Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

BETWEEN LOGIC AND TRUST: HOW PEOPLE PERCEIVE AI IN HIGH-STAKES CONTEXTS

Muhammad Annas Khan Niazi[1], Muhammad Hasnain[*2], Aswad Sheeraz[3], Adam Azhar[4], Sabeeh Shahid[5], Wasiq Shahzad[6]

[1, *2,3,4,5,6]*Department of Computer Science, Beacounhouse College Programme, Multan, Pakistan*

[*2]khawajahasnain666@gmail.com

DOI: https://doi.org/10.5281/zenodo.18060469

Keywords
Artificial Intelligence, Trust in AI, Explainable AI, Human-AI Collaboration, Transparency, Algorithmic Fairness, Ethical Decision-Making

Article History
Received: 28 October 2025
Accepted: 12 December 2025
Published: 26 December 2025

Corresponding Author: *
Muhammad Hasnain

Abstract
This paper examines how individuals perceive and place trust in artificial intelligence (AI) in its role in making decisions in high-stakes fields such as healthcare, finance, education, and autonomous transport. The research was conducted using a structured, anonymous, scenario-based questionnaire that was self-administered by 197 individuals aged between 15 and 39, in order to understand participants' preference for human assistance over AI assistance in the decision-making process, as well as the impact of transparency on perceived trustworthiness. The findings reveal a strong preference for human supervision. A total of 69.5% of respondents believed that a senior surgeon without AI assistance was more trustworthy than a mid-level surgeon with AI assistance, and 72.6% of respondents preferred human–AI cooperation over fully autonomous financial management. Education was the only sector in which participants expressed trust in AI-based admissions systems, with 17.8% fully believing in such systems and 47.7% expressing partial or conditional belief. The findings further reveal that transparency is a significant factor influencing acceptance, as 76.6% of respondents indicated that they would place greater trust in AI systems if clearer explanations were provided. Overall, the results indicate that although people are aware of the efficiency of AI, they remain cautious about allowing it to operate independently and prefer systems in which automation is balanced with human responsibility.

1. Introduction

This study aims to examine the level of confidence individuals place in artificial intelligence (AI) in major, high-stakes decision-making contexts, particularly how AI is perceived in such serious situations. In recent years, artificial intelligence has begun to play an increasingly significant role in everyday life, influencing fields such as healthcare, finance, education, and the development of self-driving vehicles. However, AI has also attracted attention due to the risks it poses and, as a result, has not yet achieved the widespread acceptance that might be theoretically expected. By considering factors such as transparency, human supervision, specific contexts, and the interests involved, this paper seeks to determine whether contextual changes can help individuals place greater trust in AI. This research aims to lay a foundation for future work in the field of artificial intelligence by contributing to the development of systems that are perceived as more reliable and accurate.

1.1Background

In recent years, artificial intelligence (AI) has moved from being a background computational tool to becoming a central technology in

everyday life. AI now has the capacity to significantly alter individuals' lives by participating in decisions ranging from medical diagnoses to stock trading and university admissions. These systems are often praised for their precision, speed, and ability to reduce human bias. However, as AI's influence continues to grow, increased scrutiny and concern have followed, with many questioning whether a non-human agent can be trusted to make decisions that have the potential to affect health, safety, or opportunity. Attitudes toward AI and trust are known to vary widely; some individuals view algorithmic judgment, free from human biases, as more objective and fair, while others perceive it as impersonal, opaque, and potentially intimidating. As algorithmic systems increasingly shape organizational and societal decision-making in the 21st century, a key challenge lies in reconciling their efficiency and objectivity with human concerns about fairness, trust, and emotional impact (Lee). Therefore, it is important to understand how ordinary individuals respond to AI-driven decisions.

1.2 Problem Statement

As AI continues to be utilized increasingly in high-leverage situations, trust in AI remains inconsistent. While substantial research has focused on improving the technical accuracy, transparency, and explainability of algorithmic systems, comparatively less attention has been paid to how affected individuals perceive fairness, justice, and the human impact of algorithmic decisions (Binns et al.). In practice, this means that even when AI systems produce superior outcomes, individuals may still reject them if the decisions are perceived as unfair or lacking human judgment.

This pattern is already evident across society, as individuals from diverse age groups are increasingly likely to interact with, work alongside, and adapt to AI-driven systems. Public attitudes toward AI, both in the present and over time, will play a decisive role in shaping how these systems are adopted and how ethical guidelines are developed in the years to come. However, there remains limited empirical evidence demonstrating how the general public differentiates between levels of trust in AI across varying contexts. Understanding how people respond to the use of AI in different situations is therefore essential to predicting how society as a whole will adapt to automated decision-making in the future.

1.3 Research Objectives

This study aims to examine how individuals perceive the reliability and fairness of AI across a range of real-world contexts. More specifically, the study seeks to measure trust in four distinct fields: medicine, education, finance, and ethics. In addition, the paper analyzes the role of transparency to determine whether more open AI models inspire greater trust. The study further aims to examine the collected data in order to identify areas of doubt that remain influenced by skepticism. This paper provides meaningful data for policymakers and developers, which can be used to design AI systems that are more morally consistent with societal values and, by implication, more psychologically acceptable. These objectives integrate empirical evidence with design relevance. The findings are not intended to present differing viewpoints, but rather to support the informed development of AI governance frameworks that can foster trust among the general population.

1.4 Hypothesis

Building on early observations and contemporary discourse, the study proposes the following hypotheses:

- Respondents will demonstrate higher levels of trust when AI is paired with human oversight rather than granted complete autonomy.
- AI systems that demonstrate transparency in their decision-making processes will be trusted more than those that do not.
- Trust levels will vary significantly across situations involving moral dilemmas or domains in which human life may be at stake (e.g., medicine and autonomous driving), compared to more analytical and impersonal contexts such as finance.
- Demographic variables, such as field of study, will significantly influence levels

of trust, with individuals from disciplines such as engineering being more likely to trust AI systems.

These hypotheses further frame the inquiry by examining not only whether humans trust AI, but also the conditions under which such trust is sustainable.

2. Literature Review

2.1 Medicine: People Still Prefer Human Doctors

Despite evidence that AI-assisted doctors demonstrate higher surgical success rates, a large majority of individuals continue to place sole trust in human experts. This tendency is particularly evident in life-or-death scenarios. A survey conducted by Palo Alto, California–based Innerbody Research (2024) found that 52% of participants still preferred a human doctor over an AI system, even when the AI provided more accurate results (Stempniak). This finding further underscores the importance of emotional comfort and human judgment in shaping trust. Trust levels, however, tend to increase when a human is assigned to oversee an AI system. A 2025 report by WiseDocs found that trust increased from 16% to nearly 60% when humans were involved in the decision-making process (Wisedocs). This indicates that individuals are significantly more willing to trust AI systems when human validation is present.

2.2 Finance: People Want Both AI and Human Input

AI has the capability to predict profitable trades more accurately than human investors, yet many individuals continue to struggle with fully trusting it. Research indicates that people often exhibit "algorithm aversion," meaning they are reluctant to rely on AI even when it outperforms human judgment (Dietvorst et al.).

2.3 Education: Is AI Fair?

Previous research suggests that students often respond negatively to rigid algorithmic decision-making systems in educational contexts, particularly when recommendations are based on objective criteria such as grades, with many expressing reluctance or rejection of AI-based guidance (Das et al.). This finding aligns with ethics studies indicating that many individuals are concerned that AI systems lack empathy and may appear "cold" (Lee). However, providing clear explanations of AI decisions has been shown to foster greater trust, even in sensitive situations that can significantly impact an individual's life, such as college admissions (Binns et al.).

2.4 Explainability: Clear Explanations = More Trust

AI systems that provide intelligible and coherent explanations are likely to be perceived as more trustworthy compared to those that function as opaque black boxes. Prior research in explainable AI indicates that explainability enables users to develop more accurate mental models of system behavior, thereby enhancing perceptions of fairness, reliability, and competence (Hoffman et al.). Explainability helps users form sound mental representations of the decision-making process, which supports the perception of fairness, reliability, and competence. However, the mere presence of an explanation is insufficient; explanations must be clear, understandable, and relevant to users' needs. Unclear or poorly constructed explanations may undermine credibility by increasing uncertainty. Consequently, effective explainability is increasingly recognized as a key factor in fostering appropriate levels of trust and acceptance in AI-driven decision-making systems.

2.5 Self-Driving Cars: People Want Ethical AI, But Also Safety

In cases where an autonomous vehicle encounters an ethical dilemma, particularly situations in which harm is unavoidable—such as deciding whether to prioritize the safety of a passenger or a pedestrian—public opinion is considerably divided on how AI should make such decisions. Many respondents, and in some cases the majority, endorse a utilitarian ethic, a finding consistent with results from MIT's Moral Machine project (Awad et al.). However, these same individuals are often unwilling to ride in a vehicle programmed to sacrifice them. This illustrates a paradox: while people desire AI systems to behave morally, they simultaneously want to feel personally safe.

2.6 Oversight Matters Most

Across all domains, one factor remains consistent: individuals feel most comfortable when a human is in charge. As noted previously, Wisedocs' report found that trust in AI quadrupled when expert humans were assigned to work alongside the systems. These findings indicate a preference for hybrid systems combining AI and human oversight, rather than fully autonomous AI systems (Wisedocs).

3. Methadology

In the data collection process, the research employed a mixed-mode survey approach. A questionnaire was designed based on seven different scenarios, each intended to examine variations in trust depending on the context. The survey also collected demographic information from participants. It was administered both online and in person, using Google Forms and printed copies distributed to college students. The printed responses were subsequently uploaded to Google Forms to centralize the data. This approach allowed for broader access to students, including those who might have been reluctant to use online forms.

A total of 197 responses were collected, of which 196 were analyzed after eliminating one invalid response. Most respondents were late teenagers in high school and young adults attending university. The sample included a mix of male and female students from various academic streams. Participants ranged in age from 15 to 39 years, with a median age of 17, and the majority falling within the 16–19 age bracket, corresponding to the average age of high school students. Consequently, the analysis was conducted on two groups: individuals under 18 and those aged 18 or above. This grouping was intended to identify potential age-related trends and examine whether differences in opinions about AI exist between minors and adults.

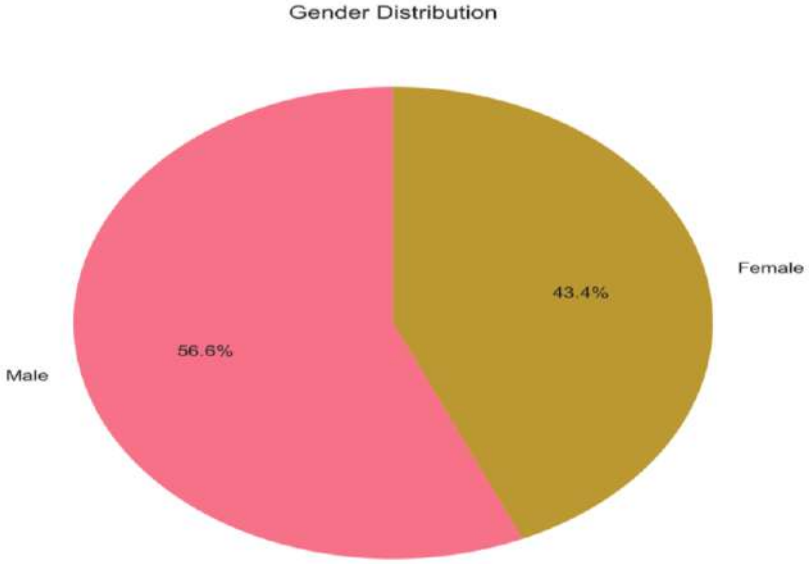

Figure 1: Gender distribution of the survey respondents (n=196). The sample was ~56.6% male and 43.4% female

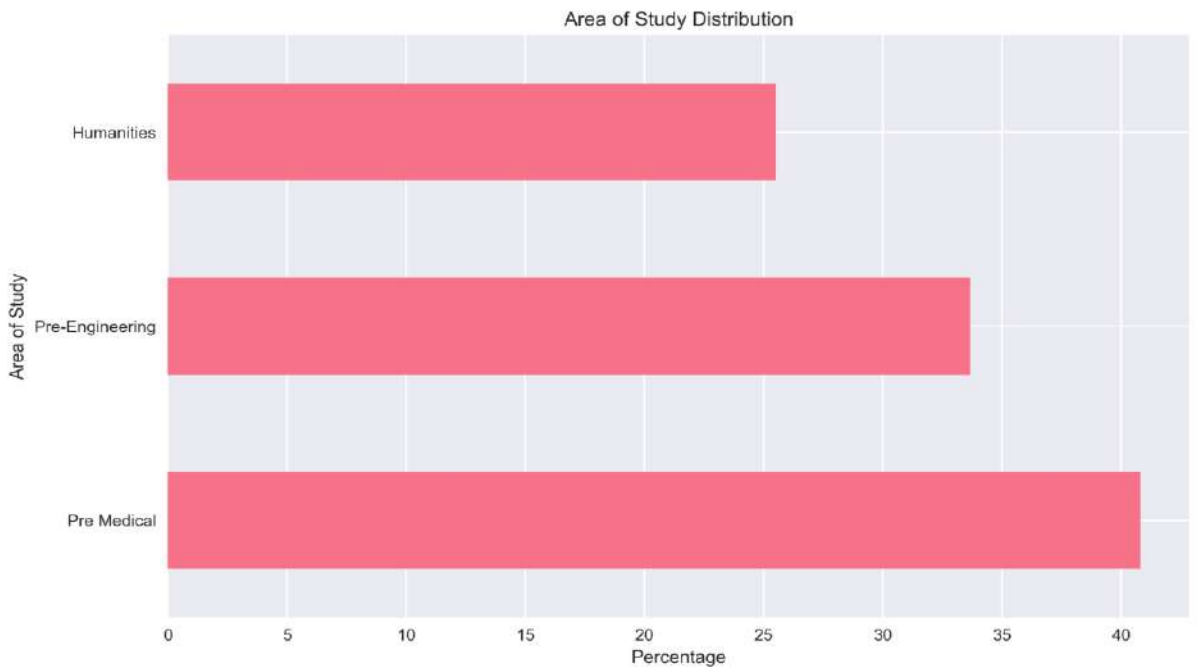

Figure 2: Academic background of respondents. Our sample covered 3 main study streams: ~41% from Medical, ~34% from Engineering (including those studying computer science), and ~25% from Humanities

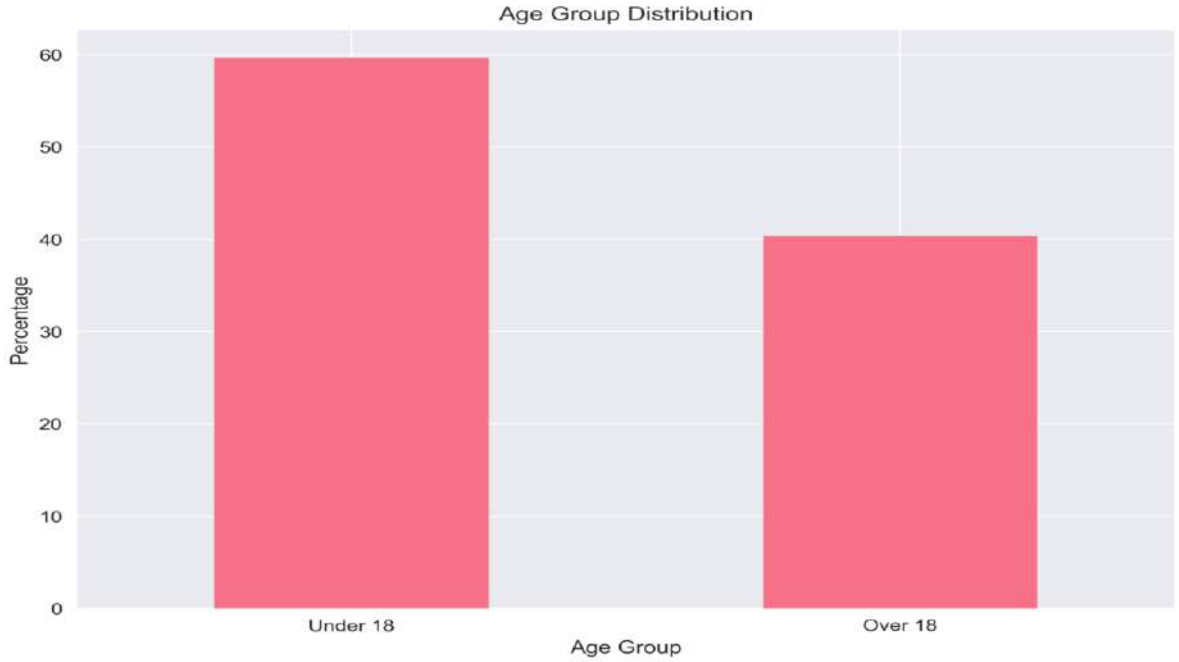

Figure 3: Age group breakdown of respondents. Around 60% were under 18, typical of high school students, the remaining 40% were 18 or older

Respondents' experience with AI systems was assessed to better understand their level of familiarity. This was measured using a scale ranging from 1 to 10, where 1–3 indicated "basic familiarity with ChatGPT or a similar widely available LLM" and 8–10 indicated being "comfortable building AI systems." The survey then presented various hypothetical real-world scenarios requiring trust in AI to make difficult, high-stakes decisions across domains such as medicine, finance, and education. Each scenario was followed by multiple-choice questions, asking, for example, whether respondents would trust AI over a human in situations such as a life-

saving surgery or whether an AI driving system should prioritize the safety of its passenger or a pedestrian. Participants were also asked whether greater transparency in AI decision-making would increase or decrease their trust in the system. The questionnaire was reviewed by peers to minimize ambiguity or bias and to ensure clarity.

The survey was conducted over the course of a week in late September 2025. Online respondents accessed the form via a shared link, while in-person participants completed physical questionnaires. Participation was entirely voluntary, and no personally identifying information was recorded, except for optional name and email fields used solely for validation and to prevent duplication; these were removed prior to analysis. Participants were instructed to answer honestly and based on their own perceptions, with reassurances that no responses were right or wrong.

Once collected and compiled, the data were exported to Google Forms for further analysis. Various descriptive statistical methods were employed to interpret the data and calculate response frequencies. Visualization tools, including Python's Matplotlib and Google Forms' built-in functions, were used to generate charts and graphs representing the data. In particular, the distribution of responses in each trust scenario was plotted, and visualizations were used to identify patterns among demographic groups. Cross-tabulations and percentage comparisons highlighted notable differences, such as whether males and females differed in their trust toward AI in specific situations, or whether science students were more trusting of AI than those in the humanities. These analyses were intended to identify general trends. Formal statistical significance testing was limited due to the small sample size; however, observed percentage gaps were noted as potential indicators of meaningful differences. All figures included in this paper were generated from the survey data and illustrate either the composition of the sample or response patterns related to trust.

4. Results

4.1 Participant Demographics

The study drew from a diverse demographic in terms of age, educational background, and gender. Participants ranged in age from 15 to 39 years. As shown in Figure 3, 60% of respondents were under 18, while the remaining 40% were 18 or older. Gender representation was relatively balanced, with 56.6% identifying as male and 43.4% as female. In terms of field of study (Figure 2), broad representation was achieved: medical students constituted the largest group (~41%), followed by engineering students (~34%) and humanities students (~25%). (The engineering category included students both with and without computer science in their course selection.)

Participants reported varying levels of experience with AI. Among those who responded, the majority indicated moderate familiarity with AI, such as regularly using AI chatbots like ChatGPT for academic and other purposes. Very few considered themselves highly experienced, with the ability to construct AI systems or modify existing frameworks to meet specific needs, and only a small number identified as complete beginners in operating AI systems. This contextual information is useful for interpreting participants' judgments and responses regarding AI trust.

4.2 Trust in AI vs Human Decisions Across Scenarios

Respondents were asked to imagine undergoing a high-risk medical procedure, with the choice between a senior human surgeon with a 95% success rate and no AI assistance, or an intermediate surgeon assisted by an expert AI system with a 99% success rate. Despite the higher success rate associated with the AI-assisted option, approximately 70% of respondents felt more comfortable entrusting their lives to the experienced surgeon without AI support, while only around 27% chose the intermediate surgeon with AI assistance. A small fraction (~3%) were unsure. These results indicate a clear trend of strong preference for human expertise over AI, even when statistical outcomes favor AI assistance, highlighting the cautious attitude many individuals adopt toward AI in high-stakes situations.

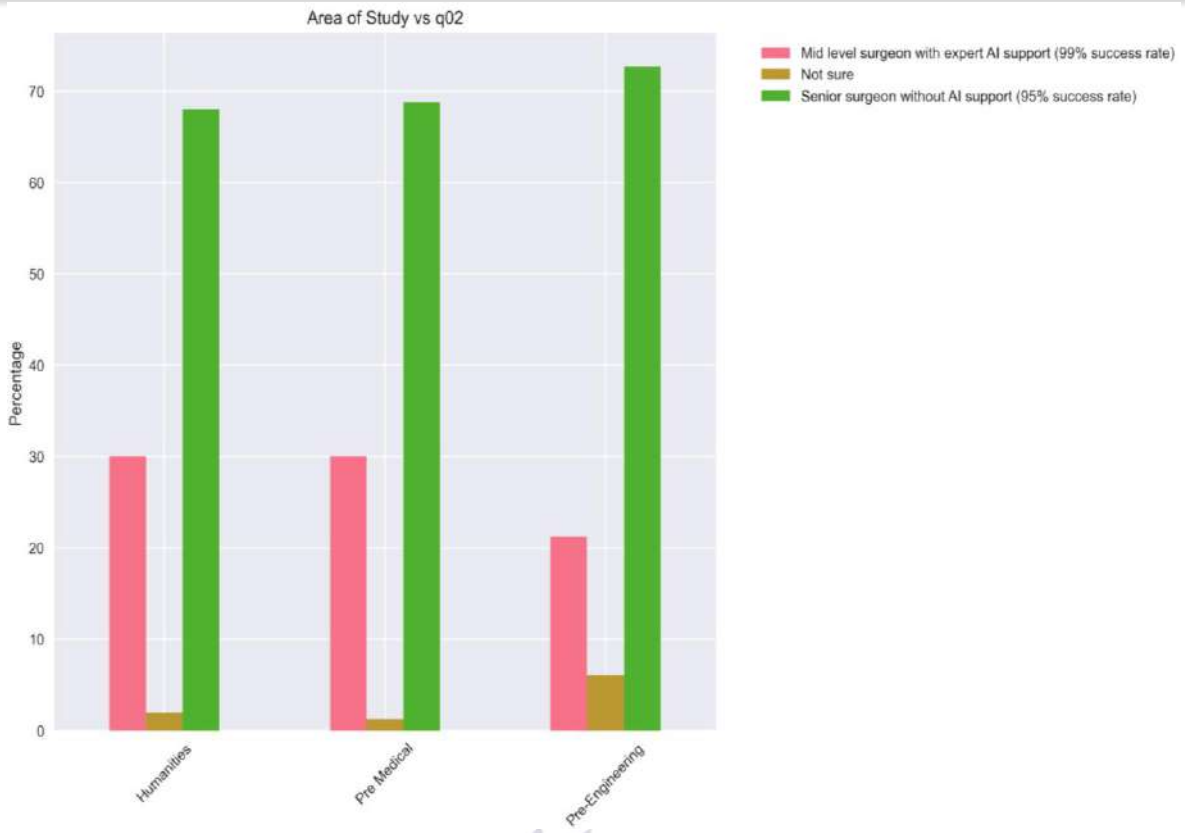

Figure 4: Trust preferences in a high-risk medical scenario. Most respondents chose the senior human surgeon despite the higher success rate of the AI-assisted option, reflecting continued caution toward AI in life-critical decisions.

In a different scenario, respondents were asked to decide how to manage their stocks, with options including an AI system with 85% accuracy, a human investor with 65% accuracy, or a combination of both. Surprisingly, 68% of participants chose the hybrid approach, combining the statistical strengths of AI with human intuition, indicating that while the majority recognize AI's superior ability to identify patterns in large datasets, they still prefer some human involvement. Approximately 17% of respondents indicated they would trust AI alone, while around 14% preferred to rely solely on human judgment, accepting a lower success rate of 65%. These results suggest a gradual development of trust in AI systems, with their usefulness being accepted but only to a limited extent.

Demographic differences were also observed. Female participants were more inclined to choose the hybrid option (79%) compared to male participants (60%). Conversely, male respondents were more likely than female respondents to rely entirely on AI (~22% vs. ~10%). These findings indicate that women were generally more cautious about allowing AI to manage their finances, although the combination of AI and human decision-making was preferred across all groups.

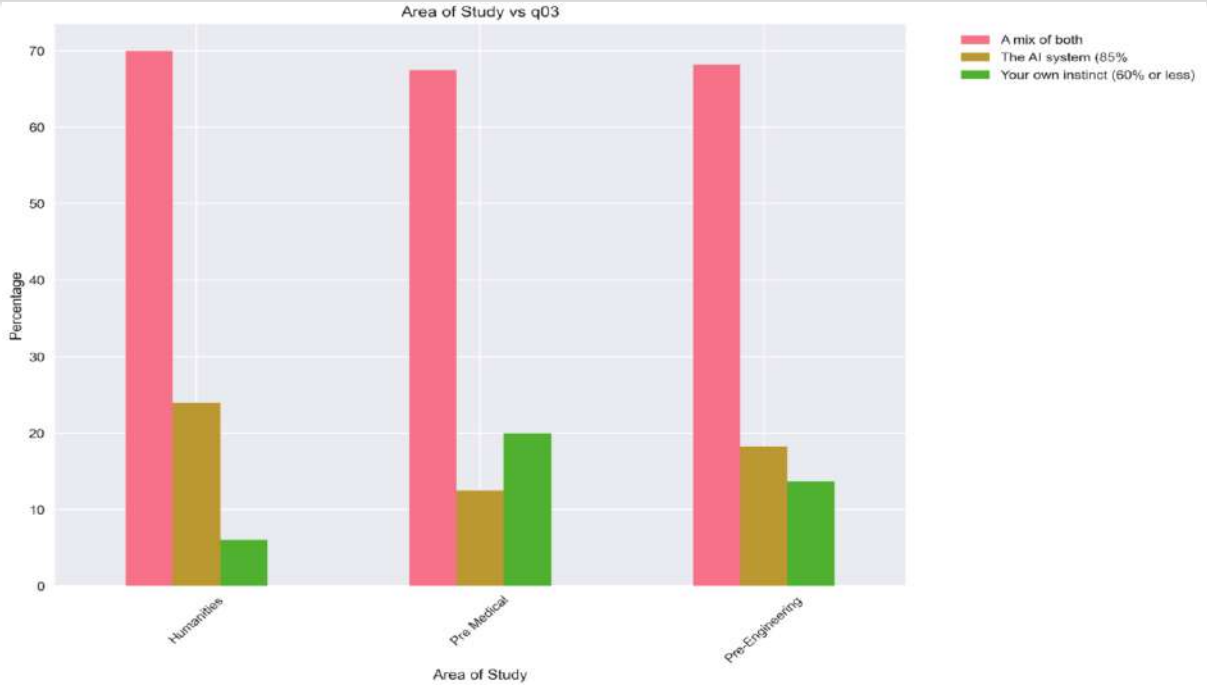

Figure 5: Trust in financial decision-making. A majority of respondents preferred a hybrid approach combining AI guidance with human judgment, showing partial trust in AI but reluctance to rely on it independently.

The next question focused on the perceived fairness of a rigid AI-driven decision system, specifically in the context of university admissions, where only students with grades ≥85% were accepted and all others were rejected without further consideration. Participants were asked to evaluate the fairness of such a system. This ethical dilemma left many respondents uncertain, with the most common response being "Maybe" (47%). Another substantial group responded with a firm "No," while only a small minority (~18%) fully trusted the system's fairness. Additionally, approximately 3% of participants were unsure and unable to provide a definitive response.

The data indicates that many individuals still harbor reservations regarding AI judgments in highly personal contexts, such as university admissions. Around 79% of respondents expressed concerns that the algorithm might be overly rigid, potentially overlooking important qualities that a human admissions officer would consider, such as a student's personal challenges or background. Differences were also observed by field of study: humanities students were more reluctant to endorse or trust AI systems in such ethical matters compared to science students. Overall, the findings reflect a broader hesitation to trust AI in ethical contexts where human compassion and emotional intelligence—areas in which AI is currently limited—play a critical role.

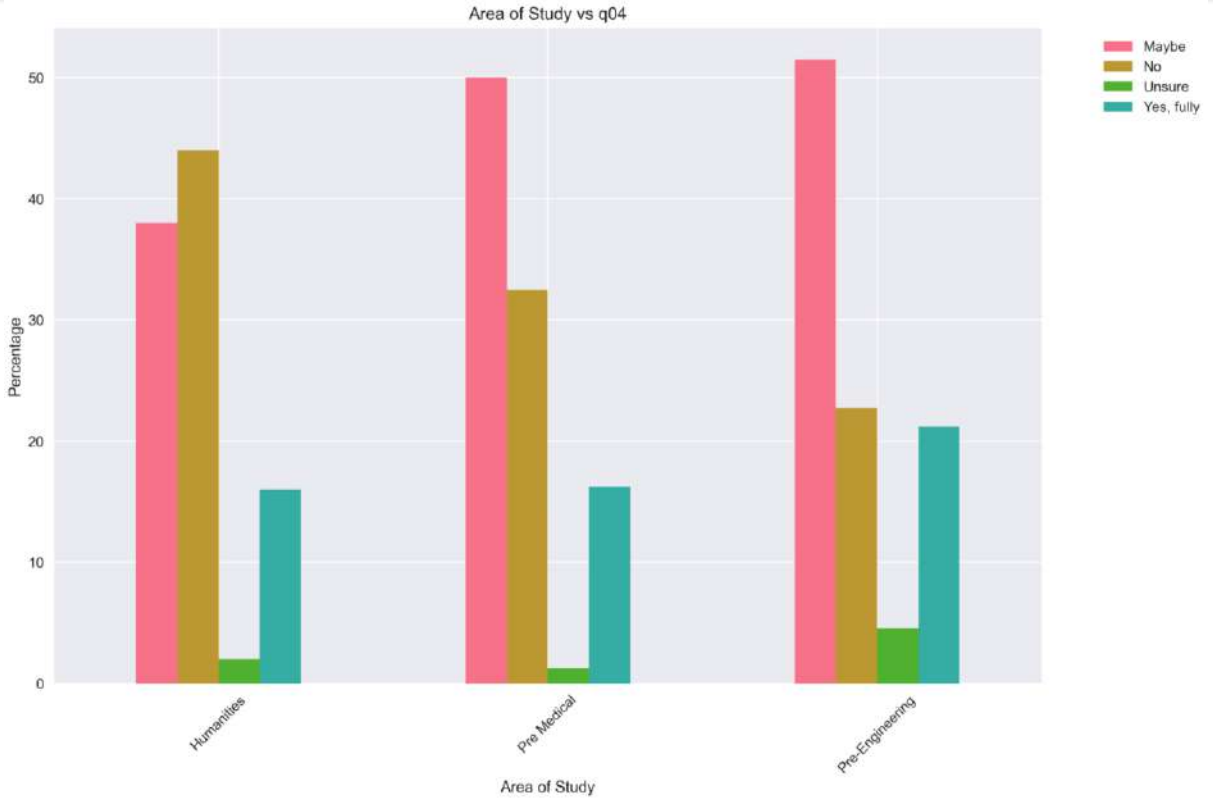

Figure 6: Perceived fairness of an AI-based admissions system. Nearly half of respondents expressed uncertainty, with many concerned that a rigid AI cutoff lacks empathy and contextual understanding.

The next scenario presented an ethical dilemma in which an AI driving system must choose whether to prioritize the safety of its passenger or that of a pedestrian. In this situation, if the AI prioritizes the passenger, the pedestrian will be harmed, and vice versa, creating an effectively no-win scenario. Participants were asked which choice they believed AI should make in such a dilemma. The majority (67%) of respondents trusted the AI to prioritize the pedestrian's safety over that of the passenger, while the remaining 33% believed the AI should prioritize the passenger. This 2:1 ratio suggests a preference for a utilitarian approach, with insights from separate discussions indicating that the risk to the passenger is implicitly accepted by choosing to ride in an AI-operated vehicle, whereas the pedestrian has no stake in the decision and should be spared.

It is important to note that those who favored passenger safety may perceive prioritizing the pedestrian as equivalent to the AI deliberately harming its passenger, which could cause unease regarding the dilemma. Interestingly, unlike prior scenarios, minimal variation was observed across genders or fields of study, with all groups reporting approximately 65–70% in favor of saving the pedestrian. These results indicate that the general population trusts AI to make ethically appropriate decisions in such dilemmas, rather than choosing self-serving options. However, further discussions regarding AI morality are necessary before establishing standardized ethical guidelines, as complex dilemmas, such as the "Trolley Problem," continue to challenge moral reasoning.

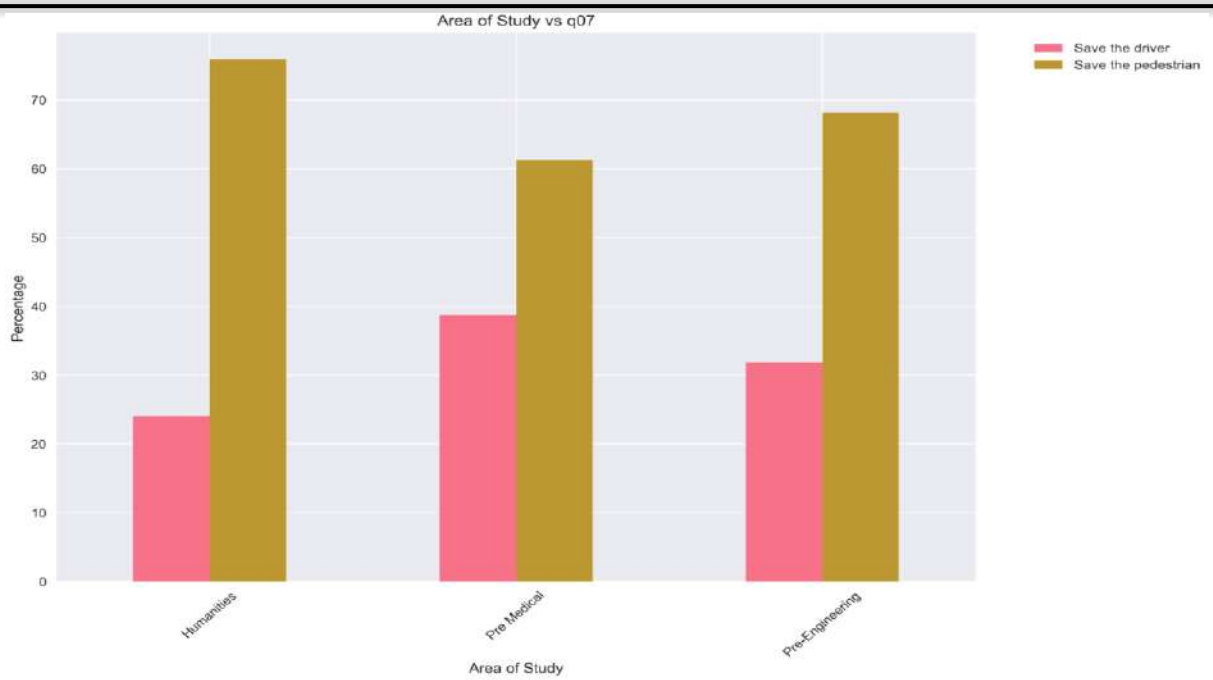

Figure 9: Ethical trust in an autonomous vehicle dilemma. A majority believed the AI should prioritize the pedestrian over the passenger, suggesting support for utilitarian decision-making in moral conflict scenarios.

The next survey question examined whether providing explanations of AI decision-making would influence respondents' trust in the system. The responses revealed a clear trend: both male and female participants reported that understanding the reasoning behind AI decisions would increase their trust. Specifically, 22% of respondents indicated that explanations would "strongly increase" their trust, while an overall majority of 76% stated that explanations would "increase" their trust in AI. A small minority (~3%) reported that clear explanations could actually reduce their trust. This group may have concerns that revealing AI's reasoning could expose flaws or biases, thereby undermining confidence in the system.

Overall, these findings indicate that transparency and clarity in AI decision-making enhance user comfort and willingness to rely on AI systems. This question highlights a clear need and demand for an AI transparency mandate, which could foster greater trust in AI-driven decisions.

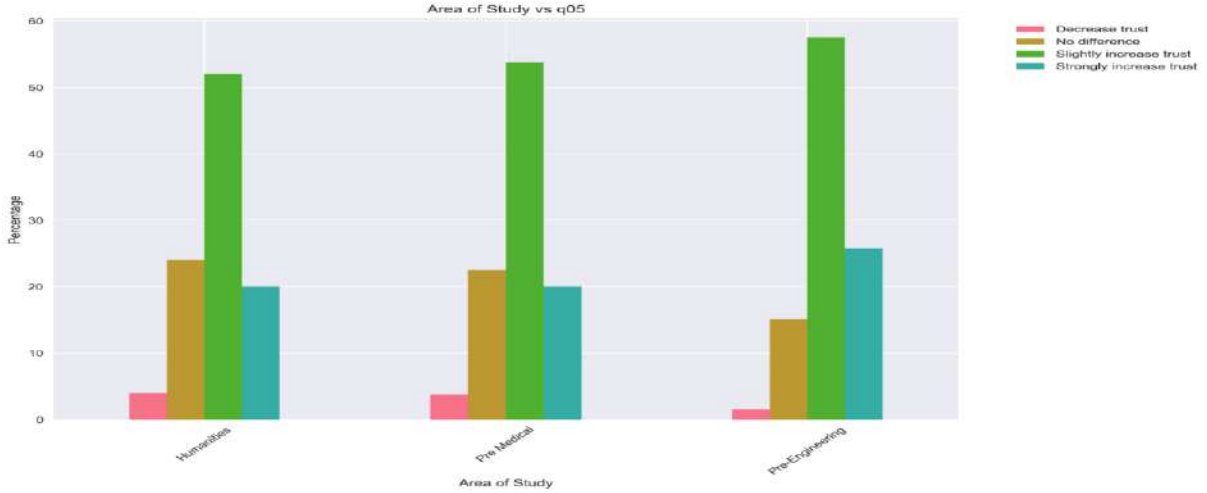

Figure 7: Impact of transparency on AI trust. Most respondents reported that clear explanations of AI decisions would increase their trust, highlighting the importance of transparency in high-stakes applications.

Finally, participants were asked whether they believe AI—with explanations and reasoning provided for its decisions—should play a greater role in environments such as finance, medicine, or education. Consistent with trends observed in previous questions, the majority opted for a compromise, supporting AI use only with strict human oversight. Specifically, approximately 72% of respondents answered, "Yes, but only with human oversight," indicating that while people recognize the utility of AI systems, they also acknowledge the limitations of these systems and their inability to fully replace human intuition and decision-making. Additionally, 14% of respondents answered, "Yes, definitely," expressing support for AI operating independently in major fields, while the remaining 14% were strongly opposed to AI integration in high-stakes scenarios. Overall, around 86% of respondents demonstrated cautious optimism toward AI and its future societal role, yet they did not believe AI is sufficiently developed to fully replace human judgment. Another notable trend was that virtually no engineering students opposed AI integration. Nonetheless, the majority across all groups favored the option of "only with human oversight," highlighting ongoing concerns regarding the deployment of AI in critical tasks.

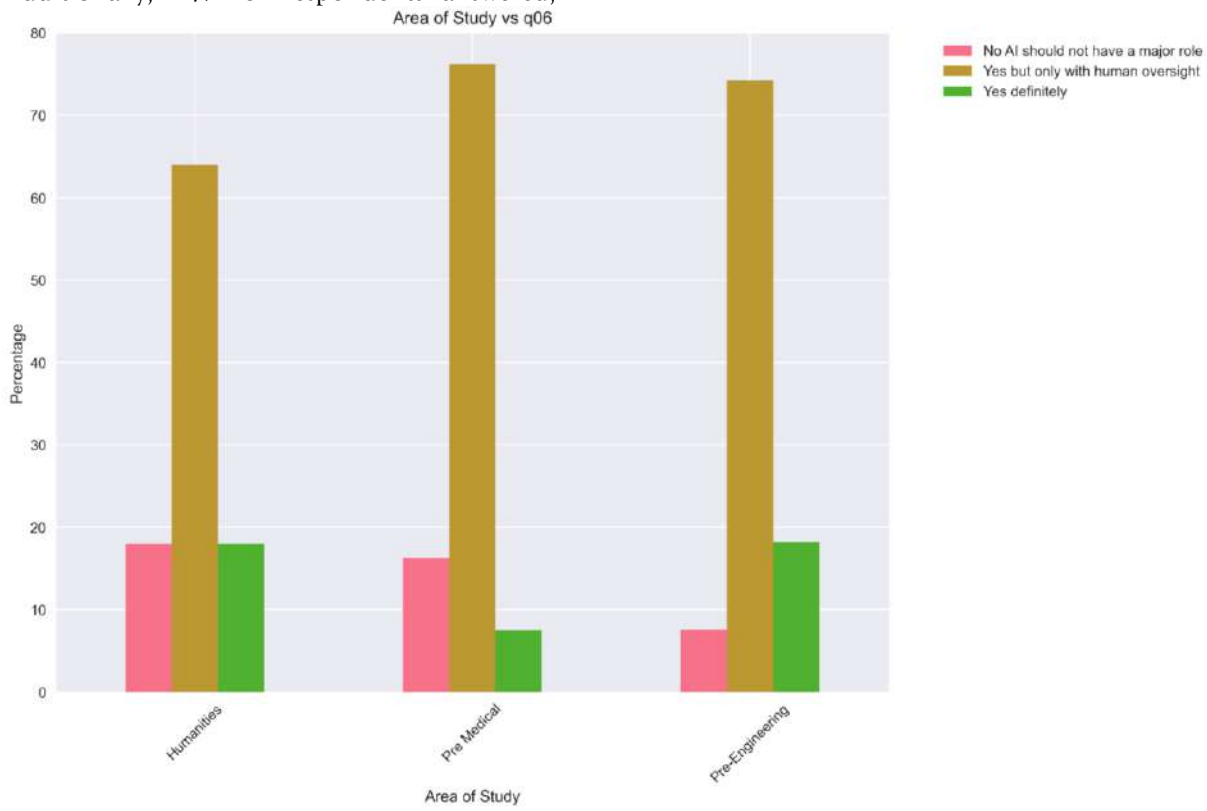

Figure 8: Acceptance of AI in critical fields. While many respondents supported AI involvement in areas such as healthcare and finance, most insisted on human oversight, indicating a preference for shared control rather than full autonomy.

Overall, these results present a nuanced view of public opinion regarding AI in high-stakes situations. In summary, participants exhibit cautious optimism toward AI, while still emphasizing the necessity of human supervision, particularly in critical scenarios. Regarding transparency, providing clear explanations and reasoning for AI decisions appears to enhance acceptance and facilitate the integration of AI into daily life. In ethical dilemmas, participants generally prefer that AI make the morally correct choice, although real-world applications of these preferences may vary. The discussion section will interpret these findings within a broader context and compare them with existing research.

5. Discussion

The survey's findings illustrate a cautiously open approach to AI. Respondents recognize AI as a capable tool with significant advantages over humans in certain areas, such as pattern recognition; however, they do not fully trust it and remain skeptical about relinquishing complete control in critical matters. For the time being, participants prefer AI to assist human experts rather than replace them entirely. While some domains are more conducive to AI adoption than others, each scenario presents its own nuances. Several key insights emerge from the research.

In the high-stakes medical scenario (Figure 4: Trust in AI for High-Leverage Situations), a clear majority (70%) chose the experienced human surgeon over the AI-assisted surgeon, despite the latter's higher statistical success rate. This highlights a common skepticism toward AI, particularly in life-or-death contexts. These results align with findings from Palo Alto, California-based Innerbody Research (2024), which reported that 52% of respondents preferred a human doctor over AI in medical decision-making (Stempniak). In our study, the higher preference for the human surgeon may be attributed to the framing of the scenario as "trusted doctor vs. unknown doctor with AI assistance" rather than "AI vs. no treatment." External research also indicates that age and digital literacy can influence trust in AI, with younger generations generally more open to AI in medicine. For instance, in a prior survey, nearly 82% of Gen Z participants chose an AI diagnosis over a human's, compared to only 57% of baby boomers. While the majority of our Gen Z sample still favored the human doctor, the 27% who selected AI may reflect this broader trend (Stempniak). As AI continues to prove its reliability, trust in high-stakes and everyday applications is likely to grow, although human oversight remains important for reassurance. This is supported by the Zendesk Global Consumer Survey (2025), which found that while AI is acceptable for routine tasks, over 55% of respondents prefer human intervention in urgent or stressful situations, such as critical medical care (Zendesk Global Survey).

One of the most prevalent trends in the data is a preference for the "human-in-the-loop" model. In high-stakes fields, approximately 72% of respondents supported AI use only with human oversight. Industry research, such as Wisedocs' 2025 report on AI in insurance claims, found that professionals' trust in AI increased from 16% to 60% when expert humans reviewed AI decisions—nearly a fourfold increase (Wisedocs). Participants in our survey implicitly agreed with this sentiment, acknowledging AI's superior processing capabilities and pattern recognition while emphasizing the need for human supervision due to AI's lack of emotional intelligence and common sense. This aligns with global research indicating that AI is more trusted when humans are available to intervene (Bach et al.). In practice, this could involve AI suggesting potential courses of action with final approval by a human expert, combining AI's speed and accuracy with human reasoning.

Our survey also revealed a clear correlation between trust in AI and understanding its reasoning. Approximately 75% of respondents stated that their trust would increase if AI provided clear explanations for its decisions. This supports ongoing research in Explainable AI (XAI) and prior findings, such as McKinsey's 2024 report, which emphasizes that understanding the rationale behind AI decisions increases acceptance (McKinsey & Company). However, the quality of explanations is crucial; confusing or superficial explanations may reduce trust. Developers of AI for high-stakes situations, such as medical diagnosis or loan approvals, should prioritize making system reasoning traceable and transparent. A small number of respondents (~3%) reported that explanations could decrease their trust, potentially due to exposure of AI's flaws. Transparency about limitations and uncertainty may, however, increase trust overall, as openness generally fosters confidence in AI systems.

Concerns regarding rigid algorithms, such as the "85% grade cutoff" for university admissions, highlight the challenge of AI fairness. Many respondents answered "maybe" or "no," indicating skepticism that strict rule-based systems can achieve fairness. Humans value context, emotions, and exceptions in decisions that significantly impact lives, such as college admissions or legal hearings. While rule-based AI may provide consistency, it may lack justice

and empathy. A small minority (18%) trusted AI entirely, potentially reflecting a preference for meritocratic, bias-free decisions. Research suggests some individuals prefer algorithmic decisions over human judgment to avoid emotional bias (Chugunova and Luhan). Nonetheless, most respondents favored a humanized approach or hybrid model to account for fairness and context, which AI developers must consider in sensitive domains.

The autonomous vehicle scenario provided insights into public expectations of AI in ethical dilemmas. Approximately two-thirds of participants believed AI should spare the pedestrian at the expense of the passenger, reflecting a utilitarian approach to minimize overall harm. One-third opted to prioritize the passenger, demonstrating the subjective nature of ethics influenced by culture and upbringing. Trust in AI, therefore, depends not only on accuracy and safety but also on moral alignment. AI developers must consider both transparency and ethical principles in system design. Younger respondents in the study indicated that socially responsible AI is more trustworthy, aligning with research suggesting that AI acceptance will hinge on ethical and accountable behavior (Fel et al.).

Demographic patterns also emerged. Female respondents were generally more cautious about granting AI full autonomy, particularly in university admissions and financial decisions, consistent with prior research showing women exhibit lower trust in autonomous systems (Dávila et al.). Age differences were observed as well: older participants, many of whom were recent graduates, were slightly more optimistic about AI in high-stakes scenarios and more accepting of rigid algorithmic decisions, potentially due to greater familiarity and understanding of AI. Field of study influenced trust, with engineering students more supportive of AI and humanities students more skeptical, particularly in scenarios involving ethics or high-risk decisions. Across all groups, however, participants recognized AI's capabilities while emphasizing the need for human intervention to compensate for its lack of empathy and emotional intelligence.

These findings align with global research, including a 2025 global AI study by the University of Melbourne and KPMG, which found that although many people regularly use AI, less than half of respondents were willing to trust AI systems — underscoring trust as a critical challenge for broader adoption (University of Melbourne & KPMG). Trust increases when conditions such as security, transparency, and human oversight are met, a pattern corroborated by our data. Similarly, studies by Forbes reported that 64% of consumers would trust an AI diagnosis over a doctor's, though differences in context, AI domain (e.g., radiology vs. surgery), and participant demographics explain variations in trust levels. As younger, tech-savvy generations become the majority, trust in AI may continue to rise as familiarity and technology adoption increase (Hsieh).

In conclusion, trusting AI is not a binary decision; context matters, and each case presents unique trade-offs between human intuition and AI's computational power. The findings suggest that participants will trust AI when it is transparent, supervised by humans, and aligned with ethical principles. Developers and stakeholders should prioritize explainability, hybrid decision-making models (AI + human), and rigorous ethical checks to bridge the trust gap and facilitate AI adoption in critical domains. Expert oversight has been shown to quadruple trust, highlighting that the public does not inherently distrust AI but expects responsible deployment.

Over time, public perceptions of AI are likely to evolve as systems become more transparent and the population becomes better informed about their capabilities. While fears associated with AI may diminish with exposure, high-profile failures could negatively impact trust and slow adoption. The cautious optimism observed among younger participants reflects recognition of AI's potential, tempered by ethical concerns and the need for responsible integration. The following section will discuss the limitations of this study and avenues for future research.

6. Limitations

While this study provides valuable insights into current perceptions of AI, several limitations and shortcomings must be acknowledged.

- Sample Bias: The sample was primarily drawn from a single high school, with

most participants in their late teens. A small minority of respondents (ages 30–39) were included, but their representation was insufficient to generalize findings to the broader population. Views of high school students may differ significantly from those of older adults who are more familiar with or even involved in AI development. Additionally, educational background and cultural factors may influence perceptions of AI; since the study was conducted in a relatively small geographic area (Multan, Pakistan), attitudes in other regions may vary.

- Limited Question Response: Question 1 (AI familiarity) was omitted in some early survey forms, leading to missing data for a portion of respondents. Some participants also skipped certain questions, introducing uncertainty and potential bias in interpreting trends.
- Hypothetical Nature of Questions: All survey questions were hypothetical and carried no real-world consequences. This may produce discrepancies between stated preferences and actual behavior. For example, participants who indicated they would save a pedestrian in an autonomous driving scenario might act differently if personally at risk. Similarly, trust in AI diagnoses may not translate into real-world reliance on AI without seeking a human second opinion. The intent–behavior gap is a common limitation in survey research.
- Survey Framing and Information: Scenarios were simplified and provided limited context. Trust decisions could depend on additional details, such as the surgeon's full experience or the mechanics of an AI system. Framing effects, such as presenting "senior surgeon vs. mid-level surgeon with AI," may have biased responses. Ambiguities in question phrasing (e.g., "Which decision would you trust the AI to make?") could also influence responses. Different wording or additional context might yield different outcomes.
- Lack of Deep Psychological or Qualitative Insights: The study was purely quantitative and did not include open-ended questions or interviews. While inferences about reasoning were drawn in the discussion, respondents' motivations, emotions, and thought processes were not directly captured. Understanding why 70% distrusted an AI-assisted surgeon, for example, would provide richer insight. Qualitative data would strengthen interpretations of trust and skepticism.
- Limited Statistical Depth: Due to the sample size and scope, analyses were mostly descriptive. Advanced statistical tests (e.g., chi-square tests, regression analysis) were not performed, limiting the ability to identify significant differences or control for confounding factors. Observed trends, such as minor gender differences, should therefore be interpreted cautiously.
- Cross-Sectional Design: Data was collected at a single point in time (late 2025). The study does not account for temporal changes in trust or external events (e.g., high-profile AI failures) that may influence opinions. A longitudinal approach would be necessary to understand how trust in AI evolves over time and with experience.

Given these limitations, findings should be interpreted as insights into the sample rather than a definitive measure of broader public opinion. Future research could address these constraints by sampling a larger, more diverse population, incorporating qualitative methods (e.g., interviews), and experimenting with different question framings to assess the impact on responses.

Despite these limitations, the study provides a valuable starting point for understanding perspectives on AI in high-stakes situations. Observations regarding the importance of oversight and transparency can inform future research and policy. Additionally, these findings align with broader academic surveys, offering local validation of global trends in public trust toward AI.

7. Conclusion

The findings of this study reveal that trust in AI is not a binary concept; rather, it exists across a spectrum and is highly context-dependent. Trust must be earned over time, particularly in high-stakes situations. The study sought to explore the conditions under which people are willing to rely on AI, with the overarching conclusion being: "AI will be trusted, but only when certain conditions are met."

Key conclusions include:

- Human Involvement: Respondents demonstrated greater comfort with AI when a human expert was involved in the decision-making process. The most acceptable deployment of AI, at least for the foreseeable future, appears to be as an assistant to human experts rather than as a fully autonomous agent. This hybrid approach leverages the precision and speed of AI while retaining the judgment and oversight of humans, mitigating concerns regarding errors and accountability.
- Importance of Transparency: A clear majority of participants indicated that understanding the reasoning behind AI decisions enhances trust. This finding reinforces the critical role of Explainable AI (XAI), particularly in high-stakes scenarios. Users prefer AI systems that are transparent rather than "black boxes," and comprehension of AI reasoning significantly increases acceptance. Developers should prioritize integrating features such as decision logs and clear explanations to overcome potential trust barriers, regardless of AI accuracy.
- Preference for Human Judgment in Life-or-Death Scenarios: In situations involving potential loss of life, intangible qualities such as accountability, empathy, and intuition built on experience were consistently valued over marginal statistical improvements offered by AI. Psychological comfort plays a substantial role, suggesting that AI should function as a support tool rather than a replacement in these scenarios. While public perception may evolve alongside advancements in AI, trust will initially require substantial demonstration of reliability combined with human oversight.
- Cautious Optimism in Other Domains: In less personal or lower-stakes contexts, such as financial decision-making, participants were more willing to rely on AI but generally preferred combining it with human input. Partial trust in AI reflects recognition of its superior data-processing and pattern recognition capabilities, while maintaining reliance on human intuition. Collaborative systems that enable interaction between AI and human users align with these preferences and are likely to be better received.
- Ethical Alignment Matters: Responses to moral dilemma scenarios indicate that the public holds AI to strict ethical standards, expecting it to act in morally appropriate ways. Developers must consider ethical alignment as central to fostering trust, ensuring that AI behavior corresponds with societal norms and values. Transparent ethical policies, particularly in applications such as autonomous vehicles, are critical to public confidence.

In conclusion, the study highlights an optimistic yet cautious trajectory for AI adoption. While participants recognize the potential of AI to improve outcomes in critical situations, there is a consistent demand for accountability, transparency, and human oversight. Trust in AI is conditional upon its alignment with human values, ethical behavior, and explainable decision-making. Developers should prioritize these aspects, incorporating domain experts in the development process and focusing on transparency, fairness, and oversight. Policymakers should consider mandates that enforce transparency and ethical compliance for AI in high-risk domains.

Finally, it is important to note that trust is dynamic and may evolve as individuals become more familiar with AI systems. Wider public

exposure, positive experiences, and consistent demonstration of reliability will likely increase trust, while high-profile AI failures could have a negative impact. Continuous engagement with the public, through surveys and other mechanisms, provides developers with insights into user concerns and expectations.

Overall, trust in AI is essential for its integration into everyday life. Establishing transparency, oversight, and ethical alignment is the foundation for gaining public confidence. Only after trust is established can AI realize its full potential in high-stakes domains and everyday applications.

Acknowledgements

We would like to express our sincere gratitude to Mehvish Shafiq, PhD, for her invaluable support in reviewing our research paper and providing insightful feedback that greatly enhanced its quality.

We also extend our heartfelt thanks to the Beaconhouse College Program, Multan, and specifically Principal Sumera Nasheed, for their cooperation and assistance in facilitating the survey in their school. Their support was instrumental in enabling the successful collection of our data.

Finally, we are grateful to Sabeeh Shahid, who not only helped us connect with the Beaconhouse administration and coordinate our meeting with the principal, but also served as our mentor throughout the project, guiding us through every stage of this research.

REFERENCES

Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance." Frontiers in Computer Science, vol. 5, 06 Feb. 2023, https://doi.org/10.3389/fcomp.2023.1096257

Lee, Min Kyung. "Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management." Big Data & Society, vol. 5, no. 1, 2018, pp. 1–16, doi:10.1177/2053951718756684.

Binns, Reuben, et al. "'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–14.

Stempniak, Marty. "Patients May Soon Trust AI More Than Doctors." Radiology Business, 31 May 2024, RadiologyBusiness.com.

Wisedocs. AI in Claims: The 4x Trust Effect of Human Oversight. Wisedocs Research Report, 2025. This report highlights how combining artificial intelligence with expert human oversight significantly increases trust among claims professionals, www.wisedocs.ai/reports/2025-ai-in-claims-survey

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." Journal of Experimental Psychology: General, vol. 144, no. 1, 2015, pp. 114–126.

Awad, Edmond, Sohan Dsouza, Rakesh Kim, Johannes Schulz, Joseph Henrich, Azim Shariff, et al. "The Moral Machine Experiment." Nature, vol. 563, no. 7729, 2018, pp. 59–64.

Das, Hemanshu, Sofoklis Goulas, and Faidra Monachou. "Why Students Reject AI for Human Counselors in College Applications: A Field Experiment." SSRN Electronic Journal, 2 June 2025, https://ssrn.com/abstract=5282793

Chugunova, Marina, and Wolfgang J. Luhan. "Ruled by Robots: Preference for Algorithmic Decision Makers and Perceptions of Their Choices." Public Choice, vol. 202, no. 1, 2025, pp. 1–24, https://doi.org/10.1007/s11127-024-01178-w

Fel, Stanisław, Jarosław Kozak, and Piotr Horodyski. "Responsibility and AI: Exploring Technology Acceptance Models." Journal of Innovation & Knowledge, vol. 10, no. 6, 2025, article 100852, Elsevier, https://doi.org/10.1016/j.jik.2025.100852

Dávila, Antonio, Ismail El Fassi, Daniel Oyon, and Nicolas Rudolf. "Gender, Knowledge, and Trust in Artificial Intelligence: A Classroom-Based Randomized Experiment." Scientific Reports, vol. 15, article 41066, 2025, https://doi.org/10.1038/s41598-025-25002-7

Global survey reveals growing consumer trust in personal AI assistants. Zendesk Global Survey, commissioned by Zendesk and conducted by YouGov, 30 July 2025, Zendesk.com.

Bach, Tita Alissa, et al. "A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective." arXiv Preprint, 18 Apr. 2023, arXiv.org/abs/2304.08795.

Building AI trust: The key role of explainability. McKinsey & Company, 26 Nov. 2024, McKinsey.com.

Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025. University of Melbourne and KPMG, 2025. Report.

Hsieh, Paul. "Patients May Soon Trust Artificial Intelligence More Than Humans." Forbes, 30 June 2024, forbes.com/sites/paulhsieh/2024/06/30/patients-may-soon-trust-artificial-intelligence-more-than-humans/.