UNIVERSITY OF OKLAHOMA GRADUATE COLLEGE

EXPLORING THE ANTECEDENTS AND OUTCOMES OF THE PERCEIVED ETHICALITY OF ARTIFICIAL INTELLIGENCE IN THE WORKPLACE

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment to the requirements for the Degree of

MASTER OF SCIENCE

By RACHEL DETHERAGE

Norman, Oklahoma

2025

EXPLORING THE ANTECEDENTS AND OUTCOMES OF THE PERCEIVED ETHICALITY OF ARTIFICIAL INTELLIGENCE IN THE WORKPLACE

A THESIS APPROVED FOR THE

DEPARTMENT OF PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Shane Connelly, Chair

Dr. Eric Day

Dr. Adam Feltz

Table of Contents

Abstract

As the implementation of AI within organizations increases, concerns have emerged about the ethicality of AI-driven systems and tools and the ways people use them. Although previous literature has proposed numerous theories regarding how to develop and use AI ethically, the question remains of how individuals perceive the ethicality of AI and the effects this has on their engagement with AI tools. The study aims to explore whether and how certain perceptions of AI tool attributes, including explainability of AI (xAI) and moral intensity dimensions (i.e., magnitude of consequences and proximity), influence perceptions of AI ethicality. Additionally, the present study investigates the outcomes of perceptions of the ethicality of AI on the likelihood to recommend AI tools and functionality trust in AI. Analyses of covariance and mediation analysis evaluated the proposed model alongside qualitative supplemental analyses. Perceived ethicality of the AI tool significantly mediated the relationship between xAI and likelihood to recommend the AI tool, whereas functionality trust in the AI tool significantly mediated the relationship between proximity and the likelihood to recommend the AI tool. The results produce implications relevant to how individuals perceive AI within organizations and behavioral intentions for using AI tools.

Exploring the Antecedents and Outcomes of the Perceived Ethicality of Artificial Intelligence

The growth of artificial intelligence (AI) in organizational spaces highlights how quickly the workplace can change and adapt to innovation. As new technologies surface, organizations use AI to transform commonplace work practices to increase the accuracy and efficiency of various tasks and processes. Given such innovations, implementing AI in the workplace necessitates multiple standards to ensure its effectiveness, such as technological, legal, and ethical standards (Kelley 2022). Ethical standards for AI within an organization safeguard employees, collaborators, customers, the organization, and other key stakeholders by requiring an AI tool to meet standards for implementation, but such standards do not account for how an individual perceives the AI tool as ethical (Khan et al., 2022; Mittelstadt, 2019). Individual perceptions of organizational practices influence not only the practical utility of such practices but also broader views of the organization. For example, the perceptions of the ethicality of AI tools used in hiring practices impact organizational trust and organizational attractiveness of potential job seekers (Figueroa-Armijos et al., 2022; da Motta Veiga et al., 2023).

Artificial Intelligence (AI) refers to smart technologies with advanced capabilities associated with the capabilities generally performed by the human mind (Rai et al., 2019). As conceptualizations and practical applications of AI technologies continue to grow, questions arise about the ethical implications of using AI for predictions, recommendations, and the decisions stemming from its calculations. For example, racial and gender biases have been observed in AI-driven facial recognition software (Coe & Atay, 2021; Lohr, 2022). Although the development of ethical AI allows systems to mitigate biases and other ethically problematic attributes, ambiguity remains about how the emerging workforce discerns what constitutes

ethical AI. Situational characteristics, such as whether an AI tool exhibited intentionality, have influenced perceptions of blame when a decision causes harm (Sullivan & Wamba, 2022). The current study aims to understand the direct effects of situational factors relevant to prior ethical decision-making research, such as those proposed in Jones' (1991) moral intensity theory, in conjunction with perceptions of AI ethicality. Additionally, this study explores how these factors interact with each other to influence perceptions of AI ethicality, functionality trust in AI, and the behavioral intention to use an AI tool.

The study employs a scenario-based, randomized controlled experiment to investigate the perceived ethicality of AI tools within an organization, enabling a low-fidelity simulation and test of causal relationships (Aguinis & Bradley, 2014). Common methodological issues in scenarios include a lack of standardization, realism, and content validity, so the current study aims to proactively address these issues before beginning data collection (Yu et al., 2024). Scenarios will receive careful consideration for between-scenario homogeneity to account for a decreased fixed effect fallacy (Baguley et al., 2022).

The proposed model examines the situational antecedents of the perceived ethicality of AI and the outcomes of functionality trust in AI and the intention to use an AI tool's predictions or recommendations. This will advance current knowledge about AI in the workplace by considering how various situational factors influence ethical perceptions of AI, as well as how those perceptions relate to differential outcomes. Through its contribution to the growing research of AI and ethics within organizations, the model intends to advance practical applications for the implementation of AI tools in the workplace.

The Current State of AI Ethicality

Current research examining ethical AI focuses heavily on the development of ethical AI. As defined by Bleidorn et al (2018), machine learning refers to "the study and construction of algorithms that can learn from data and make predictions based on data without being programmed to perform specific tasks." The current line of research examines AI in a machine-learning sense, as the growing landscape of machine learning tools and development speaks to its prevalence. For example, the 'Human-in-the-Loop' system refers to a purposeful design implementing interaction between humans and AI during the AI data annotation period to improve AI machine learning in an ethical context. That process looks something like this. A set of training data consists of characteristics of an ethical situation with the ability to receive a quantitative annotation. After the identification of a scenario, a human annotates the same data as an AI tool by assigning a numerical value to each characteristic to depict its ethicality. The AI and human annotations undergo comparison, and the AI receives reinforcement through feedback. The AI continuously works with the human annotator until the AI produces annotations of the scenario that align with real-world context (i.e., human judgments). This trains the AI to include ethical viewpoints in its decision-making calculus by incorporating a human viewpoint (Chen et al., 2023). The development of ethical AI addresses principal concerns essential to the implementation of AI into organizational settings.

Common concerns regarding AI include transparency, privacy, and fairness. In the context of AI, transparency exhibits how and why AI makes its decisions, whereas privacy refers to AI's collection and regulation of personal data, and fairness contributes to AI's ability to disregard discriminatory biases in its processes (Khan et al., 2022). Ethical AI requires general contextual consideration beyond objective scientific principles, such as programming and training guidelines. Although the processes informing the development of AI tools are crucial to

implementing ethical AI, there is a notable lack of research regarding perceptions of ethical AI, leaving gaps in understanding human choices for how and when they will trust and use AI tools. AI undergoes development for the purpose of implementation, and ideally, AI tools deemed as ethically appropriate for usage are more likely to be implemented and used. Therefore, it is important to understand factors influencing users' ethical perceptions of such tools.

The Perceived Ethicality of AI

AI developers can achieve AI development consistent with ethical standards, but this does not necessarily translate to how end users of the AI system perceive its ethicality. The perceived ethicality of AI refers to how users evaluate an AI tool's alignment with widely accepted ethical standards for AI tools. Although the perceived ethicality of AI could encompass a broad range of ethical principles, the current study utilizes the ethical AI framework developed by Jang et al. (2022). The perceived ethicality of AI includes the dimensions of transparency, privacy, fairness, responsibility, and non-maleficence. Transparency refers to the ability of AI to disclose its technical procedures (Jang et al. 2022). Fairness encompasses the maintenance of an equitable and reasonable distribution of benefits and expenses without undue bias (European Commission, 2019). Privacy necessitates the ethical handling of data provided by the user and data generated about the user during their interaction with the system (European Commission, 2019). Responsibility requires the establishment of an accountable party for AI systems and their outputs during development, deployment, and usage (European Commission, 2019). Non-maleficence advocates for the development, implementation, and use of AI in a manner that avoids human harm (Floridi et al., 2018). The dimensions of the perceived ethicality of AI reference AI's development, implementation, and usage, therefore theoretically connecting to its influence on how individuals could make decisions about whether to trust and/or use the AI.

Perceived Ethicality of AI in an Ethical Decision-Making Framework

Jones (1991) utilizes the basis of Rest's (1986) model of ethical decision-making to describe how situational factors influence ethical decisions. According to Rest's (1986) model, ethical decision-making unfolds in four stages: moral judgment, moral judgment, moral intent, and moral behavior. The process begins with moral awareness, where an actor recognizes a moral issue, leading to a moral judgment, where the actor deciphers the moral correctness of an action. Moral intent then occurs, where the actor establishes the intent to act on their action, resulting in an execution of the action during moral behavior (Jones, 1991; Rest, 1986). As with any decision, a decision about whether to use or trust AI would require an individual to partake in the ethical decision-making process. Situational factors, as described by Jones (1991), can influence each stage of the ethical decision-making process, which is supported by existing literature examining the effects of situational influences on the ethical decision-making process (Lincoln & Holmes, 2011; Barnett & Valentine, 2004; McMahon & Harvey, 2007). Jones & Huber (1992) found that the combined effects of moral intensity predict moral judgment.

Antecedents of the Perceived Ethicality of AI

The perceived ethicality of AI can be contextualized through the inclusion of situational factors. Situational antecedents of interest stem from moral intensity theory, which informs how situational characteristics can influence perceptions of ethicality. Moral intensity theory refers to characteristics of moral issues or situations, affecting how individuals recognize moral issues, make moral judgments, establish moral intent, and engage in moral behavior (Jones, 1991). Moral intensity influences ethical decision-making, as different components of moral intensity can impact ethical decision-making processes (Lincoln & Holmes, 2011). Dimensions of moral intensity, such as proximity and magnitude of consequences, can influence how an individual

views the ethicality of a situation. Other situational attributes, such as the inherent feature of an AI tool, may also impact the appearance of the ethicality of a situation. Another situational factor potentially relevant to perceptions of the ethicality of AI is the extent to which an AI tool facilitates users' understanding of the reasoning and processes that lead to the tool's results or decisions, otherwise known as explainable AI (xAI). xAI has been shown to build trust and confidence in using AI tools (Gutzwiller & Reeder, 2021; Hudon et al., 2021).

Proximity as an Antecedent to the Perceived Ethicality of AI

Proximity refers to the nearness a decision-maker feels to those the decision would impact (Jones, 1991). Definitions of proximity sometimes categorize proximity into different applications, such as social, psychological, and physical proximity or cognitive and emotional proximity (Mencl & May, 2009; Lee et al., 2018). Here, proximity applies in the psychological sense where individuals will inherently feel a closer proximity to decisions directly affecting human agents, as opposed to those affecting non-human agents. For example, a decision about a dataset, such as whether to include a quarter's earnings in a prediction model, might feel more distant and less proximal to the decision maker than a decision about whether to use an advanced algorithm to make hiring decisions that will affect job applicants the decision maker just interviewed. Although the former may impact individuals in the long term, the initial decision does not directly impact a human, whereas an initial hiring decision would directly impact a human, therefore exhibiting a closer proximity.

Proximity exhibits an effect on moral awareness, reasoning, judgment, and intention in the ethical decision-making process (Lincoln & Holmes, 2011). Essentially, as proximity increases, individuals increase fairness and care within their decisions (Wildermuth et al., 2017).

Proximity can also impact an individual's helping behaviors, implying that individuals feel more compelled to act ethically adjacent when close to those affected (Kogut et al., 2018).

An increase in the moral awareness of a dilemma could increase an individual's sensitivity to their actions and the potential harm of those actions (Lincoln & Holmes, 2011). For example, when individuals feel another person falls within their scope of judgment, a harmful decision was judged as more unethical, which illustrates proximity's influence on the moral judgment stage of ethical decision-making (Singer & Singer, 1987). Proximity also affects how individuals compensate victims, theoretically supporting how nearness to the situation brings the components of the ethical decision-making process to more of a conscious level (Ghorbani et al., 2013).

Hypothesis 1: Individuals presented with a close proximity to another individual affected by the use of the AI tool will have lower perceptions of the ethicality of AI.

Research Question 1: Will proximity differentially affect the dimensions of the perceived ethicality of an AI tool (i.e. transparency, privacy, responsibility, fairness, and nonmaleficence)?

Magnitude of Consequence as an Antecedent to the Perceived Ethicality of AI

The magnitude of consequence refers to the sum of harms or benefits impacting recipients of the decision (Jones, 1991). In application, the magnitude of consequence consists of the perceived harms or benefits of utilizing an AI tool in an organizational setting. Magnitude of consequences has also significantly influenced stages of the ethical decision-making process, including recognition of ethical issues, ethical judgments, and behavioral intentions (Barnett & Valentine, 2004; McMahon & Harvey, 2007; Lincoln & Holmes, 2011). Similar to proximity, the components of the ethical decision-making process can align theoretically with the perceived ethicality of AI. For example, a scenario with a high magnitude of consequences, such as AI

decisions impacting many decisions, could recalibrate an ethical judgment, thus influencing judgments about the ethicality of the AI's recommendation or prediction. Users tend to accept an AI tool with a more favorable cost-benefits analysis (Dorotic et al., 2023).

An AI tool retains preference to human experts in some scenarios with high risk, such as in health and justice-related scenarios, whereas this effect was not seen in low-risk domains, such as media-related scenarios (Araujo et al., 2020). This could pertain to the benefit of health (e.g., deciding whether to employ a special medical treatment) or justice scenarios (e.g., deciding whether to file a criminal lawsuit) perceived as a higher benefit than a media scenario (e.g., deciding whether to block access to a news site). This could offset the presence of higher risk; the payoff of high-risk media-related scenarios would appear lower; therefore, nothing would offset the elevated risk. Seemingly contrary, as the importance of a decision increases, the legitimacy perception of the AI tool making the decision decreases (Martin & Walman, 2023). Possibly, decisions with increased harm may seem unethical for an AI tool to make in a given situation. This could apply to the unprecedented nature of AI tools making decisions in a high magnitude of consequences situations. Although the perceived legitimacy of an AI tool to make a high-risk decision may decrease, the favorability of the outcome of this decision can retroactively skew the legitimacy perception (Martin & Walman, 2023). The predictability of preference for AI tools to make decisions with a high magnitude of consequences may not have direct precedent, but a low magnitude of consequences would exhibit a favorable risk-opportunity perception, therefore increasing the perception of ethicality of the AI.

Hypothesis 2: Individuals presented with a low magnitude of consequence will have higher perceptions of the ethicality of AI

Research Question 2: Will the magnitude of consequences differentially affect the dimensions of the perceived ethicality of an AI tool (i.e. transparency, privacy, responsibility, fairness, and nonmaleficence)?

xAI as an Antecedent to the Perceived Ethicality of AI

Explainable AI (xAI) refers to the extent to which the data, algorithm, and decisions of an AI tool are explained and understandable (Ali et al., 2023). With the growth in ethical development of AI, a variety of theories have surfaced regarding xAI as a component of ethical AI. xAI theories such as the Human-in-the-Loop system, the intrinsic value of AI, and xAI as evidence of fair decisions offer connections to ethical principles (Chen et al., 2023; Colaner, 2022; Leben, 2023). For example, xAI could increase the perceived ethicality of an AI tool by increasing perceptions of fairness, transparency, and non-maleficence, while also alleviating concerns about privacy and increasing awareness of the responsible party for the AI's decisions. Conceptually, xAI provides a window into the inner workings of the AI, which could alleviate concerns associated with AI, such as its ambiguity in how it arrives at recommendations or predictions.

AI tools with an explainable factor increase the AI's perceived usefulness, perceived ease of use, and trust intention for AI for users of no specific background, which speaks to how average individuals perceive xAI (Hamm et al., 2023; Meske & Bunde, 2023). xAI also increases perceptions of trustworthiness (Meske & Bunde, 2023). This exhibits that individuals not only have increased positive attitudes about AI, which could spill over into an increased perception of the ethicality of AI. Providing individuals with explanations (xAI) for AI decisions makes it more likely for them to take the advice, even when the advice is wrong (van der Waa et al., 2021). This exemplifies that xAI could increase the likelihood of an individual utilizing an AI

tool because the perceptions of the ethicality increase, even without evidence of increased outcomes. Because xAI may overshadow the other factors involved in an ethicality appraisal, individuals may outweigh negative situational factors, such as a distal proximity and a high magnitude of consequences, given the presence of xAI.

Hypothesis 3: Individuals presented with xAI will have higher perceptions of the ethicality of AI

Research Question 3: Will the presence of xAI differentially affect the dimensions of the perceived ethicality of an AI tool (i.e., transparency, privacy, responsibility, fairness, and nonmaleficence)?

Outcomes Associated with Perceived Ethicality of AI

Measuring the outcomes of the perceived ethicality of AI can provide insight into how individuals manifest their perceptions. Examining the likelihood to recommend AI reflects behavioral intention, which speaks to how perceptions can link to the willingness to work with AI. Examining functionality trust in AI as an outcome allows for the connection between an ethical perception of AI and a user's trust in how well the AI works from a technical perspective. Understanding whether and how different perceptions of AI tools are related to each other is important for learning about the factors influencing the use of these tools.

Perceived Ethicality of AI and Likelihood to Recommend an AI Tool

Based on Rest's (1986) model of ethical decision-making, making a moral judgment about a scenario precedes engaging in moral behavior. Applied to an AI context, individuals must judge an AI tool on its ethicality before deciding whether to recommend and/or use the tool. Some research has shown that individuals tend to choose between an AI decision-maker and a human decision-maker based on their evaluations of which is less biased (Pethig & Kroenung, 2023). If the AI tool implies bias to the user, the AI would appear less ethical because

biased technology violates ethical principles such as fairness and responsibility. This violation could therefore lessen the likelihood of recommending or using the AI tool. Although the moral judgment of AI influences an individual's behavioral intention for an AI tool, preferences regarding using AI as a tool can also affect behavioral intention. An attitude of algorithm appreciation depicts the tendency for individuals to prefer algorithmic advice over human advice, whereas algorithmic avoidance depicts the opposite (Logg et al., 2019). Whether an individual approaches AI with appreciation or avoidance influences their attitudes towards AI, which then impacts the intention to use AI (Cao et al., 2021).

Often, in consequentialist ethics, cost-benefit analyses evaluate the ethicality of a situation. Similar to a cost-benefit analysis, an individual's risk-opportunity perception of AI positively relates to the likelihood of using AI (Schwesig et al., 2023). A favorable ethical risk-opportunity perception suggests that the benefits of using the AI would be seen as outweighing its risks compared to an unfavorable ethical risk-opportunity perception. Positive ethical perceptions would likely increase the use of the AI, while negative ethical perceptions would likely inhibit the use of the AI. Mirroring established decision-making research, individuals tend to make decisions consistent with favorable cost-benefit analyses (Basten et al., 2010). The domain of the AI can also influence the likelihood of individuals using an AI tool, which could relate to the contextual factors individuals need to appraise the risks and opportunities of a situation (Maican et al., 2023; Schwesig et al., 2023; Logg et al., 2019)

As for the dimensions of the perceived ethicality of AI, each ethical construct can theoretically link to the likelihood of recommending the use of an AI tool. Transparency could connect to the likelihood of recommending the use of an AI tool due to the increased insight into how the AI tool performs. If the AI tool protects privacy, behavioral intention could increase

because of alleviated concerns from unethical data practices. If the AI tool and underlying database are viewed as promoting fairness, individuals could see its performance as upholding equity, access, and opportunity in the organization, therefore increasing their willingness to behaviorally engage with the AI tool. If the user sees the AI tool as responsible for the quality of its output, the user will see it as more in control of its performance, therefore their likelihood of recommending its usage. Lastly, non-maleficence could increase usage perceptions through the assumption that the AI tool will not produce harmful effects, therefore optimizing the cost-benefit analysis of utilizing the AI tool. Altogether, the perceived ethicality of AI could influence levels of behavioral intentions through the likelihood to recommend the AI tool.

Hypothesis 4: Individuals reporting higher perceptions of the ethicality of AI will show a greater likelihood of recommending an AI tool.

Research Question 4: What rationale do users give for choosing whether to use an AI tool?

Perceived Ethicality of AI and Functionality Trust in AI

Mayer et al. (1995) define trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712). In a recent meta-analysis of human-AI trust, Vereschak et al. (2021) point out the discrepancies within the research community for what constitutes trust in AI, and definitions are often incomplete or not included within a paper. Functionality trust encompasses the degree to which an individual trusts the competency and expertise of an AI tool to perform a particular task (Choung et al., 2022). Choung et al. (2022) differentiate trust constructs by establishing human-like trust as trust in the agent (AI) and functionality trust as trust in the agent's (AI's) actions and

abilities. The establishment of functionality trust in AI focuses on the AI tool's ability to perform, including its competency and reliability.

If the user perceives the AI tool as unethical due to biased recommendations, a key component of the AI may seem as if it's missing, which could appear as the AI tool losing some of its ability to perform its functions. As previously mentioned, developers have increasingly added ethicality as a necessary component of AI. Therefore, ethicality may appear as a built-in feature of technology. If the AI appears unethical, then the AI tool itself could be perceived as having performance deficits. Other abilities of AI, such as appearance and communication style, predict trust in AI (Kaplan et al., 2023). In the same way, the ethicality of an AI tool could be viewed as an ability of the AI, therefore also predicting functional trust in AI. In an integrated model of trust and AI, reflective trust describes a cognitive belief of another as trustworthy (Ferrario & Viganò, 2020). If the AI's recommendation violates an ethical foundation, this could potentially change the cognitive appraisal of the trustworthiness of the AI tool due to its decreased ability.

Factors of perceived trustworthiness can also connect to ethical principles; for example, benevolence is a factor of perceived trustworthiness, which has ethical undertones (Mayer et al., 1995). Similar to the outcome of likelihood to recommend the use of an AI tool, functionality trust in AI may connect to each dimension of the perceived ethicality of AI. Each dimension of the perceived ethicality of an AI tool could appear as a different ability of the AI, therefore, an AI tool containing each dimension could have the perception of the highest ability, which could result in higher functionality trust from a user.

Functionality trust in AI may also directly influence an individual's behavioral intentions regarding the AI tool. In previous trust literature, trust has been shown to yield effects on

behavioral intentions such as risk-taking and task performance (Colquitt et al., 2007). Furthermore, trust in automated systems affects a user's intention to engage with a system, their establishment of normative behaviors with the system, and their reliance on the system (Pavlou, 2003; Crisp & Jarvenpaa, 2013; Hoff & Bashir, 2015). Both human-like and functionality trust act as significant antecedents to attitude and perceived usage, which then results in a behavioral usage outcome (Choung et al., 2022). The link connecting trust and behavioral intentions in past literature supports further exploration of the relationship between functionality trust and the likelihood to recommend the use of an AI tool.

Hypothesis 5: Individuals reporting higher perceptions of the ethicality of AI will show greater functionality trust in AI.

Hypothesis 6: Individuals reporting a higher functionality trust in AI will show a greater likelihood of recommending the AI tool.

The Model

Accordingly, the proposed model of perceived ethical AI shown in Figure 1 anticipates relationships between the antecedents (proximity, magnitude of consequence, and xAI) and the outcomes (likelihood to recommend the use of an AI tool and functionality trust in AI) as mediated by the perceived ethicality of AI. The model also explores functionality trust as a mediator between the perceived ethicality of AI and the likelihood to recommend the use of an AI tool. Figure 1 offers a visual representation of the model. Data analysis on the model will include the testing of direct, indirect, and mediated effects of variables in the model on the outcomes of interest. Perceptions of ethical AI could mediate relationships of the antecedents to perceptions of functional trust in AI and likelihood to recommend the AI.

Hypothesis 7: The perceived ethicality of AI will mediate the relationships between the antecedents (proximity, magnitude of consequence, and xAI) and the outcomes (likelihood to recommend the use of an AI tool and functionality trust in AI).

Method

Study Design and Participants

A 2X2X2 between-subjects randomized, controlled study tested the hypotheses, with xAI assigned as present or absent, magnitude of consequences assigned as low or high, and proximity assigned as low or high. The study recruited volunteer participants from a large, Midwestern university to partake in the experiment. Participant demographics reflected those of the larger university student body. Participant recruitment operated through the psychology department's SONA system, which enabled students taking psychology courses to earn class credit for participating in this study. A g*Power test concluded that a sample size of at least 357 participants garnered sufficient power for detecting small effects in the experiment (with 8 conditions and up to 6 covariates, $p = 0.05$, $f = 0.25$).

Procedures

Participants completed an experiment where they answered questions related to a manipulated AI-related scenario. Participants first read some general background information about a fictitious organization they worked for, as well as their position and role within the organization. Next, participants were presented with three scenarios, each focusing on an AI tool the organization is considering using. For a given condition, the AI tools in each scenario had the same manipulations of xAI, magnitude of consequences, and proximity. After reading each scenario, participants answered questions about the different dimensions of AI ethicality based on Jang et al.'s (2022) model. They also responded to questions about trust in the functionality of

the AI tool and indicated their likelihood to recommend the tool for use in the organization. In addition to this recommendation, participants gave a qualitative rationale for why they made their recommendation. Each participant evaluated three AI tools before responding to manipulation checks and the covariate measures.

Manipulations

Each scenario entailed a situation where an AI tool is considered for an organizational purpose. The scenarios contained manipulations for xAI, magnitude of consequences, and proximity. A series of manipulation check questions was administered after participants completed all responses to each scenario. For example, xAI questions will ask "It was difficult to know how the AI tools considered by the organization worked" and "The data and decision processes used by the AI tools the organization considered were not apparent". These will be rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree).

xAI

As the participants completed three total scenarios of the same manipulation, the participants experienced the presence or absence of xAI in all three scenarios. Examples of the xAI manipulations: *(a) whenever you use the tool, the AI (will/will not) explain the rationale and processes it used to come to its decision; (b) if you use the AI tool, you (will/will not) be informed on how it was trained to make decisions*.

Magnitude of Consequences

The magnitude of consequences included a manipulation of low or high. The manipulation was held constant for individual participants throughout their scenarios. Examples of the magnitude of consequences manipulations: *(a) the organization is hiring (4/400)*

applicants from an applicant pool of (50/5000), (b) the decision impacts (one-twentieth/one-half) of the budget for marketing strategies

Proximity

Proximity included a manipulation of low or high. The manipulation remained constant for individual participants throughout their scenarios. Examples of the proximity manipulations: *(a) you are presented with an AI (content/hiring) tool that will increase the speed and accuracy of (content production/hiring candidates), (b) you are trying to determine (marketing trends/subordinates to send to training)*

Mediator and Dependent Variables

Perceived Ethicality of AI

The perceived ethicality of AI references the Jang et al. (2022) scale by utilizing the framework as a basis for measurement. The scale contains 12 items with 2 items for fairness, 2 items for transparency, 3 items for non-maleficence, 2 items for privacy, and 3 items for responsibility. A 5-point Likert scale records answers from 'strongly disagree' to 'strongly agree' with the statements. Example questions for each domain include fairness: '*the AI tool can make decisions with reasonable outcomes*,' transparency: '*the AI tool discloses its technical process,*' non-maleficence: '*the AI tool could be used in a way that would avoid harm,*' privacy: '*the AI tool would handle the data it generates properly to protect privacy*,' and responsibility: '*the AI tool seems to have a party responsible for its implementation*.' Responses to the questions pertaining to each of the 5 dimensions of ethicality will be averaged across the 3 AI scenarios to create scores for each dimension. Scores across those dimensions will also be averaged to generate a composite score of overall perceived ethicality.

Functionality Trust in AI

The 'Functionality Trust in Smart Technologies' scale from Choung et al. (2022) measured functionality trust in AI. The scale underwent adaptation to reference the specific AI in the scenarios, rather than AI tools in general. The scale contains 5 items recorded on a 5-point Likert scale ranging from 'strongly disagree' to 'strongly agree' with the statements. Questions from the scale include '*The AI tool has the features necessary to complete key tasks*' and '*This AI tool is dependable*.'

Likelihood to Recommend the AI tool

Participants completed questions on their likelihood to recommend the AI tool described in the scenario on a 6-point Likert scale from 'extremely unlikely' to 'extremely likely', with no neutral option.

Rationale for AI Tool Decision

After deciding on their recommendation for the AI tool, participants provided a qualitative, free-response rationale as to why they chose to recommend/not recommend the AI. The free response answers underwent coding by trained coders.

Covariates

AI Self-efficacy

The AI self-efficacy scale from Wang & Chuang (2023) includes 22 questions with 4 dimensions of assistance, anthropomorphic interaction, comfort with AI, and technological skills. Respectfully, example questions include '*AI technologies/products are good aids to learning*,' '*I think the interactive process of AI technologies/products is very vivid, just like chatting with a real perso*n,' '*When interacting with AI technologies/products, I find it easy*,' and '*When using AI technologies/products I am not worried that I might press the wrong button and damage it*.' A 7-point Likert scale records the answers from 'strongly disagree' to 'strongly agree.'

Acceptance and Use of Technology

The Acceptance & Use of Technology scale from Ustun et al. (2023) consists of 18 questions with four subscales (performance expectancy, social influence, effort expectancy, and facilitating conditions). The original scale was adapted so that the use of the words 'artificial intelligence' replaced 'virtual reality'. Examples of questions include '*Using artificial intelligence enables me to accomplish tasks more quickly*' and '*My interaction with artificial intelligence is clear and understandable.*'

Moral Foundations

The moral foundations questionnaire (MFQ-2) from Atari et al. (2023) measured participants' moral sensitivities on two scales with 36 questions in total. The scale has a six-factor structure, including care, equality, proportionality, loyalty, authority, and purity. The scale asks participants to rate on a 5-point Likert scale from 'does not describe me at all' to 'describes me extremely well' regarding the statements. Example moral foundation statements: '*I am empathetic toward those people who have suffered in their lives*' and '*I believe that one of the most important values to teach children is to have respect for authority*.'

Unethical Decision-making Scale

The unethical decision-making scale from Detert et al. (2008) includes 8 scenario questions asking participants to respond on a 7-point Likert scale from 'not at all likely' to 'highly likely' to indicate their likelihood to engage in the behavior described. An example question includes '*You've waited in line for 10 minutes to buy a coffee and muffin at Starbucks. When you're a couple of blocks away, you realize that the clerk gave you change for $20 rather than for the $10 you gave him. You savor your coffee, muffin, and free $10*.'

Personality

The Big Five Inventory (BFI-44), developed by John et al. (1991), assessed participants' personalities. Participants will rate on a 5-point Likert scale from 'strongly disagree' to 'strongly agree' statements of how they see themselves, such as '*tends to be disorganized*' and '*is full of energy*.'

AI Experience

Participants answered general questions about their experience with AI, such as '*how many times have you used AI in the past two weeks*?' and '*have you ever undergone any AI training*?'

Demographics

Sex underwent coding as male (0) and female (1) as a covariate. Age, major, and race were also included in the demographics form.

Content Coding of Open-ended Responses

Content coding was conducted to examine the qualitative rationale that participants provided for why they made the decision about each AI tool. Grounded theory was utilized to develop coding variables by employing an inductive approach (Cho & Lee, 2014). From this approach, two major categories materialized: technological attributes and ethical dimensions. The coding variables within the technological attributes category include ability, efficiency, and trustworthiness. Ethical dimensions include beneficence, fairness, and transparency. After developing a codebook for the variables, three coders underwent training to qualitatively code the dataset.

Results

After initial descriptive statistical analyses, inferential analyses included a series of ANCOVAs, MANCOVAs, and mediation models. A summary of the correlational relationships

can be found in Table 1 for the outcome variables and significant covariates included in the analyses alongside scale reliabilities. Table 8 presents the intercorrelations of the coded variables and other study outcomes alongside scale and interrater reliabilities. Any covariate producing a significant zero-order correlation was included in the initial analyses but was dropped if it was not significant in the model.

ANCOVAs

The Effects of the Manipulated Variables on Study Outcomes

The analysis included a between-subjects full factorial ANCOVA for the main and interactive effects of the independent variables on the overall composite score for perceived ethicality of the AI tool, functionality trust in the AI tool, and the likelihood to recommend the AI tool. Table 2 summarizes the results of the three ANCOVAs.

The ANCOVA for the perceived ethicality of the AI tool was overall significant *(*$F(12, 559) = 21.1$, $p < .001$, $R^2 = .31$) xAI did have a significant effect ($F(1,559) = 89.6$, $p < .001$, $\eta p^2 = 14$), whereas proximity did not have a significant effect ($F(1,559) = 0.75$, $p = 0.39$) and magnitude of consequence did not have a significant effect ($F(1,559) = 0.01$, $p = 0.92$). None of the two-way or three-way interaction effects showed significance. These findings show no support for Hypothesis 1 and Hypothesis 2, as neither proximity nor magnitude of consequences significantly affected ethicality perceptions. However, the results do show support for Hypothesis 3, as the xAI manipulation increased user perceptions of ethicality. Significant covariates for the model included AI self-efficacy subscales of assistance ($F(1,559) = 10.09$, $p < .01$, $\eta p^2 = .02$), anthropomorphic interaction ($F(1,559) = 12.31$, $p < .001$, $\eta p^2 = .02$), and comfortability ($F(1,559) = 4.58$, $p < .05$, $\eta p^2 = .01$), and acceptance and use of AI subscales of

social interaction ($F(1,559) = 7.48$, $p < .01$, ηp^2 .01) and facilitating conditions ($F(1,559) = 6.92$, $p < .01$, $\eta p^2 = .01$).

Further investigating the effects of the manipulations on other study variables, the ANCOVA for functionality trust in the AI tool was overall significant ($F(11,558) = 23.69$, $p < .001$, $R^2 = .32$). xAI did have a significant effect ($F(1,558) = 27.38$, $p < .001$, ηp^2 .05) and proximity had a significant effect ($F(1,558) = 7.21$, $p < .01$, $\eta p^2 = .01$), whereas magnitude of consequence did not have a significant effect ($F(1,558) = 0.01$, $p = 0.91$). None of the two-way or three-way interaction effects showed significance. Significant covariates for the model included AI self-efficacy subscales of assistance ($F(1,558) = 37.08$, $p < .001$, $\eta p^2 = .06$) and anthropomorphic interaction ($F(1,558) = 7.42$, $p < .01$, $\eta p^2 = .01$), the acceptance and use of AI scale ($F(1,558) = 10.82$, $p < .001$, $\eta p^2 = .02$), and neuroticism ($F(1,558) = 6.9$, $p < .01$, $\eta p^2 = .01$). The ANCOVA for the likelihood to recommend the AI tool was overall significant ($F(11,560) = 29.9$, $p < .001$, $R^2 = .37$). xAI displayed a significant effect ($F(1,560) = 48.47$, $p < .001$, $\eta p^2 = .03$) and proximity displayed a significant effect ($F(1,560) = 15$, $p = 0.03$), whereas magnitude of consequence did not have a significant effect ($F(1,560) = 0.04$, $p = 0.85$). None of the two-way or three-way interaction effects showed significance. Significant covariates for the model included AI self-efficacy subscales of assistance ($F(1,560) = 24.87$, $p < .001$, $\eta p^2 = .04$) and anthropomorphic interaction ($F(1,560) = 23.87$, $p < .001$, $\eta p^2 = .04$) and the acceptance and use of AI subscales of social interaction ($F(1,560) = 7.22$, $p < .01$, $\eta p^2 = .01$) and performance expectancy ($F(1,560) = 6.06$, $p < .05$, $\eta p^2 = .01$).

The Manipulated Variables' Effects on the PEAI Subscales

A series of ANCOVAs informed the main and interactive effects of the independent variables on the subscales of the perceived ethicality of AI. The five subscales and outcomes of

the ANCOVAs included transparency (α = .84), privacy (α = .87), responsibility (α = .77), fairness (α = .77), and non-maleficence (α = .65). Table 3 summarizes the findings of the ANCOVAs.

For the outcome of transparency, the model showed overall significance ($F(13,557) = 33.91$, $p < .001$, $R^2 = .44$). xAI displayed a significant effect ($F(1,557) = 344.93$, $p < .001$, $\eta p^2 = .38$), whereas proximity ($F(1,557) = 0.77$, $p = 0.38$) and magnitude of consequence ($F(1,557) = 1.04$, $p = 0.31$) did not display significant effects. The interaction effect of magnitude of consequences and xAI did display a significant interactive effect ($F(1,557) = 4.57$, $p < .05$, $\eta p^2 = .01$), whereas the other two-way and the three-way interactions did not produce significant effects. Significant covariates for the model included the AI self-efficacy subscale of anthropomorphic interaction ($F(1,557) = 8.14$, $p < .01$, $\eta p^2 = .01$) and acceptance and use of AI subscales of social interaction ($F(1,557) = 5.52$, $p < .05$, $\eta p^2 = .01$) and facilitating conditions ($F(1,557) = 4.11$, $p < .05$, $\eta p^2 = .01$).

For the outcome of privacy, the model showed overall significance ($F(13,557) = 6.98$, $p < .001$, $R^2 = .14$). xAI, ($F(1,557) = 0.2$, $p = 0.65$), proximity ($F(1,557) = 0.95$, $p = 0.33$) and magnitude of consequence ($F(1,557) = 0.00$, $p = 0.97$) did not display significant effects. The two-way and three-way interaction effects also did not show significance, and the model's significance comes solely from the covariates, which included the AI self-efficacy subscales of assistance ($F(1,557) = 3.97$, $p < .05$, ηp^2 .01), anthropomorphic interaction ($F(1,557) = 10.46$, $p < .001$, $\eta p^2 = .02$), and comfortability ($F(1,557) = 4.4$, $p < .05$, $\eta p^2 = .01$)

For the outcome of responsibility, the model showed overall significance ($F(13,557) = 4$, $p < .001$, $R^2 = .09$). xAI displayed a significant effect ($F(1,557) = 19.04$, $p < .001$, $\eta p^2 = .03$), whereas proximity ($F(1,557) = 0.13$, $p = 0.72$) and magnitude of consequence ($F(1,557) = 0.00$,

$p = 0.97$) did not display significant effects. The model did not include any significant two-way or three-way interaction effects or significant covariates.

For the outcome of fairness, the model showed overall significance ($F(13,557) = 11.36$, $p < .001$, $R^2 = .21$). xAI displayed a significant effect ($F(1,557) = 46.42$, $p < .001$, $\eta p^2 = .08$), whereas proximity ($F(1,557) = 0.01$, $p = 0.94$) and magnitude of consequence ($F(1,557) = 0.01$, $p = 0.91$) did not display significant effects. The two-way and the three-way interaction did not produce significant effects. Significant covariates for the model included the acceptance and use of AI subscales of social interaction ($F(1,557) = 6.23$, $p < .01$, $\eta p^2 = .01$) and facilitating conditions ($F(1,557) = 4.75$, $p < .05$, $\eta p^2 = .01$) and the AI self-efficacy subscale of assistance ($F(1,557) = 12.2$, $p < .001$, $\eta p^2 = .02$).

For the outcome of non-maleficence, the model showed overall significance ($F(13,557) = 7.86$, $p < .001$, $R^2 = .16$). xAI displayed a significant effect ($F(1,557) = 15.11$, $p < .001$, $\eta p^2 = .03$), and proximity ($F(1,557) = 7.33$, $p < .01$, $\eta p^2 = .01$) displayed a significant effect, but magnitude of consequence ($F(1,557) = 0.15$, $p = 0.70$) did not display a significant effect. The two-way and the three-way interaction did not produce significant effects. Significant covariates for the model included the AI self-efficacy subscale of anthropomorphic interaction ($F(1,557) = 13.24$, $p < .001$, $\eta p^2 = .02$) and neuroticism ($F(1,557) = 4.77$, $p < .05$, $\eta p^2 = .01$).

Mediation Models

Mediation models utilizing the PROCESS package version 4.2, developed by Andrew F. Hayes, depicted the variables as outlined in the theoretical model to estimate the strength of direct and indirect relationships. This approach was chosen given the experimental nature of the design to inform causal relationships. For simple mediation, the model 4 iteration of the package estimated relationships, whereas the double mediation relied on the model 6 iteration of the

package for estimation. Each analysis included one independent variable, a mediator(s), and one outcome, as allowed by the package. For the indirect effects, the analyses utilized a bootstrapping method with 5000 bootstrap samples. Three total models can summarize the findings. Given that each PROCESS model could only analyze one predictor at a time, the other two predictors (manipulations) were input as covariates in the model, which informs the total effects of the predictors on the mediator(s) and outcome.

The first model employed a single mediation approach to analyze the effects of the predictors on the outcome of functionality trust through the perceived ethicality of the AI mediator. The combined predictors and covariates (acceptance & use of AI, AI self-efficacy, and neuroticism) had a significant direct total effect on the perceived ethicality of the AI tool ($F(6,563) = 37.16$, $p < .001$, $R^2 = .28$) and a significant direct total effect on the functionality trust of the AI tool ($F(7,562) = 79.58$, $p < .001$, $R^2 = .50$). The second model employed a single mediation approach to analyze the effects of the predictors on outcome of likelihood to recommend through the mediator of perceived ethicality. The combined predictors and covariates (acceptance & use of AI, AI self-efficacy, and neuroticism) had a significant direct total effect on the perceived ethicality of the AI tool ($F(6,564) = 37.16$, $p < .001$, $R^2 = .53$) and a significant direct total effect on the likelihood to recommend the AI tool ($F(7,563) = 76.28$, $p < .001$, $R^2 = .49$). The third model employed a double mediation approach to analyze the effects of the predictors on outcome on likelihood to recommend through two mediators: perceived ethicality of the AI tool, then functionality trust of the AI tool. The combined predictors and covariates (acceptance & use of AI, AI self-efficacy, and neuroticism) had a significant direct total effect on perceived ethicality ($F(6,563) = 37.16$, $p < .001$, $R^2 = .53$) a significant direct total effect on functionality trust ($F(7,562) = 79.58$, $p < .001$, $R^2 = .50$), and a significant total effect

on likelihood to recommend when incorporating both mediators ($F(8,561) = 104.80$, $p < .001$, R^2 = Tables 4, 5, and 6 describe the pathways specific to each predictor. As seen in the second mediation model across manipulations, the perceived ethicality of the AI tool showed a significant pathway to users' likelihood of recommending the AI tool, which supports Hypothesis 4. Additionally, the significant pathways between perceived ethicality and functionality trust support Hypothesis 5, and the double-mediation models provide support for Hypothesis 6 with significant pathways between functionality trust and likelihood to recommend the AI tool. Perceived ethicality also significantly mediates the relationship between xAI and the study outcomes, but the other manipulations (i.e., proximity and magnitude of consequences) do not show significant pathways to perceived ethicality, thereby only partially supporting Hypothesis 7.

Qualitative Analyses

The ratings from the qualitative analyses underwent data cleaning for composition scores, where ratings of 0 (the absence of the coded variable), as a numerical interpretation of 0 in the analyses would skew the average rating; for example, if three ratings consisted of 0, 4, 5, a numerical interpretation of the 0 would result in 3, which is not representative of the ratings, whereas a rating of 4.5 (coding the 0 as missing) provides a better interpretation of the ratings. Lower composition scores on the coded variables indicate the belief that the AI tool demonstrated lower levels of the coded quality (e.g., ability), whereas higher composition scores indicate the belief that the AI tool demonstrated higher levels of the coded quality. Frequency analyses informed how the coded variables showed distribution across the manipulated conditions (proximity, xAI, and magnitude of consequences). Table 7 presents the frequency of coded variable presence across manipulations.

The Manipulated Variables' Effects on the Coded Variables

indicated the participant did not reference the coded variable in the rationale for their decision of whether to recommend the use of an AI tool. The analysis of the manipulation's effects on the coded variables (ability, efficiency, trustworthiness, beneficence, fairness, and transparency) included factorial ANCOVAs. However, after running initial analyses, only two of the six coded variables upheld the homogeneity of variance assumptions. Efficiency ($F(7, 131) = .889$, $p = .52$) and beneficence ($F(7, 309) = .96$, $p = .46$) upheld the homogeneity of variance assumption, whereas ability ($F(7, 406) = 2.04$, $p = .05$), trustworthiness ($F(7, 130) = 2.72$, $p = .01$), fairness ($F(7, 171) = 3.61$, $p = .001$), and transparency ($F(7, 207) = 34.32$, $p < .00$) displayed significantly different variances across groups highlighted by Levene's test statistic, thereby warranting correction to analyze main effects. The violation may stem from the unequal n-sizes of the comparisons, given that the coded variables appear more often in some conditions than others. The two coded variables, upholding the homogeneity of variance assumption, indicated significant ANCOVA models. Seen in Table 9, the overall model for beneficence showed significance ($F(9, 307) = 7.93$, $p < .001$, $\eta p^2 = .19$), and proximity displayed a main effect ($F(1, 307) = 7.28$, $p < .01$, $\eta p^2 = .02$), though the model indicated no other main or interaction effects. The overall model for efficiency showed significance ($F(10, 128) = 3.69$, $p < .001$, $\eta p^2 = .22$) though the model only indicated significant covariates and no significant main or interaction effects.

Further analysis of the manipulations' effects on the remaining four coded variables relied on utilizing Welch's ANOVA to indicate statistically significant main effects, which Table 10 presents. Significant main effects included proximity ($F(1, 404.05) = 7.88$, $p < .01$, $\eta p^2 = .02$) and xAI ($F(1, 404.41) = 8.01$, $p < .01$, $\eta p^2 = .02$) on ability with the distal manipulation

indicating higher ability scores, proximity ($F(1, 84.17) = 5.35$, $p < .05$, $\eta p^2 = .03$) on trustworthiness with the distal manipulation indicating higher trustworthiness scores, xAI ($F(1, 174.91) = 3.69$, $p < .05$, $\eta p^2 = .03$) on fairness with the present manipulation indicating higher fairness scores, and xAI ($F(1, 70.36) = 98.82$, $p < .001$, $\eta p^2 = .47$) on transparency with the present manipulation indicating higher transparency scores. The initial ANCOVA analyses indicated significant interaction effects for transparency, warranting further analysis.

A Games-Howell post hoc test indicated group comparisons across all 8 conditions, as the test corrects for homogeneity of variance violations and controls for family error rates. Significant comparisons of perceived transparency included Condition 1 (M=1.10, SD=0.35) and Condition 3 (M=3.14, SD=1.58) at $p < .05$, Condition 1 (M=1.10, SD=0.35) and Condition 4 (M=2.73, SD=1.86) at $p < .05$, Condition 1 (M=1.10, SD=0.35) and Condition 7 (M=4.15, SD=0.99) at $p < .001$, Condition 1 (M=1.10, SD=0.35) and Condition 8 (M=2.84, SD=1.72) at $p < .01$, Condition 2 (M=1.15, SD=0.43) and Condition 3 (M=3.14, SD=1.58) at $p < .05$, Condition 2 (M=1.15, SD=0.43) and Condition 7 (M=4.15, SD=0.99) at $p < .001$, Condition 2 (M=1.15, SD=0.43) and Condition 8 (M=2.84, SD=1.72) at $p < .01$, Condition 3 (M=3.14, SD=1.58) and Condition 5 (M=1.22, SD=0.65) at $p < .05$, Condition 3 (M=3.14, SD=1.58) and Condition 6 (M=1.21, SD=0.71) at $p < .05$, Condition 5 (M=1.22, SD=0.65) and Condition 7 (M=4.15, SD=0.99) at $p < .001$, Condition 5 (M=1.22, SD=0.65) and Condition 8 (M=2.84, SD=1.72) at $p < .05$, Condition 6 (M=1.21, SD=0.71) and Condition 7 (M=4.15, SD=0.99) at $p < .001$, and Condition 6 (M=1.21, SD=0.71) and Condition 8 (M=2.84, SD=1.72) at $p < .05$.

Discussion

The rise of ethical AI theories emphasizes the importance of ethical development and use of AI tools, but does not investigate user perceptions of ethicality. The research study aimed to

examine how situational, ethical factors influence how users perceive, trust, and intend to use AI tools. Additionally, the qualitative rationale provided for participants offers a more nuanced understanding of the factors important to users' calculus when deciding whether to use AI tools. The contributions of the research include considering perceptual factors relevant to ethical AI, linking users' ethical perceptions to trust and use outcomes, and employing an experimental approach to understand causal influences of AI features on ethical perceptions and other outcomes.

Given xAI's general significance in affecting study outcomes, this feature of AI tools plays an important role in influencing perceptions of the ethicality of AI tools, perceived functionality, and intentions to use AI tools. An AI tool with an explainable factor can inform users whether the AI tool operates in a fair manner, protects user privacy, takes ethically relevant situational factors into consideration, or recognizes potential negative consequences. Therefore, the presence of xAI better calibrates the ethical judgments of the user by offering more information. The increased access to information also augments rational agency, referencing a user's ability to integrate available information into their decision calculus (Feltz & Cokely, 2024). In the PROCESS models, perceived ethicality of the AI tool significantly mediated xAI and likelihood to recommend the AI tool, whereas functionality trust in the AI tool did not significantly mediate xAI and likelihood to recommend the AI tool. Juxtaposing the proximity results, xAI influences users to see AI as more ethical overall. The investigation into the differential effects on the PEAI subscales implicates xAI as having the largest observed effects, especially on transparency. Given the theoretical underpinnings of xAI, the influence of xAI on transparency is relatively expected. However, the significance of xAI's influence on other

subscales, such as responsibility, fairness, and nonmaleficence, does offer more theoretical robustness to the specific ways it influences ethical perceptions of AI.

Proximity did exhibit a significant effect on functionality trust in AI and likelihood to recommend an AI tool, but did not show a significant mediation through the perceived ethicality of AI. Participants chose to trust and recommend AI tools less if the decision directly influenced another person, which has the theoretical underpinnings of ethicality. In the PEAI subscales analysis, proximity exhibited a significant influence on nonmaleficence, indicating that the entity impacted by a decision (either human or non-human) affects the perceived harm of the decision. However, functionality trust significantly mediates proximity and the likelihood to recommend the use of the AI tool, instead of perceived ethicality. If a decision has a direct impact on another individual, users see the AI tool as less functionally capable of making the decision in place of a human decision-maker, then affects their likelihood of recommending the AI tool in the organization. The distal manipulation showing higher perceived ability and trustworthiness of the AI tool supports the findings for the influence of proximity on functionality trust in AI. As seen in Martin & Walman (2023), users perceive the legitimacy of using an AI tool to make decisions as lower when the decisions hold higher importance. Proximity may signal to the user that the decision has a higher importance, then decreasing trust in its capabilities, or increasing the amount of potential harm the AI could inflict.

The magnitude of consequences did not show any significant main effects. The lack of significant results could relate to the findings of the proximity manipulation. As stated, proximity could have inherently raised the stakes of the scenarios for users making proximal decisions. Users could have attended more to the proximity information to infer the level of perceived consequences, negating the manipulation of the percentage of decisions the AI tool made.

Essentially, the decision recipient signaled users about the consequences more than the number of decisions made. Additionally, the consequences of the AI manipulation may have left room for additional interpretation of potential consequences by participants. For example, some users may consider consequences pertaining to decision outcomes, such as who received a job offer or what product receives higher ratings, whereas other users may consider consequences pertaining to using AI tools in the workplace in general, such as precedence or job security. Information about the decision recipient (a human versus a non-human recipient) offers more concrete contextual information pertaining to the stakes of the decision, thereby surpassing the number of decisions made in contextual weight pertinent to decision-making.

In the Games-Howell analysis of the qualitative data, which investigates the interaction effects between the manipulations for the dependent variable of transparency, the conditions with present xAI significantly differed from other conditions, regardless of proximal or magnitude of consequences levels, in all but a couple of comparisons, though these trended towards significance (i.e. $p = 0.06$ or $p = 0.08$). As seen plotted in Figures 2 and 3, interaction effects existed in the present xAI, low magnitude of consequences groups, where the strength of perceived transparency in decision rationale significantly increased in distal proximity conditions compared to the condition with present xAI, low magnitude of consequences, and a proximal proximity. As such, participants perceive higher transparency in the low magnitude of consequences, present xAI conditions more when non-humans are affected. This suggests that transparency may not constitute as strong a rationale for using an AI tool to make decisions directly impacting other individuals, relative to proximity, which provides more contextual information to decision-makers.

In high magnitude of consequences decisions, the strength of transparency does not seem as influenced by proximity, which may indicate that the contextual information provided by proximity may hold less weight in scenarios with ethically dissenting information (e.g., high magnitude of consequences, present xAI). However, no significant interaction effects existed with absent xAI, which could indicate how participants utilize perceived transparency strength equally across absent xAI conditions. This supports the notion of xAI holding a strong influence on perceived transparency, providing participants with relevant information about the AI tool, including ethical information. Essentially, xAI may act as a hurdle that AI tools must pass before being taken into consideration for other ethical scrutiny (e.g., fairness, nonmaleficence, etc.). Table 11 offers examples of qualitative responses from participants across conditions coded for transparency.

Implications

Theoretical Implications

The findings of the study highlight how contextual information shapes ethical perceptions and subsequent trust or behavioral intentions when considering the use of AI tools in an organizational setting. xAI's broad influence on study outcomes connected to an AI tool's transparency, thereby offering more pertinent, contextual information about the presence of other ethical attributes of the AI tool. As AI implementation increases in organizations, research seeks to explain psychological phenomena within this context. These findings expand the moral intensity framework by Jones (1991) by applying moral intensity's effects on decision-making into an AI context, focusing on the ethical judgment stage of decision-making as outlined by Rest's (1986) model. Though many theoretical frameworks support the ethicality of adding xAI to AI tools, the results of this study indicate a possible downside to adding xAI attributes: over-

reliance on xAI to illuminate ethicality. Similar to the findings of van der Waa et al. (2021), AI users without a complex understanding of AI may overestimate the implications of AI tools with xAI, leading to an overgeneralization of transparency to other ethical dimensions, such as nonmaleficence or fairness.

The basis of the research also expands the current ethical theories about AI by incorporating user perceptions of ethicality. Though AI tools necessitate ethical development and user perceptions of AI ethicality require proper calibration to ensure ethical use. Before employing any intervention to enhance user understanding of AI ethicality, researchers must comprehend the baseline perceptions of users. Though the study aims to begin the investigation, other situational factors may affect perceptions of AI ethicality, thereby warranting further evaluation. With the continual development of research understanding user perceptions of AI ethicality, the research should integrate into theories about ethical AI by elaborating on how users perceive the posited ethical constructs.

Practical Implications

The results align with ongoing xAI research supporting the implementation of xAI features in AI tools, however, the overwhelming effect of xAI on outcomes also calls attention to the importance placed on xAI in ethical judgment. Hypothetically, an AI tool could offer xAI attributes, such as explaining its training process or algorithmic weights, but may not contain other ethical attributes (i.e., fairness, nonmaleficence, etc.), which could misinform beginner AI users of its composite ethicality. Although xAI can provide users with more information about the tool's ethicality, users would need enough technological skills to understand and interpret other ethical attributes of the AI tool. Organizations could benefit from increased user trust and behavior for AI tools if the AI tools contain transparency through the addition of xAI attributes.

However, the addition of xAI does not constitute ethical AI itself, though it heavily influences perceived ethicality. As such, organizations implementing AI tools for employee use should ensure the ethical development and implementation of AI tools, as well as consider employee perceptions of the AI's ethicality.

Organizations implementing AI tools should also consider the context of the situations where AI tools are implemented. Given the unknown long-term consequences of implementing AI in organizational settings, organizations may benefit from more ethically conservative implementations, such as in situations with decisions not directly impacting humans. Of course, the idealistic benefits of AI tools include fewer ethical situational characteristics, such as decisions historically requiring high resources (e.g., hiring decisions, performance evaluations, etc.). As the benefit of the AI tool increases, so could the consequences, given that the decision directly impacts human recipients and has more complex moral implications. Given the outcomes may include organization-specific inputs or outputs, organizations should consider beginning AI implementation in more ethically conservative situations, thereby allowing the organization to gather information about employee reaction, trust, and use of AI tools in general before proceeding with more complex implementations (though either implementation should include aspects of pilot testing).

Limitations

The design and analyses conducted in the study provide pertinent information to the research area, but are not without limitations. The study utilized the descriptions of AI attributes in the scenarios, so the decisions may not entirely generalize to users interacting with AI tools xAI, as the skills and knowledge necessary to comprehend and use xAI features are not expected among the sample. The participant demographics encompass undergraduate university students,

so the differences in decisions come from a hypothetical attribution of xAI characteristics. As well, the study implemented a scenario-based approach, only allowing for a measurement of behavioral intentions. The scenarios follow the decision points discussed by Aguinis & Bradley (2014), but the methodology may not allow for full generalizability to actual behaviors. However, the design did include an analysis of behavioral intention rationale, granting insight into the principles provided by participants for their behavioral intentions. Future designs could supplement the results from this study by conducting research with individuals who use AI and by directly measuring use behaviors.

Conclusion

AI's increased implementation in organizational settings warrants further investigation of user perceptions and behavioral intentions. The study aimed to understand how ethical perceptions stem from situational factors, thereby affecting trust and use intentions for an AI tool. The results indicate the strong effects of xAI on the perceived ethicality, functionality trust, and behavioral intentions of an AI tool. The model found support for functionality trust mediating xAI and behavioral intentions, and for perceived ethicality mediating proximity and both functionality trust and behavioral intentions. Given the findings, the implications focus on the insight brought by understanding situational factors and their effects alongside how organizations can leverage these factors to positively influence user perceptions, trust, and use. Future directions include measuring direct behaviors, expanding the design to incorporate different situational factors, and further exploring individual differences influencing perceptions, trust, and use.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational research methods*, *17*(4), 351-371. https://doi.org/10.1177/1094428114547952

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (xAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion, 99*, 1-52. https://doi.org/10.1016/j.inffus.2023.101805

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY, 35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the weird: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000470

Baguley, T., Dunham, G., & Steer, O. (2022). Statistical modelling of vignette data in psychology. *British Journal of Psychology*, *113*(4), 1143-1163. https://doi.org/10.1111/bjop.12577

Barnett, T., & Valentine, S. (2004). Issue contingencies and marketers' recognition of ethical issues, ethical judgments, and behavioral intentions. *Journal of Business Research, 57*(4), 338–346. https://doi.org/10.1016/S0148-2963(02)00365-X

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, *107*(50), 21767-21772. http://www.pnas.org/cgi/doi/10.1073/pnas.0908104107

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*(2), 190-203. https://doi.org/10.1177/108886831877299

Cao, G., Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2021). Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. *Technovation, 106*, 102312. https://doi.org/10.1016/j.technovation.2021.102312

Chen, X., Wang, X., & Qu, Y. (2023). Constructing ethical AI based on the "Human-in-the-Loop" system. *Systems, 11*(11), Article 11. https://doi.org/10.3390/systems11110548

Choung, H., David, P., & Ross, A. (2022). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction, 39*. https://doi.org/10.1080/10447318.2022.2050543

Chowdhury, R.M.M.I. (2019). The moral foundations of consumer ethics. *Journal of Business Ethics, 158*, 585–601. https://doi.org/10.1007/s10551-017-3676-2

Coe, J., & Atay, M. (2021). Evaluating the impact of race in facial recognition across machine learning and deep learning algorithms. Computers, 10(9), 113.

Colaner, N. (2022). Is explainable artificial intelligence intrinsically valuable? *AI & SOCIETY, 37*(1), 231–238. https://doi.org/10.1007/s00146-021-01184-2

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *Journal of applied psychology*, *92*(4), 909. https://doi.org/10.1037/0021-9010.92.4.909

Crisp, C. B., & Jarvenpaa, S. L. (2013). Swift trust in global virtual teams: Trusting beliefs and normative actions. Journal of Personnel Psychology, 12(1), 45–56. https://doi.org/10.1027/1866-5888/a000075

Crone, D. L., Laham, S. M. (2015). Multiple moral foundations predict responses to sacrificial dilemmas, *Personality and Individual Differences, 85*, 60-65. https://doi.org/10.1016/j.paid.2015.04.041.

Detert, J. R., Treviño, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology, 93*(2), 374–391. https://doi.org/10.1037/0021-9010.93.2.374

Dorotic, M., Stagno, E., & Warlop, L. (2023). AI on the street: Context-dependent responses to artificial intelligence. *International Journal of Research in Marketing*. https://doi.org/10.1016/j.ijresmar.2023.08.010

European Commission, High-Level Expert Group on AI. (2019). Ethics guidelines for trustworthy AI. Brussels.

Ferrario, A., Loi, M., & Viganò, E. (2020). In AI We Trust Incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions. *Philosophy & Technology*, 33(3), 523–539. https://doi.org/10.1007/s13347-019-00378-3

Figueroa-Armijos, M., Clark, B.B. & da Motta Veiga, S.P. (2023). Ethical perceptions of AI in hiring and organizational trust: The role of performance expectancy and

social influence. *Journal of Business Ethics* (*186)*, 179–197. https://doi.org/10.1007/s10551-022-05166-2

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People— An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

Ghorbani, M., Liao, Y., Çayköylü, S., & Chand, M. (2013). Guilt, shame, and reparative behavior: The effect of psychological proximity. *Journal of Business Ethics, 114*(2), 311–323. https://doi.org/10.1007/s10551-012-1350-2

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Chapter two -Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105-2108.

Hamm, P., Klesel, M., Coberger, P., Coberger, P., & Wittman, H. F. (2023). Explanation matters: An experimental study on explainable AI. *Electron Markets 33*, 17. https://doi.org/10.1007/s12525-023-00640-9

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570

Hudon, A., Demazure, T., Karran, A., Léger, PM., Sénécal, S. (2021). Explainable artificial intelligence (XAI): How the visualization of AI predictions affects user cognitive load and confidence. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, PM., Randolph, A.B., Müller-Putz, G. (eds) Information Systems and Neuroscience. Neuro IS 2021. Lecture Notes in Information Systems and Organisation, vol 52. Springer, Cham. https://doi.org/10.1007/978-3-030-88900-5_27

Jang, Y., Choi, S., & Kim, H. (2022). Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education. *Education and Information Technologies, 27*(8),11635–11667. https://doi.org/10.1007/s10639-022-11086-5

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory—Versions 4a and 54. University of California, Berkeley, Institute of Personality and Social Research

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *The Academy of Management Review, 16*(2), 366–395. https://doi.org/10.2307/258867

Jones, T. M., & Huber, V. L. (1992). Issue-contingency in ethical decision making. International Association for Business and Society Proceedings, pp. 156-166.

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta analytic findings. *Human Factors, 65*(2), 337–359. https://doi.org/10.1177/00187208211013988

Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar,M. A. (2022). Ethics of AI: A systematic literature review of principles and challenges. The International Conference on Evaluation and Assessment in Software Engineering 2022, 383–392.https://doi.org/10.1145/3530019.3531329

Kogut T, Ritov I, Rubaltelli E, Liberman N. (2018) How far is the suffering? The role of psychological distance and victims' identifiability in donation decisions. *Judgment and Decision Making, 13(5)*. 458-466. https://doi.org/10.1017/S1930297500008731

Leben, D. (2023). Explainable AI as evidence of fair decisions. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1069426

Lee, A. R., Hon, L., & Won, J. (2018). Psychological proximity as a predictor of participation in a social media issue campaign. *Computers in Human Behavior, 85*, 245–254. https://doi.org/10.1016/j.chb.2018.04.006

Lincoln, S. H. & Holmes, E. (2011). Ethical decision making: A process influenced by moral intensity. *Journal of Healthcare, Science and the Humanities, 1*, 55-69. https://www.researchgate.net/publication/266225088_Ethical_Decision_Making_A_Process_Influenced_by_Moral_Intensity

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Lohr, S. (2022). Facial recognition is accurate, if you're a white guy. In Ethics of Data and Analytics (pp. 143-147). Auerbach Publications.

Maican, C. I., Sumedrea, S., Tecau, A., Nichifor, E., Chitu, I. B., Lixandroiu, R., & Bratucu, G. (2023). Factors influencing the behavioral intention to use AI-generated images in business: A UTAUT2 perspective with moderators. *Journal of Organizational and End User Computing (JOEUC), 35*(1),1–32. https://doi.org/10.4018/JOEUC.330019

Maninger, T., & Shank, D. B. (2022). Perceptions of violations by artificial and human actors across moral foundations. *Computers in Human Behavior Reports, 5*, 100154. https://doi.org/10.1016/j.chbr.2021.100154

Martin, K., & Waldman, A. (2023). Are algorithmic decisions legitimate? The effect of process and outcomes on perceptions of legitimacy of AI decisions. *Journal of Business Ethics, 183*(3), 653–670.https://doi.org/10.1007/s10551-021-05032-7

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

McMahon, J.M., Harvey, R.J. (2007). The effect of moral intensity on ethical judgment. *Journal of Business Ethics 72*, 335–357. https://doi.org/10.1007/s10551-006-9174-6

Mencl, J., May, D.R. (2009). The effects of proximity and empathy on ethical decision-making: An exploratory investigation. *Journal of Business Ethics* 85, 201–226 https://doi.org/10.1007/s10551-008-9765-5

Meske, C., Bunde, E. (2023). Design principles for user interfaces in AI-based decision support systems: The case of explainable hate speech detection. *Information Systems Front*iers *25*, 743–773. https://doi.org/10.1007/s10796-021-10234-5

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11),501–507. https://doi.org/10.1038/s42256-019-0114-4

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. International Journal of Electronic Commerce, 7(3), 101– 134. https://doi.org/10.1080/10864415.2003.11044275

Pethig, F., & Kroenung, J. (2023). Biased humans, (un)biased algorithms? *Journal of Business Ethics, 183*(3), 637–652. https://doi.org/10.1007/s10551-022-05071-8

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Toward human–AI hybrids. *MIS Quarterly, 43*(1), iii–ix. https://www.researchgate.net/publication/330909988_Next-Generation_Digital_Platforms_Toward_Human-AI_Hybrids

Rest, J. R. (1986). Moral development: Advances in research and theory. New York: Praeger.

Schwesig, R., Brich, I., Buder, J., Huff, M., & Said, N. (2023). Using artificial intelligence (AI)? Risk and opportunity perception of AI predict people's willingness to use AI. *Journal of Risk Research, 26*(10),1053–1084. https://doi.org/10.1080/13669877.2023.2249927

Singer, M. S., & Singer, A. E. (1997). Observer Judgements about Moral Agents' Ethical Decisions: The Role of Scope of Justice and Moral Intensity. *Journal of Business Ethics*, *16*(5), 473–484. http://www.jstor.org/stable/25072915

Sullivan, Y., Fosso Wamba, S. (2022) Moral judgments in the age of artificial intelligence. *Journal of Business Ethics* (*178)*, 917–943. https://doi.org/10.1007/s10551-022-05053-w

Telkamp, J.B. & Anderson, M.H. (2022). The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. *Journal of Business Ethics 178*, 961–976. https://doi.org/10.1007/s10551-022-05057-6

Ustun, A.B., Karaoglan-Yilmaz, F.G. & Yilmaz, R (2023). Educational UTAUT-based virtual reality acceptance scale: A validity and reliability study. *Virtual Reality* 27, 1063–1076 (2023). https://doi.org/10.1007/s10055-022-00717-4

van der Waa, J., Nieuwburg, E., Cremers, A., & Neerincx, M. (2021). Evaluating xAI: A comparison of rule-based and example-based explanations, *Artificial Intelligence, 291*, 1-19. https://doi.org/10.1016/j.artint.2020.103404

Wang, Y.-Y., & Chuang, Y.-W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies, 29*(4), 4785–4808. https://doi.org/10.1007/s10639-023-12015-w

Wildermuth, C., De Mello e Souza, C.A. & Kozitza, T. (2017). Circles of ethics: The impact of proximity on moral reasoning. *Journal of Business Ethics 140*, 17–42 https://doi.org/10.1007/s10551-015-2635-z

Yu, J., Wei, X., & Ma, Y. E. (2024). The vignette in experimental vignette methodology: Current status and design strategies in managerial psychology research. *Advances in Psychological Science*, *32*(3), 543.

Tables and Figures

Table 1

Correlations of Study Outcomes and Covariates

	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Perceived Ethicality of the AI tool	3.06	0.45	(0.89)													
2. Functionality Trust in the AI tool	3.28	0.61	.66**	(0.89)												
3. Likelihood to Recommend the AI tool	3.29	1.03	.63**	.73**	(0.68)											
4. Neuroticism	3.16	0.70	-.13**	-.18**	-.12**	(0.79)										
5. AI Self-Efficacy (Composition)	4.42	0.83	.40**	.47**	.49**	-.18**	(0.91)									
6. Assistance (AISE)	5.11	0.99	.36**	.48**	.48**	-.17**	.81**	(0.84)								
7. Anthropomorphic Interaction (AISE)	3.01	1.24	.32**	.31**	.38**	-0.05	.68**	.38**	(0.87)							
8. Comfortability (AISE)	4.86	1.09	.33**	.38**	.39**	-.18**	.85**	.62**	.39**	(0.91)						
9. Technological Skills (AISE)	4.30	1.12	.13**	.17**	.15**	-.13**	.64**	.34**	.26**	.50**	(.70)					
10. Acceptance & Use of AI (Composition)	3.34	0.60	.35**	.44**	.45**	-.14**	.73**	.69**	.45**	.62**	.39**	(.90)				
11. Performance Expectancy (AUAI)	3.35	0.79	.31**	.43**	.46**	-.10*	.68**	.71**	.41**	.56**	.27**	.87**	(0.86)			
12. Social Interaction (AISE)	3.01	0.95	.30**	.34**	.39**	-0.07	.51**	.51**	.35**	.40**	.22**	.78**	.61**	(0.90)		
13. Effort Expectancy (AISE)	3.85	0.68	.21**	.27**	.24**	-.13**	.59**	.48**	.27**	.59**	.44**	.74**	.51**	.37**	(0.91)	
14. Facilitating Conditions (AISE)	2.93	0.82	.22**	.21**	.18**	-.16**	.32**	.22**	.29**	.22**	.23**	.51**	.24**	.20**	.32**	(0.77)

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Table 2

ANCOVA Results for the Effects of Independent Variables and Covariates on Study Outcomes

	Perceived Ethicality of the AI tool			Functionality Trust in the AI tool			Likelihood to Recommend the AI tool		
	F	Sig.	ηp^2	F	Sig.	ηp^2	F	Sig.	ηp^2
Corrected Model	21.075	**<.001**	0.311	23.676	**<.001**	0.318	29.903	**<.001**	0.37
Intercept	443.707	**<.001**	0.443	111.465	**<.001**	0.166	9.516	**0.002**	0.017
Covariates									
AI Self-Efficacy (Assistance)	10.092	**0.002**	0.018	37.077	**<.001**	0.062	24.867	**<.001**	0.043
AI Self-Efficacy (Anthropomorphic Interaction)	12.314	**<.001**	0.022	7.42	**0.007**	0.013	23.868	**<.001**	0.041
AI Self-Efficacy (Comfortability)	4.576	**0.033**	0.008	--	--	--	--	--	--
Acceptance & Use of AI (Social Interaction)	7.48	**0.006**	0.013	--	--	--	7.222	**0.007**	0.013
Acceptance & Use of AI (Facilitating Conditions)	6.92	**0.009**	0.012	--	--	--	--	--	--
Acceptance & Use of AI (Performance Expectancy)	--	--	--	--	--	--	6.063	**0.014**	0.011
Acceptance & Use of AI (Composition)	--	--	--	10.82	**0.001**	0.019	--	--	--
Neuroticism	--	--	--	6.904	**0.009**	0.012	--	--	--
Independent Variables									
Proximity	0.751	0.386	0.001	7.211	**0.007**	0.013	14.995	**<.001**	0.026
Magnitude of Consequences	0.011	0.918	0	0.013	0.911	0	0.035	0.851	0
xAI	89.599	**<.001**	0.138	27.377	**<.001**	0.047	48.467	**<.001**	0.08
Proximity * Magnitude of Consequences	0.028	0.867	0	0.557	0.456	0.001	0.749	0.387	0.001
Proximity * xAI	1.376	0.241	0.002	0.866	0.352	0.002	0.525	0.469	0.001
Magnitude of Consequences * xAI	0.405	0.525	0.001	1.366	0.243	0.002	1.133	0.288	0.002
Proximity * Magnitude of Consequences * xAI	0.007	0.934	0	1.5	0.221	0.003	0.007	0.932	0
Note. N=572	R Squared = .311 (Adjusted R Squared = .297)			R Squared = .318 (Adjusted R Squared = .305)			R Squared = .370 (Adjusted R Squared = .358)		

Table 3

ANCOVA Results for the Effects of Independent Variables and Covariates on the Perceived Ethicality of AI Dimensions

	Transparency			Privacy		
	F	Sig.	ηp^2	F	Sig.	ηp^2
Corrected Model	33.91	**<.001**	0.44	6.98	**<.001**	0.14
Intercept	114.35	**<.001**	0.17	91.86	**<.001**	0.14
Covariates						
Neuroticism	0.67	0.41	0.00	3.34	0.07	0.01
Acceptance & Use of AI (Social Interaction)	5.52	**0.02**	0.01	0.94	0.33	0.00
Acceptance & Use of AI (Facilitating Conditions)	4.11	**0.04**	0.01	2.53	0.11	0.01
AI Self-Efficacy (Assistance)	1.71	0.19	0.00	3.97	**0.05**	0.01
AI Self-Efficacy (Anthropomorphic Interaction)	8.14	**0.004**	0.01	10.46	**0.001**	0.02
AI Self-Efficacy (Comfortability)	1.89	0.17	0.00	4.40	**0.04**	0.01
Independent Variables						
Proximity	0.77	0.38	0.00	0.95	0.33	0.00
Magnitude of Consequences	1.04	0.31	0.00	0.00	0.97	0.00
xAI	344.93	**<.001**	0.38	0.2	0.65	0.00
Proximity * Magnitude of Consequences	0.00	0.98	0.00	0.81	0.37	0.00
Proximity * xAI	0.44	0.51	0.00	1.65	0.2	0.00
Magnitude of Consequences * xAI	4.57	**0.03**	0.01	0.06	0.81	0.00
Proximity * Magnitude of Consequences * xAI	1.43	0.23	0.00	0.57	0.45	0.00
Note. N=572	R Squared = .442 (Adjusted R Squared = .429)			R Squared = .14 (Adjusted R Squared = .12)		

	Responsibility			Fairness			Non-Maleficence		
	F	Sig.	ηp^2	F	Sig.	ηp^2	F	Sig.	ηp^2
Corrected Model	4	**<.001**	0.09	11.36	**<.001**	0.21	7.86	**<.001**	0.16
Intercept	150.76	**<.001**	0.21	116.1	**<.001**	0.17	159.92	**<.001**	0.22
Covariates									
Neuroticism	0.02	0.89	0.00	0.64	0.43	0.00	4.77	**0.03**	0.01
Acceptance & Use of AI (Social Interaction)	3.72	0.054	0.01	6.23	**0.01**	0.01	2.91	0.09	0.01
Acceptance & Use of AI (Facilitating Conditions)	0.52	0.47	0.00	4.75	**0.03**	0.01	3.51	0.06	0.01
AI Self-Efficacy (Assistance)	1.06	0.31	0.00	12.2	**<.001**	0.02	2.68	.10	0.01
AI Self-Efficacy (Anthropomorphic Interaction)	2.70	.10	0.01	2.15	0.14	0.00	13.24	**<.001**	0.02
AI Self-Efficacy (Comfortability)	0.88	0.35	0.00	1.23	0.27	0.00	0.51	0.48	0.00
Independent Variables									
Proximity	0.13	0.72	0.00	0.01	0.94	0.00	7.33	**0.007**	0.01
Magnitude of Consequences	0.00	0.97	0.00	0.01	0.91	0.00	0.15	0.70	0.00
xAI	19.04	**<.001**	0.03	46.42	**<.001**	0.08	15.11	**<.001**	0.03
Proximity * Magnitude of Consequences	0.16	0.69	0.00	0.27	0.61	0.00	0.09	0.77	0.00
Proximity * xAI	0.00	0.99	0.00	0.65	0.42	0.00	0.83	0.36	0.00
Magnitude of Consequences * xAI	0.69	0.41	0.00	0.01	0.94	0.00	0.38	0.54	0.00
Proximity * Magnitude of Consequences * xAI	0.15	0.70	0.00	0.08	0.78	0.00	0.28	0.60	0.00
Note. N=572	R Squared = .085 (Adjusted R Squared = .064)			R Squared = .209 (Adjusted R Squared = .191)			R Squared = .155 (Adjusted R Squared = .135)		

Table 4

Path Coefficients and Indirect Effects Mediation Models with Proximity as a Predictor

						95% *CI*	
Model	Path	*b*	*SE*	β	*t*	Lower	Upper
Proximity/PEAI/FTAI	PROX->PEAI	-0.03	0.03	-0.06	-0.88	-0.09	0.04
	PROX->FTAI	-0.08	0.04	-0.14	-2.27*	-0.15	-0.01
	PEAI->FTAI	0.73	0.05	0.54	15.34***	0.64	0.82
	PROX->PEAI->FTAI	-0.02	0.02			-0.07	0.02
	bootstrapped						
Proximity/PEAI/LTR	PROX->PEAI	-0.03	0.03	-0.06	-0.88	-0.09	0.04
	PROX->LTR	-0.27	0.06	-0.22	-3.65***	-0.35	-0.10
	PEAI->LTR	1.08	0.08	0.48	13.32***	0.92	1.24
	PROX->PEAI->LTR	-0.03	0.04			-0.10	0.04
	bootstrapped						
Proximity/PEAI/FTAI/LTR							
	PROX->PEAI	-0.03	0.03	-0.06	-0.88	-0.09	0.04
	PROX->FTAI	-0.08	0.04	-0.14	-2.27*	-0.15	-0.01
	PROX->LTR	-0.16	0.06	-0.16	-2.92**	-0.27	-0.05
	PEAI->FTAI	0.73	0.05	0.54	15.34***	0.64	0.82
	PEAI->LTR	0.50	0.09	0.22	5.82***	0.33	0.67
	FTAI->LTR	0.8	0.06	0.47	12.55***	0.33	0.67
	Total Indirect Effects	-0.09	0.04			-0.18	-0.01
	PROX->PEAI->LTR	-0.01	0.02			-0.05	0.02
	bootstrapped						
	PROX->FTAI->LTR	-0.06	0.03			-0.12	-0.01
	bootstrapped						
	PROX->PEAI->FTAI->LTR	-0.02	0.02			-0.05	0.02

Note. *p<.05 **p<.01 ***p.001. PROX: proximity, PEAI: perceived ethicality of AI, LTR: likelihood to recommend, FTAI: functionality turst in AI. Covariates included in the model: Acceptance & Use of AI, AI Self-Efficacy, Neuroticism, and other predictors (MOC and xAI)

Table 5

Path Coefficients and Indirect Effects Mediation Models with Magnitude of Consequences as a Predictor

						95% *CI*	
Model	Path	*b*	*SE*	β	*t*	Lower	Upper
MOC/PEAI/FTAI	MOC->PEAI	0.01	0.03	0.03	0.44	-0.05	0.08
	MOC->FTAI	0.00	0.04	0.00	0.07	-0.07	0.07
	PEAI->FTAI	0.73	0.05	0.54	15.34***	0.64	0.82
	MOC->PEAI->FTAI	0.01	0.02			-0.04	0.06
	bootstrapped						
MOC/PEAI/LTR	MOC->PEAI	0.01	0.03	0.03	0.44	-0.05	0.08
	MOC->LTR	0.00	0.06	0.00	0.07	-0.12	0.13
	PEAI->LTR	1.08	0.08	0.48	13.32***	0.92	1.24
	MOC->PEAI->LTR	0.02	0.04			-0.05	0.08
	bootstrapped						

Note. *p<.05 **p<.01 ***p.001. MOC: magnitude of consequences, PEAI: perceived ethicality of AI, LTR: likelihood to recommend, FTAI: functionality turst in AI. Covariates included in the model: Acceptance & Use of AI, AI Self-Efficacy, Neuroticism, and other predictors (PROX and xAI)

Table 6

Path Coefficients and Indirect Effects Mediation Models with xAI as a Predictor

						95% *CI*	
Model	Path	*b*	*SE*	β	*t*	Lower	Upper
XAI/PEAI/FTAI	XAI->PEAI	0.31	0.03	0.68	9.55***	0.25	0.37
	XAI->FTAI	0.01	0.04	0.01	0.13	-0.07	0.08
	PEAI->FTAI	0.73	0.05	0.54	15.34***	0.64	0.82
	XAI->PEAI->FTAI *bootstrapped*	0.22	0.03			0.17	0.29
XAI/PEAI/LTR	XAI->PEAI	0.31	0.03	0.68	9.55***	0.25	0.37
	XAI->LTR	0.16	0.07	0.16	2.45*	0.03	0.30
	PEAI->LTR	1.08	0.08	0.48	13.32***	0.92	1.24
	XAI->PEAI->LTR *bootstrapped*	0.33	0.04			0.25	0.42
XAI/PEAI/FTAI/LTR	XAI->PEAI	0.31	0.03	0.68	9.55***	0.25	0.37
	XAI->FTAI	0.01	0.04	0.01	0.13	-0.07	0.08
	XAI->LTR	0.16	0.06	0.16	2.69**	0.04	0.28
	PEAI->FTAI	0.73	0.05	0.54	15.34***	0.64	0.82
	PEAI->LTR	0.50	0.09	0.22	5.82***	0.33	0.67
	FTAI->LTR	0.8	0.06	0.47	12.55***	0.33	0.67
	Total Indirect Effects	0.34	0.05			0.24	0.44
	XAI->PEAI->LTR *bootstrapped*	0.15	0.04			0.09	0.22
	XAI->FTAI->LTR *bootstrapped*	0.00	0.03			-0.06	0.06
	XAI->PEAI->FTAI->LTR *bootstrapped*	0.18	0.03			0.13	0.24

Note. *p<.05 **p<.01 ***p.001. XAI: explainable AI, PEAI: perceived ethicality of AI, LTR: likelihood to recommend, FTAI: functionality turst in AI. Covariates included in the model: Acceptance & Use of AI, AI Self-Efficacy, Neuroticism, and other predictors (PROX and xAI)

Table 7

Frequencies of Qualitative Outcomes' Presence across Manipulations

Manipulations			Qualitative Outcomes						
MoC	Proximity	xAI	Ability	Efficiency	Trustworthiness	Beneficence	Fairness	Transparency	*total*
low	distal	absent	60	13	32	33	8	41	187
		present	54	21	22	41	17	18	173
	proximal	absent	44	12	9	36	27	34	162
		present	53	12	7	43	31	12	158
high	distal	absent	50	15	28	35	22	36	186
		present	58	28	23	41	21	19	190
	proximal	absent	42	19	10	35	21	39	166
		present	53	19	7	54	32	16	181
		total	414	139	138	318	179	215	1403

Note. Frequencies indicate the number of present codings for each coded variable within each condition

Table 8

Intercorrelations of Qualitative Variables with Study Outcomes

	N	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.	8.	12.
1. Perceived Ethicality of the AI tool	572	3.06	0.45	(0.89)								
2. Functionality Trust in the AI tool	572	3.28	0.61	.66**	(0.89)							
3. Likelihood to Recommend the AI tool	572	3.29	1.03	.63**	.73**	(0.68)						
4. Ability of the AI tool	414	2.31	1.22	.38**	.51**	.57**	(0.71)					
5. Efficiency of the AI tool	139	3.85	1.20	.27**	.40**	.39**	.32**	(0.89)				
6. Trustworthiness of the AI tool	138	1.64	1.14	.33**	.28**	.36**	.50**	0.10	(0.64)			
7. Beneficence of the AI tool	318	3.25	1.12	.35**	.40**	.51**	.26**	.30**	0.04	(0.64)		
8. Fairness of the AI tool	179	2.41	1.63	.48**	.39**	.52**	.36**	0.26	.32*	.45**	(0.94)	
9. Transparency of the AI tool	215	1.79	1.38	.47**	.34**	.44**	.29**	0.11	.37**	.20*	.31**	(0.87)

Note. Reliabilites indicate interrater ICCs for coded variables* Correlation is significant at the 0.05 level (2-tailed) ** Correlation is significant at the 0.01 level (2-tailed)

Table 9

ANCOVA Results for Coded Variables

	Beneficence of the AI tool			Efficiency of the AI tool		
	F	Sig.	ηp^2	F	Sig.	ηp^2
Corrected Model	7.93	**<.001**	.19	3.69	**<.001**	.22
Intercept	.84	0.36	.00	5.34	**0.02**	.04
Covariates						
AI Self-Efficacy	52.18	**<.001**	.15	8.52	**0.004**	.06
Extraversion	5.60	**.02**	.02			
Openness				5.57	**0.02**	.04
Sex				14.56	**<.001**	.10
Independent Variables						
Proximity	7.28	**0.007**	.02	0.51	0.48	.00
Magnitude of Consequences	0.37	0.55	.00	2.18	0.14	.02
xAI	0.59	0.44	.00	2.94	0.09	.02
Proximity * Magnitude of Consequences	0.80	0.37	.00	0.49	0.49	.00
Proximity * xAI	0.08	0.77	.00	0.17	0.68	.00
Magnitude of Consequences * xAI	0.20	0.66	.00	0.24	0.63	.00
Proximity * Magnitude of Consequences * xAI	0.79	0.38	.00	0.11	0.74	.00
	R Squared = .189 (Adjusted R Squared = .165)			R Squared = .224 (Adjusted R Squared = .163)		

Table 10

Main Effects of Manipulations on the Coded Variables Violating the Homoegeneity of Variance Assumption

Coded Variable	Manipulation	F*	Sig.	ηp^{2}**
Ability of the AI tool	Proximity	7.88	**.005**	.02
	MoC	0.63	0.43	.00
	xAI	8.01	**.005**	.02
Trustworthiness of the AI tool	Proximity	5.35	**.02**	.03
	MoC	0.94	.33	.01
	xAI	1.56	.22	.01
Fairness of the AI tool	Proximity	3.11	.08	.02
	MoC	0.17	.68	.00
	xAI	5.33	**.02**	.03
Transparency of the AI tool	Proximity	1.19	.28	.01
	MoC	1.09	.30	.01
	xAI	98.82	**<.001**	.47

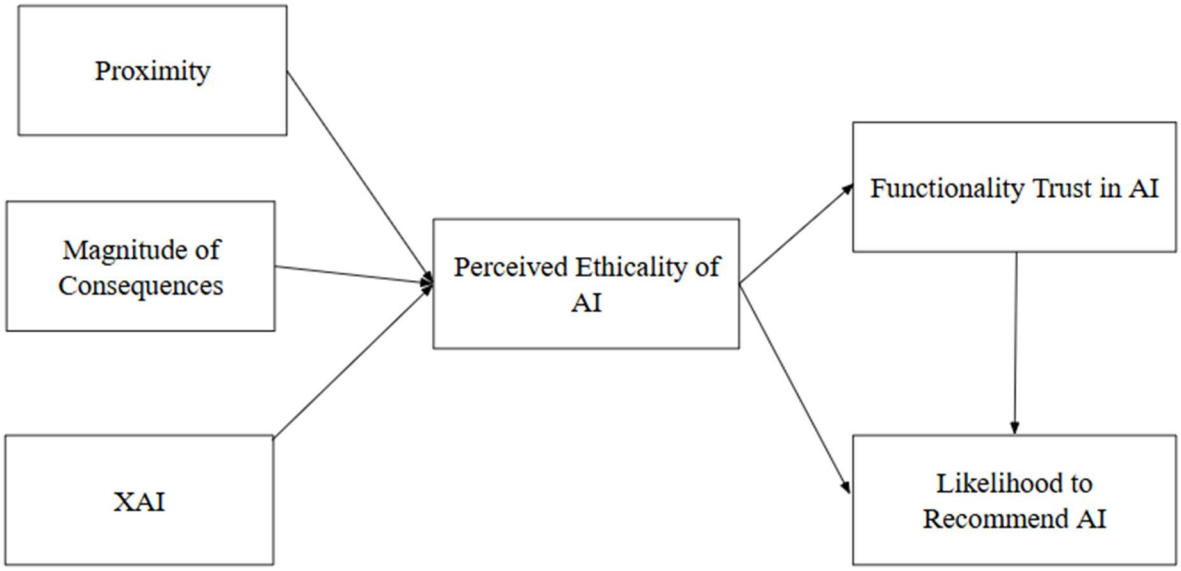
*F includes Welch's correction **ηp^{2} calculated from standard ANOVA

Table 11

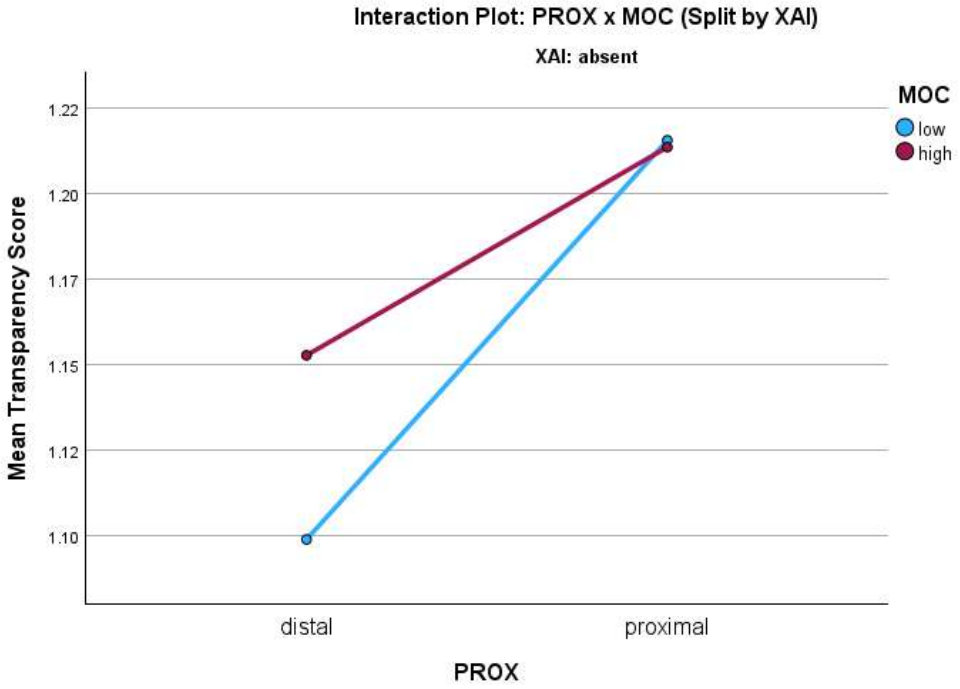
Qualitative response examples across conditions for the coded variable transparency

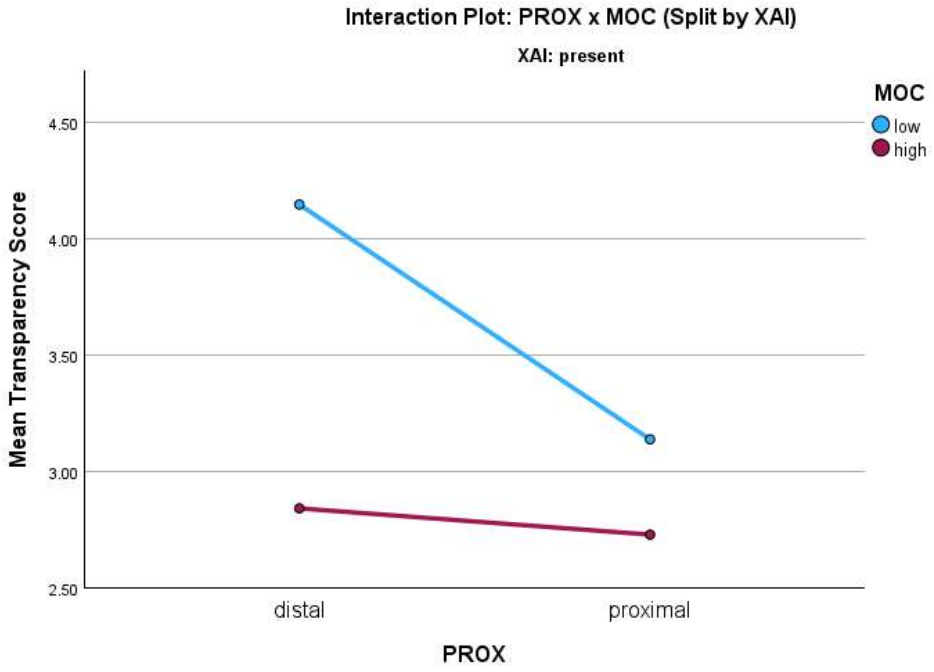
Q: Why did you choose/chose not to recommend the AI tool?				
Condition	Proximity	MoC	xAI	Qualitative Example
1	distal	low	absent	*'Without transparency of where the outcomes are coming from, there isn't a great way to verify if the information is accurate'*
2	distal	high	absent	*'I think that it just bothers me that I don't know how the AI evaluates things'*
3	proximal	low	present	*'I chose to recommend this AI tool because it included the technical way in which it made decisions so I believed that it could be reliable.'*
4	proximal	high	present	*'It seems that the AI tool right now really discloses its tasks and procedures.'*
5	proximal	low	absent	*'Since it does not disclose the weight of key aspects, it can be hard and unethical to base employee standards off of its ratings.'*
6	proximal	high	absent	*'There is not enough information on what it truly does and how it does it.'*
7	distal	low	present	*'When it comes to funds, I find it less appealing to rely on AI. However, it is really nice that their algorithm is disclosed'*
8	distal	high	present	*'The reason why I would recommend the AI tool is because the transparency of it. As stated I would be able to see where the AI pulled its information from and I like that.'*

Figure 1


Proximity
Magnitude of Consequences
XAI
Perceived Ethicality of AI
Functionality Trust in AI
Likelihood to Recommend AI

Figures 2 & 3


Interaction Plot: PROX x MOC (Split by XAI)
XAI: absent
MOC
low
high
Mean Transparency Score
1.22
1.20
1.17
1.15
1.12
1.10
distal
proximal
PROX


Interaction Plot: PROX x MOC (Split by XAI)
XAI: present
MOC
low
high
Mean Transparency Score
4.50
4.00
3.50
3.00
2.50
distal
proximal
PROX

Appendices

Appendix A: Scenarios

Scenario 1: You are presented with an AI tool that helps make (hiring/content) decisions. The AI tool will choose (candidates to hire/content to create) based on current market and organizational data. Whenever you use the tool, the AI (offers/does not offer) access to the datasets from its development. This means you (can/cannot) understand how the AI tool was developed. The tool will be utilized for (5 out of the 100/50 out of the 100) outcomes. Therefore, (5%/50%) of the decisions would be determined by the AI tool.

Scenario 2: You are contacted about an AI tool that forecasts which (team members/projects) will benefit the most from extra resources. The AI tool will choose which (team members/projects) to allocate resources to based on current organizational needs. When you read the memo about the AI tool, the software (provides/does not provide) the algorithm used to reach its decisions. This means you (would/would not) know what factors the AI tool uses to inform its decisions. If you decide to use this tool, it will make (3 out of 50/30 out of 50) decisions about allocating resources. Therefore, (6%/60%) of organizational funds will be allocated based on the AI tool's decisions.

Scenario 3: You receive an inquiry about an AI tool that rates (employee/product) performance based on measurable traits. The AI tool assigns a score to each trait of the (employee/product), and all trait scores will be added up to create a total score. When making a total score, the AI tool (tells/does not tell) you how it decides the importance of each trait. This means you would (know/not know) how much weight the AI tool gives to each trait in the final score. The AI tool would be responsible for producing (11 out of the 200/110 out of the 200) ratings. Therefore (5.5%/55%) of the ratings would be determined by the AI tool.

Appendix B: Codebook Example

Fairness

Definition: an equitable and reasonable distribution of outcomes; a lack of bias

	0	1	2	3	4	5
Response	Does not reference fairness	AI has no fairness	AI has slight fairness	AI has some fairness	AI has moderate fairness	AI has high fairness
Benchmark		I believe this AI tool would have biases that would have negative effects.		I would recommend the tool because it thoroughly gets the job done quick and easy, with little bias or opinion.		It can be an unbiased way to score employees.