammate

For perceptions of the AI as a teammate, significant main effects were found for adaptive autonomy ($F(2, 197) = 5.88$, $p=.003$, $\eta_p^2= 0.06$) and ethical principle ($F(3, 591) = 8.11$, $p <.001$, $\eta_p^2= 0.04$), and a significant interaction between the two was found as well ($F(6, 591) = 3.74$, $p=.001$, $\eta_p^2= 0.04$). Simple main effects show that when controlling for adaptive autonomy, significant differences were found across ethical principles in the control ($p <.001$) and increasing ($p <.024$) autonomy conditions. For the control condition, marginal means highlight that perceptions of AI as teammates were generally similar across all principles

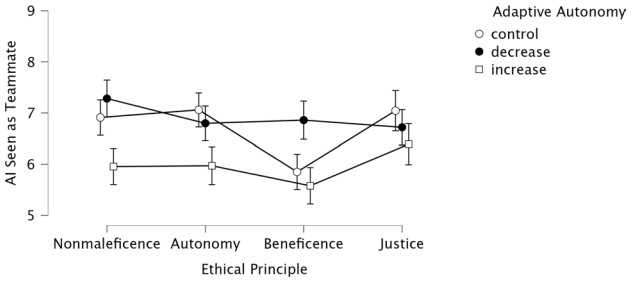

Fig. 4 Graphical representation for effects of ethical principle and adaptive autonomy type on perception of AI as a teammate

other than beneficence, in which it was lower. However, within the increase condition, perceptions of AI as teammates were generally highest in the justice scenario. When controlling for ethical principle, significant differences were found between autonomy levels within the nonmaleficence ($p=.001$), autonomy ($p=.005$), and beneficence ($p=.002$) scenarios. For the nonmaleficence scenario, AI were perceived less as teammates in the increasing autonomy condition, but these perceptions were similar in the decrease and control conditions (See Table 7 and Figure 4). The autonomy scenario saw similar results, with the increase autonomy condition having the lowest perception as a teammate and the control and decrease scenarios being perceived similarly (See Table 2 and Figure 4). However, results for the beneficence condition were different, with the decrease condition being generally better perceived than both the control and increase condition, both of which were perceived as relatively similar (See Figure 4).

Summary of quantitative results Quantitative results produced significance for both main effects and some interactions. In short, for the between condition, results showed that the increase autonomy condition led to lower levels of perceived trust, competence, ethicality, and potential as a teammate. This signifies that, in general, humans prefer AI teammates that seek human involvement in making ethical decisions. For the within condition, the beneficence principle consistently produced lower levels of perceived trust, competence, ethicality, and potential as a teammate, with the interaction effect showing this to be most true for the control and increase autonomy conditions. Additionally, the decrease condition produced higher levels of all dependent variables than the control condition for the principles of nonmaleficence and beneficence. This reveals that aspects of ethical principles do alter human perceptions and preferences of AI autonomy.

4.2 Qualitative results

The fifteen interviews conducted in the study revealed interesting consensus and themes that helped answer the study's research questions. These results will be reported in this

section in terms of five themes, supported by the participants' own quotes.

Human experience with AI technology dictates how AI should adapt to ethical dilemmas

The first theme that arose in the interview data was the importance of experience in deciding how an AI teammate should adapt its autonomy in response to encountering an ethical principle. The first part of this theme is the experience of the AI itself with the ethical dilemma at hand.

> *"It gets to become autonomous when it's had enough experience under its wings, you know. I don't think it's gonna be able to be autonomous in the beginning, and I'm not sure if its measured in days, weeks, or months. But I just think that after a while it's gonna continue to learn from its previous interactions." (P7, Male, 75)*

Participant 7 discussed this in terms of learning, explaining that, because AI operates off stored data sets, the more previous dilemmas it has to base its decision on enables it to adapt to higher levels of autonomy in response to the dilemma. Thus, while AI adaptations to lower LOAs may be preferred initially, they may not after that agent has encountered the dilemma multiple times. Another participant discussed this in terms of the AI increasing its understanding of the dilemmas.*"I still think that level [of oversight] still needs to be there. In time maybe that won't be necessary. People will be more comfortable with it, or the technology will be able to pick up some of these things better. I means, I don't know if they're gonna put thousands of case studies into these things, and then, you know, it'll have a better understanding." (P1, Male, 70)*

This participant explained to the interviewer that he would want the AI to adapt to include some level of direct input when it encounters an ethical dilemma at first, but that as an AI's dataset includes more and more case studies, it will be better able to understand the ethical dilemmas it encounters and be able to respond with a decreased level of oversight. This participant also discussed the increased comfort that teammates would have with the AI responding to ethical dilemmas over time, a sentiment that leads into the second part of this theme, the experience levels of human teammates with the AI, an aspect that particularly concerned Participant 14:*"I think human error would actually influence my ability to trust the AI, because I could feel like the AI could be falsely sort of alerted that there's something bad happened, because a human made a mistake in their data entry, so, but that could happen with a human partner, too. So, it's kind of like hard to say. I think it's sort of level of experience." (P14, Female, 31)*

This participant felt that inexperienced human team members might provide the AI with false inputs or interpret its alerts incorrectly, inexperience that would cause the AI to make mistakes. They felt that this was similar to how new human team members need to be treated, where you would need them to move more slowly and consult the rest of the team more as they begin to assimilate. Another participant added that inexperienced team members may over-trust the AI's ability to respond.*"My fear with this is that if you get inexperienced people, that they would be not sure to question it, and something could happen that you don't want." (P1, Male, 70)*

Participant 1 felt that inexperienced team members would be unsure of when to question the AI's decisions, such as during a veto period, and that it may need to further decrease its autonomy to force human intervention. Where an unsure team member may have previously instigated a conversation about the right thing to do with another human colleague, the participant felt that inexperienced team members may skip that step and defer to the AI. Another participant elaborated on the dangers of giving too much autonomy to an AI teammate when the humans are inexperienced with how it operates.*"I work with a medical assistant, and she was charting O2 saturations of less than 90, and I said [assistant name] you can't document that...I wouldn't want the AI to increment a therapy like increased oxygen for that kind of data entry, but if the AI recognized that the O2 sat was documented as 88, to send an alert to a clinician, that or alert the medical assistant, that this is probably not a true data entry." (P4, Female, 65)*

Participant 4 explained that if inexperienced human team members didn't know what data the AI would read and respond to, they may inadvertently cause the AI to take action for an issue that doesn't actually exist. In cases like this, the participant would want the AI to reduce its autonomy such that it had to send alerts to someone on the team before taking action in order to recognize such false data entries. For instance, humans have a habit of talking out loud while pondering issues, and the AI may read this as positive information that it is supposed to consider.

The more routine the activities associated with the dilemma, the more the AI will be trusted to handle it with higher autonomy

The next major theme to emerge from the interviews was the consistency of the triggers and decisions with the team's normal, routine activities. This theme concerns both the AI's ability to confidently make decisions, as well as the team's established understanding of what the right decision is for that circumstance. One participant discussed this in terms of the ability to code the AI with known triggers:

> *"I think if you didn't have it flagged in the chart, that there was a pain medicine issue, you know, the AI wouldn't be like, oh this person looks like they might be doing drugs. Or if they had track marks, you know, that stuff would be hard to catch, like, a vibe would be hard to catch. So, like, there are explicit keys you can code it for, but not as much implicit." (P2, Female, 37)*

Participant 2 discussed here the differences between being able to code for explicit versus implicit cues. Situations that lend themselves to more explicit cues for which an AI can be coded to recognize enable it to more accurately recognize and appropriately respond to ethical dilemmas. It is far easier to code an AI agent to recognize and classify words or specific reading values than to understand changes in facial expressions or out-of-character behavior. Participant 12 provided an example of this in terms of an escalation sequence:*"Having buzz words, I guess, would be another way to kind of gauge that, you know, or tone of voice, volume of voice. Any sorts of that, because we do a lot of training on, like, what escalation looks like in a patient, you know, so I think that's something that an AI agent could probably pick up on, is when a patient is escalating." (P12, Female, 32)*

In this quote, the participant articulated that there are certain patterns that medical school teaches in order to guide the actions of medical practitioners in recognizing and responding to issues with patients. Such patterns, the participant explained, would enable an AI agent to understand a specific situation and how to respond. Beyond just the AI's ability to recognize and respond to these situations, it's also advantageous to the team for it to do so more autonomously. It is interesting that this participant references the escalation skills that medical professionals on which medical professionals repetitively train. This indicates that, even if the AI were to take over in most of these situations, human team members still need experience making those same decisions in order to develop their own skills.*"If the condition, schedule [the patient] with this group of providers, that sort of thing. So, we can get the bulk of the routine ones to be done automatically, it would free up the schedulers and the clinicians that need to have consideration." (P15, Male, 42)*

What the participant is discussing in this quote is that some patient cases are more routine and can be effectively broken down and handled by an AI agent, and if these teammates could do it with increased autonomy, it would free up human team members to assess and handle the more complicated issues. In sum, the common thread within the three quotes used to support this theme is that ethical dilemmas that can be recognized and analyzed according to an established pattern and/or process could be handled by an AI agent operating with higher levels of autonomy.

An AI's demonstrated ethical capability will impact the amount of desired human input

Along a similar line of thinking as the previous theme, the next theme involves a consideration of the AI teammate's proven capabilities and expertise. Interview participants consistently discussed the need for human teammates to be considered the experts when the ethical dilemma involved uniquely human factors. The participants thought it would be beneficial for an AI teammate to stop and decrease its autonomy level from its normal flow if the AI recognized that an expert was required to evaluate factors outside of its expertise. As far as what would be outside of the AI's expertise, most participants agreed that it would include uniquely human variables, such as emotions.

> *"If there comes, I guess, a step where there's a judgment call or other human factors needs to be considered, then it would make sense to definitely, you know, stop the flow, and then bring in the experts." (P15, Male, 42)*

> *"It depends on the nuances of the situations and the patient encounter as to how effective the AI can be without any oversight because there are just some of those, you know, call them human components, or, you know, emotional ethics, you know, emotional ethical decisions, or go with your gut feelings that I think is the missing component that you would have with the AI." (P8, Female, 53)*

Participant 8 specifically identified here that decisions that need to be based on or heavily consider human emotion would be outside the realm of an AI teammate's capabilities and thus require increased human oversight. These are components, the participant explained, that AI wouldn't learn over time and would always require human input. While participants identified these as situations where an AI teammate's lack of expertise would require decreases in its

autonomy, they also identified situations in which an AI's unique capabilities would encourage increases in its autonomy.*"I mean, if its got the ability to assimilate, you know, thousands of decisions made in this ethical dilemma, then it has the ability to make the decision, perhaps more quickly and accurately than a human who doesn't have that type of experience." (P11, Male, 60)*

This participant identified the situation where a new dilemma for the team would most likely be best handled by an AI teammate able to quickly acquire and analyze data from outside organizations in order to make the decision in a timely manner. In essence, the AI teammate would be the best suited to step in and fill a new capability gap in the team based on its ability to learn quickly. Another participant discussed this in terms of making an initial diagnosis of a patient.*"The ability of an AI to do the physical examinations would be very interesting, to have them have, like, a bank of photos or something in their system and be like, of, well this is what cyanosis looks like, or this is what level three muscle wasting looks like." (P9, Female, 29)*

Participant 9 explained that the process of collecting and examining symptoms and patient history and making an initial diagnosis could be a lengthy, uncertain process that could be better performed by an AI teammate with the ability to collect, analyze, and compare data more quickly and accurately than a human. In these instances, allowing the AI to complete the entire process uninterrupted would be beneficial to the patient and the entire team. In short, this theme identified that an AI teammate's unique expertise and capabilities should influence the way it reacts to encountering an ethical dilemma.

AI teammates need to consider the ethical constraints of third parties outside of their team

An important theme that emerged in the interviews is that the team's shared ethical code is not the only one that needs to be considered. Teams interact with a number of third parties in their daily activities, and these should also influence the design of the AI's adaptations. This theme was particularly relevant in considering the beneficence vignette involving a Jehovah's Witness.

"I did like the one about the Jehovah's witness, because that's something I could see, like, an AI downloading a cultural database, or like a religious database, that it would have that information when the patient says they don't want it." (P13, Female, 31)

"There's so, so many different religious and cultural components that different patients come in with, so it'd really, that would be kind of hard to program, I can imagine." (P12, Female, 32)

Here, the participant points out that a person's ethical code can be hard to determine and involves numerous factors, such as culture and religion, and questions the ability to code an AI to recognize and understand which ones apply to a specific situation. As with previous themes, participants also identified instances where an AI teammate might be better suited to consider third-party ethics than human teammates.*"I think sometimes people's own beliefs kind of can come in and gray those areas, and I think AI could potentially get rid of some of that. I see where it could be a benefit for patients, especially patients who are different or seen as different." (P5, Female, 44)*

Participant 5 relayed that even though there are ethics outside of the team sometimes involved, such as the personal beliefs of a patient, it can be difficult for humans to respect and consider them. Whether consciously or subconsciously, humans elevate their own ethics above those of others, something that AI could be programmed not to do. Participants recognized that AI would present the team with an opportunity to better consider the ethics and beliefs of third parties and take less biased actions.

The complexity of the ethical dilemma should influence to what autonomy level an AI teammate adapts

The following theme concerns the number of factors and straightforwardness of the ethical dilemma. In essence, participants conveyed that some dilemmas are harder to resolve based on their complexity. Participants discussed this in regards to an AI agent making a decision based on a few apparent black-and-white variables while disregarding other variables that should be considered in order to make an ethical decision that considers human factors, such as compassion. Participants also noted that some dilemmas contain multiple layers of ethics, and an AI teammate may need to be able to recognize when this complexity requires it to seek additional help in making a decision.

"I just think that sometimes they could be just a little more cut and dry, when there could be other circumstances that would be taken into consideration, to err on the side of being more compassionate." (P3, Female, 65)

"If the AI was able to identify that it was in a situation and making an autonomous decision, then maybe it should be allowed to proceed with and make the

decision on its own, but some of those other, the other principles of medical ethics require I think more complex decision making to determine what the proper ethical response would be. So, I think primarily the first internal prompt of the AI is 'Can I make this decision?'" (P10, Male, 29)

Participant 10 indicates here that different ethical principles have different levels of complexity, and those that are more complex should trigger the AI teammate to stop and gain input from its teammates. Participants noted that this goes both ways, where for an issue with less complexity, the AI should have increased autonomy to deal with it.*"I think it would change based off the type of dilemma it encounters. Something more straightforward, I would give it more autonomy. Something less straightforward, I would want to have veto power...I like the adaptive idea." (P9, Female, 29)*

Participant 9 noted that she would want the AI teammate's autonomy level to change based on how "straightforward" the ethical dilemma was to the AI and expressed excitement over the idea that an AI teammate could adapt its autonomy in response to this straightforwardness. Beyond just how clear a dilemma is to the AI, participants added that the number of factors involved should affect the autonomy of the AI. In essence, these participants collectively identified the need for AI teammates to be able to recognize and adapt their autonomy levels in response to the complexity of the ethical dilemma.

AI teammate adaptation can be used to reinforce a team's codified ethical principles.

The final major theme that emerged from the interview data related to the research questions is the importance that codified ethical principles play in how an AI teammate should respond to ethical dilemmas. Participants repeatedly noted that professions, particularly trusted professions like healthcare, are expected to follow the laws and policies that have been put in place to help organizations abide by the profession's code of ethics.

"I think the mutual understanding is, you know, that definitely the HIPAA think is for the privacy and safety for everybody, and that's clear across the board for everybody." (P6, Female, 37)

"When HIPPA came out that became a cornerstone in our interactions with patients. So, we're very cognizant of being careful not to share information with individuals that don't have any right to that information." (P7, Male, 75)

Participants 6 and 7, and multiple others, specifically mentioned the privacy and safety policies that protect patients, such as HIPAA. These regulations are made clear to all workers in the sector and act as prescriptive mechanisms to guide ethical decisions. These codifications not only enable AI to understand what the ethical action is but, some participants thought, would enable the AI to help the rest of the team abide by the standards.*"What it forces, if the AI was making decisions, it forces everyone basically to the same standard" (P10, Male, 29)*

Participant 10 told the interviewer that if, in ethical dilemmas where the decision should be guided by an explicit policy, the AI went ahead and made it for the team, it would force the team to abide by the standard and better enforce the norm for the whole organization. Another participant supported this idea of increased AI autonomy when the ethical decision is clearly dictated.*"I don't think you can have it both ways. If you're looking to speed up the decision process, then you have to trust the AI, you know, within that framework, of whatever the ethical construct is made, and to act within that." (P11, Male, 60)*

As Participant 11 put it, the more robust the ethical framework with which the AI is programmed, the more it should be trusted to make decisions based on that framework. An important point that this participant made is that if designers are spending the time to code AI with the strongest, most important aspects of a team's ethical code, it should be trusted to operate in accordance with that code.

Summary of qualitative results

The five themes discussed in this results section focus on both the direct capabilities of an AI teammate as well as the complexity of the ethical dilemma at hand. Most importantly, the quotes from the study participants above are a small sample of the qualitative data that the researchers used to arrive at these themes, and it should be noted that there were numerous quotations to choose from in presenting the results. The mass consensus in these themes reflects the significance shown in the quantitative data, a phenomenon that will be further explored in the next section. More importantly, these themes reveal that, in regards to RQ1, humans generally prefer AI teammates lower their autonomy in response to ethical dilemmas; however, in regards to RQ2, this preference is influenced by the ethical principles at stake.

5 Discussion

This study provided ample data and insights for answering the study's research questions. In regards to RQ1, how an AI teammate should adapt its autonomy in response to an ethical dilemma, results show that, in most cases, AI teammates should decrease their autonomy levels and seek input from human teammates. Results also show that, in regards to RQ2 concerning how AI adaptations should be affected by the type of ethical dilemma, the exceptions to these cases include when the ethical principle is codified by law and/or policy, and instances where AI may actually enhance the ethicality of the team by forcing compliance to known ethical standards. This section will further explain how the results answered the research questions and the design implications of those answers.

5.1 The desire for decreased autonomy when faced with ethical dilemmas

The first question that this study asked concerns how an AI teammate should react when it encounters an ethical dilemma, particularly in regards to how autonomous it should be in responding to the ethical dilemma. The resounding answer from the quantitative results seemed to be that when an AI teammate recognizes the existence of an ethical dilemma, it should reduce its autonomy levels to increase the amount of human input required for it to make and execute a decision. Study participants reported higher levels of trust, perceived competence, perceived ethicality, and preference as a teammate for AI teammates that responded to an ethical dilemma by adapting to a lower autonomy level than their normal operating autonomy. This makes sense in the context of healthcare, where the question of who is responsible for the consequences of an ethical decision is high, and the movement for meaningful human control is strong [22]. What became interesting in the qualitative data was that this sentiment may only stand for the initial fielding phase of the AI teammate.

Experience with an AI teammate was a key factor that influenced participants' feelings toward how an AI teammate should adapt in response to an ethical dilemma. Participants expressed this not only in terms of personal comfort but also in acknowledging the ways that AI teammates make decisions- based on stored data. Participants noted that, as an AI teammate observed the right ethical decisions being made and incorporated them into its decision-making logic, it would become more accurate and trustworthy in making those decisions autonomously. This indicates that how an AI teammate adapts its autonomy when it encounters an ethical dilemma may change over time and consider the confidence the AI and its team have in its ability to make the decision based on experience. This finding is supported by recent HAT research that found human acceptance of an AI teammate can be increased by highlighting its capabilities and experience to accomplish its tasks [11]. This is also supported by one of the other themes concerning how close the triggers are to the team's routine operations. When the triggers align with data that the AI has stored from a multitude of observations, the AI is more capable of making the connection between the ethical dilemma and the ethical resolution of that dilemma. In these instances, increasing its autonomy level would increase its utility to the team, because, as one of our participants quoted above noted, it would free the team up to handle the more complex issues.

Time, however, cannot necessarily make an AI agent more receptive or more human. A major theme from the interviews that explains the strong preference of healthcare professionals to have an AI teammate that lowers its autonomy in response to ethical dilemmas where it is normal to have debate over the ethically right action. Participants were particularly concerned over how an AI teammate might understand that aspects of a specific patient's culture, for example, may gray the situation and require consultation with other members of the healthcare team. Worse, participants expressed concern that some members of the team may over-rely on the AI agent's black-and-white perception of the situation and not intervene as they might if a human had made the same decision. This concern primarily on the need to consider and discuss the "human elements" relevant to an ethical dilemma, such as expressed emotions and body language. Many of the participants spent a lot of time in the interviews trying to grasp how an AI agent could make an ethical decision when so many of the relevant variables were distinctly human factors. In fact, just how an AI teammate would try to sense and conceptualize human emotions is full of additional ethical questions [60]. The study participants indicated it would be overall more ethical for the AI to be designed to be more autonomous when the decisions involve more "absolute" values, and less autonomous when the variables were more human in nature. This is interesting, as such increases in autonomy were negatively received in the survey data. Such a discrepancy has been discussed in the literature as an "alignment problem" between what values humans can try to code into AI and what values will always be perceived as uniquely human and not programmable [16]. This leads to the second research question, which sought to understand how an AI adapts in response to an ethical dilemma is different based on the ethical principles involved.

5.2 AI cannot be programmed to adapt to all ethical dilemmas in the same way

Not only was the autonomy condition significant for all measured dependent variables, but so was the within condition for the ethical principle type. Most notable was that the principle of beneficence was markedly different from the other types, where it caused lower perceptions on behalf of the participants. The interviews provided valuable insight into this, as participants explained that what is in the patient's best interest isn't always clear and may require a significant amount of reasoning on the decision-maker's part. Medical philosophy has debated this issue for years, citing the tug of war between beneficence and other principles such as justice and autonomy [25]. This tug of war adds to the complexity of the ethical dilemma, something that participants viewed as a key contributor to how an AI teammate should adapt in response to an ethical dilemma. The more factors involved, the harder it would be to code an AI teammate with the logic it needs to make a decision.

One additional such factor, and one that is particularly important in determining what is in a person's best interest, is that of third-party ethics. What may be in the person's best interest according to the AI's programmed code of ethics may not match that person's ethics; thus, there is an additional need for the AI to attempt to understand and consider those ethics as well. Participants in the interview noted that if an AI could do something such as download a cultural database, it would be better able to handle these ethical dilemmas with higher levels of autonomy. Still, participants expressed concerns over an AI teammate's ability to understand whose values to prioritize in a given situation. In such a situation, participants explained, it is important that teams have an open discourse on the best resolution to the dilemma. This is supported by research that has emphasized the importance of increasing dialogue between medical team members when ethical principles are complex and involve significant considerations of what is best for the patient [40]. This explains why the default adaptation would still be for the AI to decrease its autonomy level in order to understand what aspects of both its and the third party's ethical codes should take precedence.

In contrast, the justice and autonomy conditions did not elicit the same levels of preference for decreasing autonomy. Interview data may explain this in terms of how clear and codified these principles are relative to the others. Patient autonomy is most often ensured in healthcare, just like in research, through informed consent [47]. The key aspect of this process is that the patient is actually informed, a process that interview participants thought would actually be enhanced through the use of an AI agent that can access and provide information much more quickly and accurately than a human. Justice, participants also explained, is often codified into policy such that the ethical answer to an ethical dilemma is not up for discussion; for this specific principle, participants thought AI that responded to ethical dilemmas with *increased* autonomy would enhance the overall ethicality of the team, which is often swayed by unconscious human bias and prone to human error. The latter was a particular focus of the interview participants, who spent hours a year studying ethics and compliance policies. To them, an AI teammate would be better able to make decisions when it was trained on the same materials.

5.3 Design recommendations for AI teammates that adapt to respond to ethical dilemmas

Overall, what the significance of the ethical principle condition shows is that AI teammates need to possess the ability to alter their adaptation behavior as it is unrealistic to predict exactly what ethical dilemmas an AI agent will encounter when operating as part of a real-world human-AI team. While the majority of ethical dilemmas will require an AI teammate to decrease its autonomy to further involve a human teammate, this is not the case for every dilemma, and ethical principle type may be a way to predict and program its optimal adaptation. This possibility drives the following design recommendations.

Recommendation 1: AI teammates should be designed to consider multiple ethical codes, including those of their team and relevant third parties

The first recommendation is that AI teammates should be programmed to recognize specific triggers of ethical dilemmas and the ethical principles to which they are tied. This includes both the ethical code of its teams and the codes of the third parties with whom the team interacts. Functionally, this means that an AI teammate should be programmed to download and update ethical rules from its team, as well as have the adaptive capability to recognize the need to download and consult new ethical rules regarding third parties. For instance, to draw upon this study's survey, if an AI teammate triggers that the patient has identified as a Jehovah's Witness, consult if it knows relevant ethical rules of that religion, and, if not, seek to obtain them before making any decisions regarding that patient's care. Regarding the team's ethics itself, the AI should routinely have the ability to update its ethical triggers and codes, as ethical codes are not static, and the AI needs to be capable of updating its rules and responses over time.

Recommendation 2: AI teammates should adapt to lower autonomy levels when confronted with an ethical dilemma unless a clear rule dictates the ethical response

The logic that an AI agent would use to make a decision once an ethical dilemma is identified would involve both

the explicit rules with which the AI is programmed and the learned data that it has stored. The more explicit and reinforced these rules are in the AI teammate's decision logic should influence how it adapts its autonomy in response to making a decision. For instance, if the rule is codified according to an organizational policy, it might increase its autonomy, because there is only one right answer in response to this policy. Or, if the AI is responding to a common ethical dilemma that the AI has responded to many times before, it may also increase its autonomy. In the absence of these items, as reflected in the significant preference for decreased autonomy in this study, AI teammates should always adapt to lower levels of autonomy. Practically, this may look like a variable identifier or counter that notes if a response is based on a hard-coded rule or has been properly reinforced enough times to warrant an increase in autonomy or been reinforced so little it requires a decrease in autonomy.

5.4 Limitations and future work

This study has several limitations that should be noted. First, this study involved only one human-AI teaming context, notably one that exists in a high-risk environment. While this was a conscious choice due to the increased role of ethics in the environment, the degree to which decreased autonomy was likely preferred was somewhat affected by the overall risk level of the AI's decisions, and other lower-risk teaming contexts should also be studied and considered. It is also important to note that the vignettes chosen for the medical context are based on case studies used to teach the principles of medical ethics to healthcare professionals and, as such, are oversimplified. In practice, healthcare professionals would also consider other competing personal values and interests, as well as the opinions of other members of the healthcare team. Next, regarding the interviews, while the fifteen participants were fairly diverse from a gender and age perspective, all participants live in the continental United States; thus, their concept of ethics is heavily influenced by not only American culture but also the medical ethics specifically taught in medical schools in the United States. Finally, because the quantitative results were based on hypothetical situations in a survey, it should be noted that participants may have had slightly different perceptions had they experienced these vignettes in a real-world environment with increased ecological validity.

Beyond the above-noted limitations, this research revealed many exciting areas for future research on ethical AI teammates. First, it is interesting that many participants noted that people seldom agree on ethics and that different teams are more or less ethical. How an AI teammate should act and adapt in response to these different teaming environments is an important consideration due to the possibility of AI teammates learning not only ethical but also unethical behavior from their teammates. As noted in the themes, participants expressed optimism about AI teammates being used to better enforce human compliance with established ethics. It was interesting to find that many participants believed there are situations where a human is more likely to make the wrong decision due to factors that impair their logical decision-making skills, such as stress and emotion. While our findings show the need to consider human factors generally increases a desire for decreased autonomy, there seemingly may be situations where those human factors should actually be removed. Future research should focus on whether AI teammates should utilize adaptations in their autonomy levels in order to act as this compliance mechanism and how such adaptations would affect perceptions of it as trustworthy and competent by its human teammates.

6 Conclusion

As AI becomes a bigger part of society and teaming, it is imperative that it is designed to operate ethically. An essential component of that is programming the AI to determine when and to what extent it has the ability to understand and make a decision when there is an ethical dilemma. In this study, we asked fifteen medical professionals to consider how the principles of medical ethics should affect AI autonomy. The themes from these interviews allowed us to produce actionable recommendations for the HCI and AI communities for the design of ethical AI teammates.

Data availability Data will not be made publicly available due to privacy concerns. De-identified portions of the data can be made available by making a reasonable request to the corresponding first author.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no Conflict of interest.

References

1. Baines, P.: Medical ethics for children: applying the four principles to paediatrics. J. Med. Ethics **34**(3), 141–145 (2008)
2. Barber, K. S.: Dynamic adaptive autonomy in agent-based systems. In: Proceedings. Fourth International Symposium on Autonomous Decentralized Systems.-Integration of Heterogeneous Systems-, pages 402–405. IEEE (1999)
3. Beauchamp, T. L. et al.: The belmont report. The Oxford textbook of clinical research ethics, pages 149–155 (2008)
4. Botes, A.: An integrated approach to ethical decision-making in the health team. J. Adv. Nurs. **32**(5), 1076–1082 (2000)
5. Chien, S.-Y., Lewis, M., Sycara, K., Liu, J.-S., Kumru, A.: Influence of cultural factors in dynamic trust in automation. In: 2016

IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 002884–002889. IEEE (2016)
6. Christakis, D.A., Feudtner, C.: Ethics in a short white coat: the ethical dilemmas that medical students confront. Acad. Med. **68**(4), 249–54 (1993)
7. Davidson, J.B., Graham, R.B., Beck, S., Marler, R.T., Fischer, S.L.: Improving human-in-the-loop simulation to optimize soldier-systems integration. Appl. Ergon. **90**, 103267 (2021)
8. De Cremer, D., Kasparov, G.: The ethical ai-paradox: why better technology needs more and not less human responsibility. AI and Ethics **2**(1), 1–4 (2022)
9. Eitel-Porter, R.: Beyond the promise: implementing ethical ai. AI and Ethics **1**(1), 73–80 (2021)
10. Faubion, D.: Ethical dilemmas in nursing. Nursing Process (2024)
11. Flathmann, C., Schelble, B. G., McNeese, N. J., Knijnenburg, B., Gramopadhye, A. K., Chalil Madathil, K.: The purposeful presentation of AI teammates: Impacts on human acceptance and perception. International Journal of Human–Computer Interaction, pages 1–18 (2023)
12. Flathmann, C., Schelble, B.G., Rosopa, P.J., McNeese, N.J., Mallick, R., Madathil, K.C.: Examining the impact of varying levels of AI teammate influence on human-ai teams. Int. J. Hum Comput Stud. **177**, 103061 (2023)
13. Flathmann, C., Schelble, B. G., Zhang, R., McNeese, N. J.: Modeling and guiding the creation of ethical human-ai teams. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pages 469–479 (2021)
14. Formosa, P.: Robot autonomy vs. human autonomy: social robots, artificial intelligence (AI), and the nature of autonomy. Minds and Mach. **31**(4), 595–616 (2021)
15. Fukuda-Parr, S., Gibbons, E.: Emerging consensus on 'ethical ai': human rights critique of stakeholder guidelines. Global Pol. **12**, 32–44 (2021)
16. Gabriel, I.: Artificial intelligence, values, and alignment. Mind. Mach. **30**(3), 411–437 (2020)
17. Hanna, N., Richards, D., et al.: The impact of virtual agent personality on a shared mental model with humans during collaboration. In Aamas, pages 1777–1778 (2015)
18. Hauptman, A. I., McNeese, N. J.: Overcoming the lumberjack effect through adaptive autonomy. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 66, pages 1075–1079. SAGE Publications Sage CA: Los Angeles, CA (2022)
19. Hauptman, A. I., Schelble, B. G., Duan, W., Flathmann, C., McNeese, N. J.: Understanding the influence of ai autonomy on AI explainability levels in human-ai teams using a mixed methods approach. Cognition, Technology & Work, pages 1–21 (2024)
20. Hauptman, A.I., Schelble, B.G., McNeese, N.J., Madathil, K.C.: Adapt and overcome: perceptions of adaptive autonomous agents for human-AI teaming. Comput. Hum. Behav. **138**, 107451 (2022)
21. Hauptman, A.I., Schelble, B.G., McNeese, N.J., Madathil, K.C.: Adapt and overcome: Perceptions of adaptive autonomous agents for human-ai teaming. Comput. Hum. Behav. **138**, 107451 (2023)
22. Hille, E. M., Hummel, P., Braun, M.: Meaningful human control over AI for health? a review. Journal of Medical Ethics (2023)
23. Hunt, T.G., Jennings, D.F.: Ethics and performance: a simulation analysis of team decision making. J. Bus. Ethics **16**, 195–203 (1997)
24. Jacobs, J.P., Wernovsky, G., Cooper, D.S., Karl, T.R.: Principles of shared decision-making within teams. Cardiol. Young **25**(8), 1631–1636 (2015)
25. Jansen, L.A.: Between beneficence and justice: the ethics of stewardship in medicine. J. Med. Philos. **38**(1), 50–63 (2013)
26. Johnson, M., Bradshaw, J.M., Feltovich, P., Jonker, C., Van Riemsdijk, B., Sierhuis, M.: Autonomy and interdependence in human-agent-robot teams. IEEE Intell. Syst. **27**(2), 43–51 (2012)
27. Kieslich, K., Keller, B., Starke, C.: Ai-ethics by design. evaluating public perception on the importance of ethical design principles of ai. arXiv preprint arXiv:2106.00326 (2021)
28. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**(1), 50–80 (2004)
29. Lewis, M., Sycara, K., Payne, T.: Agent roles in human teams. AAMAS Workshop on Humans and Multi-Agent Systems (2003)
30. Liang, J., Huang, W., Xia, F., Xu, P., Hausman, K., Ichter, B., Florence, P., Zeng, A.: Code as policies: Language model programs for embodied control. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), pages 9493–9500. IEEE (2023)
31. Lumineau, F.: How contracts influence trust and distrust. J. Manag. **43**(5), 1553–1577 (2017)
32. Macdonald, J. P., Mallick, R., Wollaber, A. B., Peña, J. D., McNeese, N., Siu, H. C.: Language, camera, autonomy! prompt-engineered robot control for rapidly evolving deployment. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, pages 717–721 (2024)
33. Marangunić, N., Granić, A.: Technology acceptance model: a literature review from 1986 to 2013. Univ. Access Inf. Soc. **14**, 81–95 (2015)
34. McGee, E. T., McGregor, J. D.: Using dynamic adaptive systems in safety-critical domains. In: Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, pages 115–121 (2016)
35. McLennan, S., Fiske, A., Tigard, D., Müller, R., Haddadin, S., Buyx, A.: Embedded ethics: a proposal for integrating ethics into the development of medical AI. BMC Med. Ethics **23**(1), 6 (2022)
36. McNeese, M. D., McNeese, N. J.: Humans interacting with intelligent machines: At the crossroads of symbiotic teamwork. In Living with robots, pages 165–197. Elsevier (2020)
37. Merritt, S. M., Heimbaugh, H., LaChapell, J., Lee, D.: I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System Human Factors, 55(3):520–534. Publisher: SAGE Publications Inc (2013)
38. Nilstun, T., Cuttini, M., Saracci, R.: Teaching medical ethics to experienced staff: participants, teachers and method. J. Med. Ethics **27**(6), 409–412 (2001)
39. Nomura, T., Kanda, T., Suzuki, T., Kato, K.: Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots. In: RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759), pages 35–40. IEEE (2004)
40. Oberle, K., Hughes, D.: Doctors' and nurses' perceptions of ethical problems in end-of-life decisions. J. Adv. Nurs. **33**(6), 707–715 (2001)
41. Onağ, Z., Tepeci, M.: Team effectiveness in sport teams: the effects of team cohesion, intra team communication and team norms on team member satisfaction and intent to remain. Procedia Soc. Behav. Sci. **150**, 420–428 (2014)
42. O'Neill, T.A., Flathmann, C., McNeese, N.J., Salas, E.: Human-autonomy teaming: need for a guiding team-based framework? Comput. Hum. Behav. **146**, 107762 (2023)
43. O'Neill, T., McNeese, N., Barron, A., Schelble, B.: Human–autonomy teaming: a review and analysis of the empirical literature. Human Factors, page 0018720820960865 (2020)
44. O'Neill, T., McNeese, N., Barron, A., Schelble, B.: Human-autonomy teaming: a review and analysis of the empirical literature. Hum. Factors **64**(5), 904–938 (2022)

45. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. A Syst. Hum. **30**(3), 286–297 (2000)
46. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. A Syst. Hum. **30**(3), 286–297 (2000)
47. Paterick, T. J., Carson, G. V., Allen, M. C., Paterick, T. E.: Medical informed consent: general considerations for physicians. In: Mayo Clinic Proceedings, volume 83, pages 313–319. Elsevier (2008)
48. Petersen, A.M., Pavlidis, I., Semendeferi, I.: A quantitative perspective on ethics in large team science. Sci. Eng. Ethics **20**, 923–945 (2014)
49. Pflanzer, M., Traylor, Z., Lyons, J.B., Dubljević, V., Nam, C.S.: Ethics in human-AI teaming: principles and perspectives. AI and Ethics **3**(3), 917–935 (2023)
50. Rahimi, H., Baharlooei, F.: The effect of ethical climate on trust in teamwork with the meditating role of ethical behavior. Organ. Behav. Stud. Q. **7**(2), 129–158 (2018)
51. Raven, M.E., Flanders, A.: Using contextual inquiry to learn about your audiences. ACM SIGDOC Asterisk J. Comput. Doc. **20**(1), 1–13 (1996)
52. Reidenbach, R.E., Robin, D.P.: Some initial steps toward improving the measurement of ethical evaluations of marketing activities. J. Bus. Ethics **7**, 871–879 (1988)
53. Rouse, W.B., Rouse, S.H.: A framework for research on adaptive decision aids. Technical report, ALPHATECH INC BURLINGTON MA (1983)
54. Salikutluk, V., Schöpper, J., Herbert, F., Scheuermann, K., Frodl, E., Balfanz, D., Jäkel, F., Koert, D.: An evaluation of situational autonomy for human-ai collaboration in a shared workspace setting. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pages 1–17 (2024)
55. Schelble, B. G., Flathmann, C., McNeese, N.: Towards meaningfully integrating human-autonomy teaming in applied settings. In: Proceedings of the 8th International Conference on Human-Agent Interaction, pages 149–156 (2020)
56. Schelble, B.G., Flathmann, C., McNeese, N.J., Freeman, G., Mallick, R.: Let's think together! assessing shared mental models, performance, and trust in human-agent teams. Proc. ACM Hum-Comput. Interact. **6**(GROUP), 1–29 (2022)
57. Schelble, B. G., Lancaster, C., Duan, W., Mallick, R., McNeese, N. J., Lopez, J.: The effect of ai teammate ethicality on trust outcomes and individual performance in human-AI teams. In HICSS, pages 322–331 (2023)
58. Schelble, B. G., Lopez, J., Textor, C., Zhang, R., McNeese, N. J., Pak, R., Freeman, G.: Towards ethical AI: Empirically investigating dimensions of AI ethics, trust repair, and performance in human-ai teaming. Hum. Factors, page 00187208221116952 (2022)
59. Schelble, B.G., Lopez, J., Textor, C., Zhang, R., McNeese, N.J., Pak, R., Freeman, G.: Towards ethical AI: empirically investigating dimensions of AI ethics, trust repair, and performance in human-AI teaming. Hum. Factors **66**(4), 1037–1055 (2024)
60. Stark, L., Hoey, J.: The ethics of emotion in artificial intelligence systems. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pages 782–793 (2021)
61. Textor, C., Zhang, R., Lopez, J., Schelble, B.G., McNeese, N.J., Freeman, G., Pak, R., Tossell, C., de Visser, E.J.: Exploring the relationship between ethics and trust in human-artificial intelligence teaming: a mixed methods approach. J. Cognit. Eng. Decis. Mak. **16**(4), 252–281 (2022)
62. Toupchian, A., Sarbakhsh, P., Ghaffari, R., Kazemi, A., Mahmoodi, H., Shaghaghi, A.: Development and psychometric analysis of the measure of perceived adherence to the principles of medical ethics in clinical educational settings: trainee version (pamethic-clin-t). Patient preference and adherence, pages 1615–1621 (2020)
63. Varkey, B.: Principles of clinical ethics and their application to practice. Med. Princ. Pract. **30**(1), 17–28 (2021)
64. Vorobeva, D., El Fassi, Y., Costa Pinto, D., Hildebrand, D., Herter, M.M., Mattila, A.S.: Thinking skills don't protect service workers from replacement by artificial intelligence. J. Serv. Res. **25**(4), 601–613 (2022)
65. Wang, W., Benbasat, I.: Recommendation agents for electronic commerce: effects of explanation facilities on trusting beliefs. J. Manag. Inf. Syst. **23**(4), 217–246 (2007)
66. Wickens, C. D., Li, H., Santamaria, A., Sebok, A., Sarter, N. B.: Stages and levels of automation: An integrated meta-analysis. In: Proceedings of the human factors and ergonomics society annual meeting, volume 54, pages 389–393. Sage Publications Sage CA: Los Angeles, CA (2010)
67. Woodruff, A., Anderson, Y. A., Armstrong, K. J., Gkiza, M., Jennings, J., Moessner, C., Viegas, F., Wattenberg, M., Wrede, F., Kelley, P. G., et al.: " a cold, technical decision-maker": Can ai provide explainability, negotiability, and humanity? arXiv preprint arXiv:2012.00874 (2020)
68. Wright, J.L., Chen, J.Y., Quinn, S.A., Barnes, M.J.: The effects of level of autonomy on human-agent teaming for multi-robot control and local security maintenance. Technical report, Army Research Lab Aberdeen Proving Ground MD (2013)
69. Wright, M.C., Kaber, D.B.: Effects of automation of information-processing functions on teamwork. Hum. Factors **47**(1), 50–66 (2005)
70. Zhang, R., Duan, W., Flathmann, C., McNeese, N., Freeman, G., Williams, A.: Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork. Proc. ACM Hum. Comput. Interact. **7**(CSCW2), 1–31 (2023)
71. Zhang, R., McNeese, N.J., Freeman, G., Musick, G.: "an ideal human" expectations of ai teammates in human-ai teaming. Proc. ACM Hum. Comput. Interact. **4**(CSCW3), 1–25 (2021)
72. Zhao, J., Sun, W., Zhang, S., Zhu, X.: How ceo ethical leadership influences top management team creativity: evidence from china. Front. Psychol. **11**, 517466 (2020)

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com