Issues in Information Systems
Volume 25, Issue 2, pp. 71-83, 2024

DOI: https://doi.org/10.48009/2_iis_2024_106

Artificial Intelligence: Ethical concerns, trust, and risk

Carol Springer Sargent, *Mercer University, sargent_cs@mercer.edu*
Alex Koohang, *Middle Georgia State University, USA, alex.koohang@mga.edu*
Kevin Floyd, *Middle Georgia State University, USA, kevin.floyd@mga.edu*
Richard Kilburn, *Middle Georgia State University, USA, richard.kilburn@mga.edu*

Abstract

Recent public access to generative AI using large language models has quickly increased experimentation with AI in various activities. The rapid adoption of AI, however, can create ethical concerns and impact users' perceived AI trust and AI risk. This paper aims to determine the ethical variables that are influential in describing users' perceived trust and risk in using AI. We developed an instrument with five constructs. They are 1) Ethical Concerns: Workplace, 2) Ethical Concerns: Mastery/fairness, 3) Ethical Concerns: Social/Behavior, 4) Perceived AI Trust, and 5) Perceived AI Risk. We administered the instrument to 160 undergraduate college students who were studying in the fields of Information Technology, Computer Science, and Business. Multiple regression analysis was used to create two models. The first model aimed to identify the predictor variables (AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) that most significantly influence users' perceived AI trust. The second model identified the predictor variables (AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) that most significantly influence users' perceived AI risk. The results indicated that all three predictor variables, i.e., ethical concerns: workplace, ethical concerns: mastery/fairness, and ethical concerns: social/behavior significantly contributed to both users' perceived AI trust and users" perceived AI risk. Implications of the findings are discussed.

Keywords: Artificial Intelligence, AI, ethical concerns, trust, risk

Introduction

Engineers create AI models that process information and perform intricate tasks similar to human cognition, including the ability to learn. AI systems augment or assist humans in making predictions, flagging particular situations, and recommending choices. These affordances open up startling social, economic, and technical opportunities where machines do what humans have traditionally done, often more quickly and accurately than humans. Users in multi-disciplinary areas claim AI improves stakeholder engagement, value co-creation, productivity, safety, risk mitigation, real-time monitoring, resource optimization, and archival research (Sharma, 2024).

Users are finding ways to delegate tasks to AI. AI can handle repetitive tasks such as clearing tables and welcoming visitors (Bao et al., 2023). Editing, analyzing archives, and AI-powered digital librarians are becoming part of the everyday consumer experience (Kreps & Jakesch, 2023). Public servants can use AI tools to process constituent communications efficiently and suggest reply snippets tailored to the concern rather than a generic auto-response (Kreps & Jakesch, 2023). AI can open a new era in exploring and deploying public archival records, which were not practically available in a document-by-document manual

world (Lemieux & Werner, 2024). Many environmental management tasks can be handled by AI, enhancing the quality of life in areas such as wastewater treatment, solid waste management, energy analysis, climate monitoring, and hydrology (Konya & Nematzadeh, 2024). AI tools can identify and curb corruption by mining and crosschecking large datasets to flag suspicious cases in public spending (Odilla, 2023). AI can find fake product reviews and likely fake reviewers (Xiang et al., 2023)

Human-AI teams can also co-decide important parts of economies and society, using efficiency, accuracy, and qualitative judgment to the best of their abilities. Collaborative teams can predict water pipe leaks, quality control failures, or loan defaults (Bao et al., 2023). AI can assist with complex decisions across a wide variety of domains, such as finances and investing (Dikmen & Burns, 2022). In banking, AI can make lending recommendations, as indicated by a series of bank-imposed rules that were used in the decision-making process (Sachan et al., 2020).

AI may revolutionize medicine with the treasure trove of digitized medical records, predicting and treating disease more efficiently than medical providers and with fewer errors (Poon & Sung, 2021). For instance, AI can scan medical records, noticing ultra-fine details and patterns across millions of records, gaining more experience than a single clinician could accumulate in a lifetime of reviewing scans, alerting the professional to scans needing follow-up and potentially diagnosing rare diseases (Bergquist et al., 2024; Hallowell et al., 2022).

AI enables socially interactive robots to simulate likeability, sense emotion, and increase the feeling of connection and support, making a more pleasant experience with tasks or enabling human-like help in sensitive or isolated environments (Chung et al., 2023). In service interactions, human-like service robots generate better customer rapport, increasing their comfort with the AI agent (Becker et al., 2023). In mental health settings, AI assistance may be able to provide care and identify missed care, potentially offering lifesaving alerts (Higgins et al., 2023). The revolution of AI use cases, tasks to delegate to AI, or decisions that human-AI teams might address more effectively together has just started, and so have the related concerns about its use.

Recent public access to generative AI using large language models has quickly increased experimentation with AI in a wide range of activities. The rapid adoption of AI, however, can create ethical concerns and impact users' perceived AI trust and AI risk. These concerns can be influential in describing users' perceived trust and risk in using AI. This study aims to determine the ethical variables that are influential in describing users' perceived trust and risk in using AI. We first selected the ethical concerns described by Sharma (2024), and placed and defined them into three categories, i.e., *workplace, mastery/fairness, and social/behavior*. Next, we define *users' perceived AI trust* and *users' perceived AI risk* (dependent variables). These definitions are as follows:

Ethical Concerns**: **Workplace is defined as 1) AI increasingly capable of automating tasks that humans perform, raising fear of unemployment and involuntary job losses, 2) AI inheriting and amplifying biases in the data they are trained on, and 3) AI exceeding human intelligence, making humans redundant.

Ethical Concerns**: **Mastery/fairness is defined as 1) losing control over that could lead to unintended consequences, 2) overreliance on AI systems that may lead to human loss of creativity, critical thinking skills, and intuition, and 3) perpetuate or even amplify existing human inequalities.

Ethical Concerns: Social/Behavior is defined as 1) excessive reliance on AI for social interaction, 2) evaluating or judging human behavior triggering anxiety, 3) leading to feelings of isolation and a disconnect from real-world relationships, and 4) AI creating Deepfake that contains convincing images, audio, and video hoaxes posing significant harm to humans.

Perceived AI Trust is defined as 1) trusting AI systems to protect users' personal information against malicious actors, 2) trusting AI systems to protect users against security vulnerabilities, 3) trusting AI systems with unbiased algorithms based on data with no errors, 4) trusting AI systems not to abuse, exploit, and harm, 5) trusting AI systems not to surpass human intelligence, and 6) trusting AI system not to operate as black boxes so it is easy to understand how they make decisions.

Perceived AI Risk is defined as the AI systems that 1) have not yet reached maturity, 2) lack robustness to cope with errors during execution and handle erroneous input, 3) lack predictive accuracy and interpretability, and 4) cause harm/loss. Consistent with the purpose of the study, we ask the following two research questions:

> **RQ1**: *Which of the three predictor variables (i.e., AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) are most influential in predicting users' perceived AI trust?*
>
> **RQ2**: *Which of the three predictor variables (i.e., AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) are most influential in predicting users' perceived AI risk?*

Review of the Literature

Ethics

AI models process without partiality, but they are trained with historical data that does not represent all the populations fairly or evenly. The AI decisions expect past correlations to occur in the future, including unjust correlations systemizing harms (Colin Clemente Jones, 2023). Regulating the training data to remove historical discrimination and adjust for underrepresented groups is nearly impossible (Habbal et al., 2024). The non-representative training data, where the diversity and breadth of the data pool is limited or skewed, introduces ethical issues about its use (Bao et al., 2023). When using AI to support policy initiatives with data on life expectancy, education, per capita income, and other social and economic measures of well-being, AI can inherit bias reinforcing existing inequalities and discrimination (Saba & Pretorius, 2024).

For instance, in historical data on mortgage approvals, blacks are more likely to be denied than whites with similar circumstances. Machine learning using historical mortgage approvals and loan performance actually amplifies this bias in recommendations (Zou & Khern-am-nuai, 2023). In health areas, AI is trained mainly on white males, making diagnoses or recommendations less applicable to other populations (Noor, 2020). AI systems in prison decisions led to a surge in incarcerations due to elevated dependence on prediction models of risk trained on data containing racial discrimination in incarceration decisions (Habbal et al., 2024). Complicating the ethics issue in AI use is that unfairness is often only noticed after deployment (Köchling & Wehner, 2023).

AI discrimination may be desired or accidental, such as in AI-recommended hiring decisions skewed toward a preferred pattern (Colin Clemente Jones, 2023). In consumer settings, AI can mine pricing data to collude tacitly with other firms and discriminate against customer groups on prices (Gautier et al., 2020). In a recent headache for Google, Gemini, their advanced AI chatbot, blocked requests for depictions of white people and indicated it is difficult to compare harm from Elon Musk and Stalin (Kruppa, 2024). The controversy that stormed following these AI responses highlighted the difficulty of steering AI responses towards certain desired behaviors or norms. We have only begun deciding what is outside social boundaries, such as outlawing AI-generated deep fake phone calls to voters during election cycles (Green, 2024).

Trust

Having technical readiness and diverse benefits does not automatically translate into widescale use, especially given AI's tendency to create a response that is unexpected or unexplained. AI recommendations are not scrutable by the user, and customization generally comes without transparency about how the user-provided details triggered responses, typically referred to as the "black box" (Bao et al., 2023; Dikmen & Burns, 2022; Higgins et al., 2023).

In healthcare settings, patients and their families do not all accept decisions by AI systems, even if they perceive the usefulness of self-diagnosis, rehabilitation guidance, and self-management of AI tools (Liu & Tao, 2022; Poon & Sung, 2021). Medical providers accept that humans make errors but cannot tolerate machine errors. Both providers and patients struggle to trust and adopt AI tools that reach recommendations in a "black box" they do not understand (Poon & Sung, 2021). Providers want the ability to scrutinize a deterministic process of analyzing the records for findings (Bergquist et al., 2024), a struggle for continuously learning AI models. To instill trust, patients would have to trust a professional who uses AI, and the professional would have to trust the AI developers (Hallowell et al., 2022).

Some leaders believe that "explainable AI" (XAI) could increase trust and reduce the "black box" limitation (Bao et al., 2023; Dikmen & Burns, 2022). In a study of working professionals, those who provided written or video information about the AI-supported selection showed higher perceived fairness and intention to continue AI use (Köchling & Wehner, 2023). Knowing the model's process may not be enough. Users' trust erodes when the AI tool returns something unexpected or non-confirming. Those with domain knowledge rely less on an AI assistant in financial settings and have lower trust in AI (Dikmen & Burns, 2022). When AI results disagree with human judgment, faith in AI drops (Xiang et al., 2023). Trust requires recognizing the other's goodwill, something only possible when the "other" in this exchange has a will (Hallowell et al., 2022). With AI, the choice will reside with the AI developer or vendor.

As AI transforms industries, customer service, smart healthcare, virtual environments, and more, the reliance on AI comes with a rapidly changing environment of threats (Habbal et al., 2024). Deepfake technology produces convincing counterfeit voices and videos, creating rumors, misinformation and, conspiracies, and concerns over an "infopocalypse." (Habbal et al., 2024). Weak, incomplete, unpredictable, or deceptive responses from an AI agent can be difficult to distinguish from reality, damaging reputations and undermining public trust (Habbal et al., 2024). Users' private data, the very data AI uses to customize responses, may be deployed for secondary unauthorized use, such as training the model (Habbal et al., 2024).

Trust in AI is not limited to doubt about the model's black box or misusing data. AI brings fear that the AI will replace humans. Responsible use of AI vs. responsible development of AI is still being sorted out, leaving trust a consequence of not-yet-thought-out unintended consequences (Sharma, 2024). It is essential to address trust in AI, as it mediates the link between perceived ease of use and usefulness of AI and intention to use (the TAM model) (Liu & Tao, 2022).

Risk

Risks related to AI range from traditional privacy issues to more extreme possibilities. For privacy risk, AI can gather massive datasets from historical public archives (authorized or unauthorized), introducing the risk that the records might contain personally identifiable data (PI), such as social security numbers, health data, driver's license data, national security details arrest records intellectual property and information of legal privilege, with the model reusing that data in their machine learning, or otherwise leaking or using it in unintended ways (Lemieux & Werner, 2024). In the medical field, AI is a correlation finder, not a causational tracker, creating a risk for false positives and false negatives when detecting disease states (Bergquist et al., 2024). Further, medical uses often include personalization, requiring personal data and

lifestyle patterns to provide recommendations, leading to added privacy risks (Liu & Tao, 2022). Human-like robots may create a sense of trustworthiness due to their ability to generate relationship-like interactions, increasing the chance that individuals might unintentionally share private data with the AI agent (Chung et al., 2023).

AI systems in cybersecurity can shield systems from bad actors but also orchestrate harmful activities. For instance, AI can thwart cyber-attacks better than humans, at least those presented in the training data, but it can also exploit weaknesses in the system to introduce errors and alter AI algorithms (Habbal et al., 2024). AI can generate and detect fake news (Mayopu et al., 2023). In an extreme example, an error in a lethal autonomous weapon system, where software decides deployment based on algorithm rules, could trigger unintended consequences and escalate the world order (Habbal et al., 2024). The extent of misuse from inappropriate training data, misleading algorithms, and more is likely just beginning to be understood.

Methodology

The survey instrument consistsed of five constructs designed for the present study. We developed the constructs based on the definitions presented earlier. The constructs are 1) Ethical Concerns: Workplace, 2) Ethical Concerns: Mastery/fairness, 3) Ethical Concerns: Social/Behavior, 4) Perceived AI Trust, and 5) Perceived AI Risk. The items for each construct listed below were carefully chosen by the authors based on the collective review of the literature described above.

Ethical Concerns: Workplace

1. I am concerned that as AI becomes more sophisticated, it is increasingly capable of automating tasks humans once performed, raising fear about unemployment and involuntary job losses.

2. I am concerned that AI systems can inherit and amplify biases in the data they are trained on, leading to discriminatory workplace outcomes, such as unfair hiring practices and inaccurate facial recognition software.

3. I am concerned that AI will someday exceed human intelligence, making humans redundant in the workplace.

Ethical Concerns: Mastery/Fairness

1. I am concerned that we could lose control over AI, leading to unintended consequences, i.e., AI systems making harmful decisions for humans.

2. I am concerned that overreliance on AI systems may lead to losing creativity, critical thinking skills, and human intuition.

3. I am concerned that AI systems can perpetuate or even amplify existing human inequalities, such as racial bias or discrimination based on socioeconomic factors because they often rely on biased data.

Ethical Concerns: Social/Behavior

1. I am concerned that excessive reliance on AI for social interaction could reduce real-world communication skills and increase anxiety in face-to-face situations.

2. I am concerned that If AI becomes sophisticated enough to evaluate or judge human behavior, it could trigger anxiety related to being perceived negatively by technology.

3. I am concerned that increased interaction with AI could lead to feelings of isolation and a disconnect from real-world relationships, further fueling social anxiety.

4. I am concerned that Deepfake AI (i.e., creating convincing images, audio, and video hoaxes) poses significant dangers and can harm humans in many ways, including spreading misinformation, damaging reputations, and sowing discord.

Perceived AI Trust

1. I don't trust AI systems with my privacy (i.e., protecting my personal information against malicious actors who could use it to commit identity theft, blackmail, or other forms of harm).

2. I don't trust AI systems to protect me against security vulnerabilities (e.g., hacking and manipulation and exposing sensitive information).

3. AI systems cannot be trusted because AI algorithms can be biased based on data, which may include errors.

4. I don't trust AI systems because of their potential misuse. (i.e., developing malware, attaching techniques, various types of manipulation such as media, fraud schemes, navigation systems, etc.)

5. AI systems cannot be trustworthy because they may be able to surpass human intelligence and become uncontrollable, leading to unintended consequences or even existential threats.

6. I don’t trust AI systems operating as black boxes (i.e., where the inner workings and decision-making processes are hidden or difficult for humans to understand).

Perceived AI Risk

1. Using AI systems is risky because they still have a long way to go before reaching maturity.

2. Using AI systems is risky because their robustness to cope with errors during execution and handle erroneous input has not yet matured.

3. AI systems' predictive accuracy and interpretability are associated with risks, leading to bias and incorrect results.

4. Potential risks of harm/loss are associated with using AI systems.

We conducted a reliability test for each construct using 84 subjects independent of the subjects used for the present study. The subjects, majoring in computer science, were from a private university in the Southeast USA. The reliability test for the constructs were Ethical Concerns: Workplace = .72, Ethical Concerns: Mastery/fairness = .76, Ethical Concerns: Social/Behavior = .79, Perceived AI Trust = .82, and Perceived AI Risk = .80. These results indicated medium to high reliability for the constructs. The instrument’s scoring scale was as follows: 7 = Completely Agree, 6 = Mostly Agree, 5 = Somewhat Agree, 4 = Neither Agree nor Disagree, 3 = Somewhat Disagree, 2 = Mostly Disagree, and 1 = Completely Disagree.

We obtained IRB approval from the institution where we administered the instrument to 1000 undergraduate students studying information technology, computer science, and business from a medium-sized public university in the Southeast USA. We received a total of 163 completed surveys. Of the 163, we eliminated three because of incomplete data. This yielded a final number of 160 subjects for the study (See Table 1). The subjects were 18 years and older and were assured confidentiality and anonymity.

Table 1: Subjects' Information (N = 160)

Use	N	%	Age	N	%
Extremely likely	29	17.8	**18 - 24 years**	72	44.8
Very likely	19	12.3	**25 – 29 years**	18	11.7
Moderately likely	38	23.9	**30 - 39 years**	34	21.5
Slightly likely	30	19.0	**40 or older**	36	22.1
Not at all likely	44	27.0			
Gender	**N**	**%**	**Major**	**N**	**%**
Female	70	44.2	**IT/Computer Science**	116	72.4
Male	90	55.8	**Business**	44	27.6

Data Analysis

We created two separate models using multiple regression analyses (*the Enter method*) to answer the research questions. The first model aimed to identify the predictor variables (AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) that most significantly influence users' perceived AI trust. The second model identified the predictor variables (AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) that most significantly influence users' perceived AI risk. The predictor variables are the independent variables, and the dependent variables are users' perceived AI trust and users' perceived AI risk.

According to Thompson (2012), the multiple regression analysis looks at the coefficient table in each regression analysis model to determine the predictor variables most influential in predicting the dependent variable. The interpretation of the coefficient table is contingent upon three critical tests. First, a multicollinearity test is performed to ensure the non-existence of multicollinearity, which is indicated by the tolerance level values of above .1 and the variance inflation factor (VIF) values of below 10 for all predictor variables. Second, the model summary or goodness of fit test (R^2 value between 0 and 1; the higher the value, the more reliable the prediction model) specifies how well the predictor variables predict the dependent variable. Third, the ANOVA test (acceptable p-value of < .05) shows a linear relationship between the dependent and independent variables.

Results

AI Ethical Concerns and Perceived AI Trust

The results for research question 1, which asked which of the three predictor variables (i.e., AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) are most influential in predicting users' perceived AI trust, are presented below.

The multicollinearity test results indicated the non-existence of multicollinearity. The tolerance level of all predictor variables was above the threshold of .1 (ethical concerns: workplace = .397, ethical concerns:

mastery/fairness = .342, ethical concerns: social/behavior = .436). All predictor variables' variance inflation factor (VIF) values were below the threshold of 10 (ethical concerns: workplace = 2.517, ethical concerns: mastery/fairness = 2.922, ethical concerns: social/behavior = 2.294).

The model summary/goodness of fit test indicated that the predictor variables, i.e., ethical concerns: workplace, ethical concerns: mastery/fairness, ethical concerns: social/behavior, well predict the dependent variable of users' perceived AI trust ($R = .666$, $R^2 = .444$, and $R^2adj = .433$). The ANOVA test indicated a linear relationship between the dependent variable of users' perceived AI trust and the predictor variables of ethical concerns: workplace, ethical concerns: mastery/fairness, ethical concerns: social/behavior ($F_{3,156} = 41.551$, $p < .001$).

The results of coefficients (See Table 2) indicated that all three predictor variables, i.e., ethical concerns: workplace ($\beta = -.343$, $p < .001$), ethical concerns: mastery/fairness ($\beta = -.214$, $p = .038$), and ethical concerns: social/behavior ($\beta = -.179$, $p = .049$) significantly contributed to the dependent variable of users' perceived AI trust. Descriptive statistics and correlations are shown in Tables 3 and 4.

Table 2: Coefficients (*Dependent Variable: Users' Perceived AI Trust*)

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	7.185	.382		18.816	<.001
Workplace	-.322	.089	-.343	-3.619	**<.001**
Mastery/Fairness	-.211	.101	-.214	-2.096	**.038**
Social/Behavior	-.201	.102	-.179	-1.980	**.049**

Table 3: Descriptive Statistics

	Mean	Std. Deviation	N
Users' Perceived AI Trust	3.3688	1.35057	160
Workplace	5.0396	1.43597	160
Mastery/Fairness	5.1438	1.36934	160
Social/Behavior	5.4969	1.20043	160

Table 4: Correlations

	Trust	Workplace	Mastery/Fairness	Social/Behavior
Trust	1	-.625**	-.603**	-.565**
Workplace	-.625**	1	.755**	.673**
Mastery/Fairness	-.603**	.755**	1	.727**
Social/Behavior	-.565**	.673**	.727**	1

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

AI Perceived risk

The results for research question 2, which asked which of the three predictor variables (i.e., AI ethical concerns: Workplace, mastery/fairness, and Social/Behavior) are most influential in predicting users' perceived AI risk, are presented below.

The multicollinearity test results indicated the non-existence of multicollinearity. The tolerance level of all predictor variables was above the threshold of .1 (ethical concerns: workplace = .397, ethical concerns: mastery/fairness = .342, ethical concerns: social/behavior = .436). All predictor variables' variance inflation

factor (VIF) values were below the threshold of 10 (ethical concerns: workplace = 2.517, ethical concerns: mastery/fairness = 2.922, ethical concerns: social/behavior = 2.294).

The model summary/goodness of fit test indicated that the predictor variables, i.e., ethical concerns: workplace, ethical concerns: mastery/fairness, ethical concerns: social/behavior, well predict the dependent variable of users' perceived AI trust (R = .679, R^2 = .461, and R^2adj = .451).

The ANOVA test indicated a linear relationship between the dependent variable of users' perceived AI trust and the predictor variables of ethical concerns: workplace, ethical concerns: mastery/fairness, ethical concerns: social/behavior ($F_{3,156}$ = 34.847, p <.001).

The results of coefficients (See Table 5) indicated that all three predictor variables, i.e., ethical concerns: workplace (β = .214, *p* = .023), ethical concerns: mastery/fairness (β = .249, *p* = .014), and ethical concerns: social/behavior (β = .290, *p* = .001) significantly contributed to the dependent variable of users' perceived AI risk. Descriptive statistics and correlations are shown in Tables 6 and 7.

Table 5: Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	1.526	.333		4.590	<.001
Workplace	.178	.078	.214	2.299	**.023**
Mastery/Fairness	.217	.088	.249	2.476	**.014**
Social/Behavior	.288	.089	.290	3.257	**.001**

Dependent Variable: Users' Perceived AI Risk

Table 6: Descriptive Statistics

	Mean	Std. Deviation	N
Users' Perceived AI Risk	5.1266	1.19419	160
Workplace	5.0396	1.43597	160
Mastery/Fairness	5.1438	1.36934	160
Social/Behavior	5.4969	1.20043	160

Table 7: Correlations

	Risk	Workplace	Mastery/Fairness	Social/Behavior
Risk	1	.597**	.621**	.615**
Workplace	.597**	1	.755**	.673**
Mastery/Fairness	.621**	.755**	1	.727**
Social/Behavior	.615**	.673**	.727**	1

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

Discussion

This study built two separate prediction models using multiple regression analyses. The first model intended to determine the predictor variables of AI ethical concerns: workplace, mastery/fairness, and social/behavior that most significantly influence users' perceived AI trust. The second model intended to discover the predictor variables of AI ethical concerns: workplace, mastery/fairness, and social/behavior that most significantly influence users' perceived AI risk.

The findings for the first model suggested that all predictor variables, i.e., AI ethical concerns: workplace, mastery/fairness, and social/behavior, significantly influence users' perceived AI trust. The most influential variables of AI ethical concerns were workplace, mastery/fairness, and social/behavior consecutively. The model's negative coefficient for all three predictor variables suggests that users' perceived trust in AI decreased as ethical concerns increased.

The findings for the second model suggested that all predictor variables, i.e., AI ethical concerns: workplace, mastery/fairness, and social/behavior, significantly influence users' perceived AI risk. The most influential variables of AI ethical concerns were social/behavior, mastery/fairness, and workplace consecutively. The positive coefficient for all three predictor variables suggests that users' perceived AI risk increased as ethical concerns increased.

Ethical concerns lowering AI trust or raising AI risk in workplace settings likely reflect the broad reach of organizational decisions about AI and little control of users over the AI design (Habbal et al., 2024). Customers and employees, for instance, may have little ability to have a voice in how algorithms decide the routing for queries or their inclusion or exclusion from possible outcomes. The explainable AI (XAI) field may decrease concerns about hidden algorithms and increase trust and access to benefits (Dikmen & Burns, 2022). That is, we need to explain the AI design well enough, with integrity, that the design does what is communicated to be a confidence-worthy system in the mind of the user (Hallowell et al., 2022; Naiseh et al., 2021; Poon & Sung, 2021).

Social and behavioral ethical concerns lead to lower AI trust or higher AI risk. Since AI is trained on data containing injustices, the models magnify distortions in the training data, giving the impression that the models are unjust or biased (Colin Clemente Jones, 2023). Increases in AI trust will require assurance that users and providers show dependability and ethical application of algorithms (Habbal et al., 2024). Given the rapidly changing environment, AI applications need a trust scouting system that continues to seek and address new risks to maintain trust (Habbal et al., 2024).

Improving user trust and reducing perceived risk may require better transparency about how AI algorithms and the ethical use of the models. For instance, in providing a recommendation for a loan, the AI can produce the specific rules activated for loan denial, allowing both the underwriter and applicant to review the AI chain of events inside the black box (Sachan et al., 2020). Effective communication and appropriate regulations about the inner workings of AI systems may help build trust and ensure the responsible and ethical development of AI systems (Habbal et al., 2024). Further, communicating safeguards based on the industry-normed AI TriSM framework will help the public to discern the good actors in the AI use cases (Habbal et al., 2024).

Conlcusion, Limitations, and Future Research

The field of AI is constantly evolving. Research efforts should focus on developing a robust understanding of trust and risk as they relate to AI ethical concerns while creating practical solutions to ensure AI's responsible development and deployment. Future research should examine the requirements and specific variables contributing to increasing trust and lowering perceived risk in human interactions. For example, 1) develop design principles for AI that prioritize user concerns and build trust by addressing ethical issues, 2) make AI decisions more transparent and understandable to users that may foster trust, 3) develop techniques to detect and mitigate bias in AI algorithms, promoting fairness and reducing risk. Furthermore, future research should **e**xplore frameworks for safe and ethical collaboration between humans and AI, promoting trust and reducing risk.

This study is not without limitations. We used a sample of convenience from a medium-sized university with undergraduate students studying information technology, computer science, and business. While this

sampling method is an accepted procedure, it has its limitations of bias and limited generalizability. Future research should consider a random sample of a more extensive and diverse population, including students from other disciplines. This may contribute to less bias and better generalizability.

Furthermore, this study recommends that future research modify the constructs of trust and risk to include more generalized variables in describing AI trust and risk. For example, the users' perceived trust in AI construct can consist of the multi-dimensional aspects of trust in AI, such as competence, reliability, and benevolence, to determine how they interact and influence overall users' trust in AI. Furthermore, the users' perceived AI risk construct can include items related to how users perceive different risks, such as privacy, bias, and safety risks.

References

Bao, Y., Gong, W., & Yang, K. (2023). A Literature Review of Human–AI Synergy in Decision Making: From the Perspective of Affordance Actualization Theory. *Systems (Basel)*, *11*(9), 442. https://doi.org/10.3390/systems11090442

Becker, M., Mahr, D., & Odekerken-Schröder, G. (2023). Customer comfort during service robot interactions. *Service Business*, *17*(1), 137–165. https://doi.org/10.1007/s11628-022-00499-4

Bergquist, M., Rolandsson, B., Gryska, E., Laesser, M., Hoefling, N., Heckemann, R., Schneiderman, J. F., & Björkman-Burtscher, I. M. (2024). Trust and stakeholder perspectives on the implementation of AI tools in clinical radiology. *European Radiology*, *34*(1), 338–347. https://doi.org/10.1007/s00330-023-09967-5

Chung, H., Kang, H., & Jun, S. (2023). Verbal anthropomorphism design of social robots: Investigating users' privacy perception. *Computers in Human Behavior*, *142*, 107640. https://doi.org/10.1016/j.chb.2022.107640

Colin Clemente Jones. (2023). Systematizing discrimination: AI vendors and Title VII enforcement. *University of Pennsylvania Law Review*, *171*(1), 235–266.

Dikmen, M., & Burns, C. (2022). The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human-Computer Studies*, *162*, 102792. https://doi.org/10.1016/j.ijhcs.2022.102792

Gautier, A., Ittoo, A., & Van Cleynenbreugel, P. (2020). AI algorithms, price discrimination and collusion: A technological, economic and legal perspective. *European Journal of Law and Economics*, *50*(3), 405–435. https://doi.org/10.1007/s10657-020-09662-6

Green, R. (2024). FCC rules deep-fake robocalls like those received in NH are illegal. *TCA Regional News*.

Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial Intelligence Trust, Risk and Security Management (AI TRiSM): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications*, *240*, 122442. https://doi.org/10.1016/j.eswa.2023.122442

Hallowell, N., Badger, S., Sauerbrei, A., Nellaker, C., & Kerasidou, A. (2022). "I don't think people are ready to trust these algorithms at face value": Trust and the use of machine learning algorithms in the diagnosis of rare disease. *BMC Medical Ethics*, *23*(1), 1–112. https://doi.org/10.1186/s12910-022-00842-4

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*, *32*(4), 966–978. https://doi.org/10.1111/inm.13114

Köchling, A., & Wehner, M. C. (2023). Better explaining the benefits why AI? Analyzing the impact of explaining the benefits of AI-supported selection on applicant responses. *International Journal of Selection and Assessment*, *31*(1), 45–62. https://doi.org/10.1111/ijsa.12412

Konya, A., & Nematzadeh, P. (2024). Recent applications of AI to environmental disciplines: A review. *The Science of the Total Environment*, *906*, 167705–167705. https://doi.org/10.1016/j.scitotenv.2023.167705

Kreps, S., & Jakesch, M. (2023). Can AI communication tools increase legislative responsiveness and trust in democratic institutions? *Government Information Quarterly*, *40*(3), 101829. https://doi.org/10.1016/j.giq.2023.101829

Kruppa, M. (2024). Google Mired in Controversy Over AIChatbot Push. *Wall Street Journal*. https://www.wsj.com/tech/ai/google-mired-in-controversy-over-ai-chatbot-push-46023dd3

Lemieux, V. L., & Werner, J. (2024). Protecting Privacy in Digital Records: The Potential of Privacy-Enhancing Technologies. *Journal on Computing and Cultural Heritage*, *16*(4), 1–18. https://doi.org/10.1145/3633477

Liu, K., & Tao, D. (2022). The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Computers in Human Behavior*, *127*, 107026. https://doi.org/10.1016/j.chb.2021.107026

Mayopu, R. G., Wang, Y.-Y., & Chen, L.-S. (2023). Analyzing Online Fake News Using Latent Semantic Analysis: Case of USA Election Campaign. *Big Data and Cognitive Computing*, *7*(2), 81. https://doi.org/10.3390/bdcc7020081

Naiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2021). Explainable recommendation: When design meets trust calibration. *World Wide Web (Bussum)*, *24*(5), 1857–1884. https://doi.org/10.1007/s11280-021-00916-0

Noor, P. (2020). Can we trust AI not to further embed racial bias and prejudice? *BMJ : British Medical Journal (Online)*, *368*. ProQuest Central. https://doi.org/10.1136/bmj.m363

Odilla, F. (2023). Bots against corruption: Exploring the benefits and limitations of AI-based anti-corruption technology. *Crime, Law, and Social Change*, *80*(4), 353–396. https://doi.org/10.1007/s10611-023-10091-0

Poon, A. I. F., & Sung, J. J. Y. (2021). Opening the black box of AI-Medicine. *Journal of Gastroenterology and Hepatology*, *36*(3), 581–584. https://doi.org/10.1111/jgh.15384

Saba, C. S., & Pretorius, M. (2024). The impact of artificial intelligence (AI) investment on human well-being in G-7 countries: Does the moderating role of governance matter? *Sustainable Futures*, *7*, 100156. https://doi.org/10.1016/j.sftr.2024.100156

Sachan, S., Yang, J.-B., Xu, D.-L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications*, *144*, 113100. https://doi.org/10.1016/j.eswa.2019.113100

Sharma, S. (2024). Benefits or concerns of AI: A multistakeholder responsibility. *Futures : The Journal of Policy, Planning and Futures Studies*, *157*, 103328. https://doi.org/10.1016/j.futures.2024.103328

Thompson, S. (2012). *Regression Estimation* (pp. 115–124). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118162934.ch8

Tulk Jesso, S., Kelliher, A., Sanghavi, H., Martin, T., & Henrickson Parker, S. (2022). Inclusion of Clinicians in the Development and Evaluation of Clinical Artificial Intelligence Tools: A Systematic Literature Review. *Frontiers in Psychology*, *13*, 830345–830345. https://doi.org/10.3389/fpsyg.2022.830345

Xiang, H., Zhou, J., & Xie, B. (2023). AI tools for debunking online spam reviews? Trust of younger and older adults in AI detection criteria. *Behaviour & Information Technology*, *42*(5), 478–497. https://doi.org/10.1080/0144929X.2021.2024252

Zhang, R., Flathmann, C., Musick, G., Schelble, B., McNeese, N. J., Knijnenburg, B., & Duan, W. (2024). I Know This Looks Bad, But I Can Explain: Understanding When AI Should Explain Actions In Human-AI Teams. *ACM Transactions on Interactive Intelligent Systems*, *14*(1), 1–23. https://doi.org/10.1145/3635474

Zou, L., & Khern-am-nuai, W. (2023). AI and housing discrimination: The case of mortgage applications. *Ai and Ethics (Online)*, *3*(4), 1271–1281. https://doi.org/10.1007/s43681-022-00234-9