

The Cognitive Foundations of Algorithmic Trust: Human Reasoning in High-Stakes AI Systems

Sonrisa Watts*

Independent Researcher / Emotion Encoded

October 26, 2025

Abstract

Artificial intelligence (AI) is increasingly integrated into professional and personal contexts, yet its adoption in high-stakes sectors such as healthcare, law, and finance remains constrained by human trust, affective responses, and expectations of accountability. This mixed-methods preprint presents *preliminary findings* from an ongoing study examining cognitive, emotional, and ethical determinants of AI trust. The dataset combines two survey samples ($N_1 = 69$; $N_2 = 44$) with qualitative interviews from fifteen domain experts in law, medicine, and finance. Quantitative analyses, including logistic regression and descriptive modeling, indicate that prior AI experience, perceived transparency, and emotional comfort are key predictors of trust and acceptance across decision-making domains. Familiarity consistently enhances calibrated confidence, while transparency and perceived accountability strengthen trust in financial and professional contexts. Conversely, older participants exhibited more cautious attitudes toward AI autonomy, suggesting that demographic and affective factors jointly shape trust formation. Collectively, these findings highlight that algorithmic trust evolves through interactional, cognitive, and emotional processes rather than purely technical performance.

The qualitative component enriches these patterns through seven recurring themes: emotional reasoning, meta-cognition, AI-augmented creativity and innovation, algorithms codifying injustice, radical skepticism, heuristics and cognitive biases, and algorithmic transparency and explainability.. Together, these findings illuminate how trust is negotiated through both cognitive and affective appraisals, users weigh perceived fairness, empathy, and accountability rather than technical accuracy alone. Experts across sectors consistently framed AI as a subordinate decision-support system, emphasizing the continued primacy of human oversight and moral responsibility.

Collectively, these *preliminary findings* indicate that trust in AI is neither static nor purely rational but develops iteratively through exposure, perceived social value, and institutional safeguards. The results underscore the importance of explainability, ethical alignment, and participatory design in building public confidence. Future work will expand the dataset and employ cross-cultural comparisons to test the generalizability of these early insights.

Keywords: algorithmic trust, cognition, transparency, accountability, affective reasoning, artificial intelligence, cognitive bias, heuristics, high stakes fields, emotion, psychological resistance

*Corresponding Author: sonrisawatts@gmail.com; Phone: +1 869-661-1404

Table of Contents

1	Introduction and Literature Review	3
2	Methods	5
2.1	Ethical Considerations and Participant Identification	5
2.2	Quantitative Data and Sampling	6
2.3	Qualitative Data and Expert Recruitment	6
2.4	Reproducibility and Data Availability	7
3	Results	8
3.1	Overview of Findings	8
3.2	Quantitative Data	8
3.3	Logistic Regression: AI Attitude Predictors	13
3.4	Logistic Regression: Inferential Analysis:	13
3.5	Qualitative data	13
3.6	Thematic Analysis of Interviews with Domain Experts	14
4	Discussion	15
4.1	Cognitive, Professional, and Ethical Dimensions of Trust in AI	15
4.2	Healthcare: Knowledge, Attitudes, Acceptance, and Application	16
4.3	Generative AI in Medical Decision-Making and Peer Perception	16
4.4	Primary Care and Patient Portal Messaging	17
4.5	Integration of Knowledge, Attitude, and Peer Perception	17
4.6	Law: Legitimacy, Accountability, and Trust Calibration	18
4.7	Finance: Predictive Reliability, Efficiency, and Ethical Risk	18
4.8	Cross-Domain Synthesis: Toward Context-Specific Governance	19
5	Limitations	19
6	Conclusion	19
7	Acknowledgments	20

1 Introduction and Literature Review

Artificial intelligence (AI) systems are increasingly deployed in high-stakes contexts such as medicine, law, and finance, where algorithmic decisions can directly affect health outcomes, legal judgments, and financial inclusion. Yet, widespread adoption remains constrained not only by technical reliability but by the human capacity to *trust* these systems. Trust, in this context, is neither purely cognitive nor purely emotional, it arises at the intersection of comprehension, moral reasoning, and perceived fairness (Frank et al., 2024; Witkowski, Dougherty & Neely, 2024).

While engineers often equate trust with system performance, social cognition research frames it as an evolving psychological contract between humans and technology. Trust calibration depends on whether users perceive AI as *transparent*, *accountable*, and *aligned* with human ethical standards. When algorithms are opaque or autonomous, people may experience moral discomfort or skepticism, particularly in domains where empathy, responsibility, and care are integral (Naiseh et al., 2023).

From a cognitive standpoint, human reasoning about AI can be conceptualized through Kahneman’s (2011) dual-process framework. “System 1” involves fast, intuitive judgments, often guided by emotion and moral instinct, while “System 2” reflects slow, deliberative reasoning that engages when explanations and evidence are available. AI transparency and interpretability act as cognitive cues that trigger System 2 evaluation, allowing users to reason about model logic rather than rely solely on intuition. Conversely, when outcomes feel ethically loaded or unpredictable, intuitive System 1 dominates, leading to caution, moral rejection, or overtrust depending on emotional framing.

Algorithmic trust is therefore a **cognitive-affective construct**: it blends rational evaluation of performance with emotional responses to risk, fairness, and human control. As AI increasingly mediates professional authority, whether in diagnostic reasoning, legal interpretation, or financial risk assessment, understanding how people reason about accountability and comfort becomes essential for responsible design. High-profile failures of automated systems in healthcare triage, sentencing recommendations, and loan approval have intensified calls for “explainable AI” and “human-in-the-loop” safeguards, underscoring that transparency and accountability are now ethical as well as technical imperatives.

In high-stakes domains, barriers to AI adoption are particularly pronounced. Naiseh et al. (2023) highlight that different explanation classes in AI systems can significantly impact trust calibration, with some explanations leading to over-reliance on AI recommendations [3]. Similarly, research shows that judges who rely on AI tools are generally rated lower in legitimacy compared to those who base decisions on their own expertise [35]. This phenomenon suggests that perceived fairness and human legitimacy remain crucial components of algorithmic trust, especially when automation challenges professional identity.

Beyond the technical hurdles, one of the biggest hurdles for AI adoption remains the human psychological barrier, particularly in high-stakes domains. While the public readily embraces AI for low-stakes tasks, such as summarizing information or drafting non-critical communications, this acceptance sharply reverses when decisions directly impact health, liberty, or finance. In these critical areas, a strong preference for advice from human experts over AI persists, suggesting AI is often not viewed as a primary advisor. This reluctance stems not from AI’s quantitative or analytical capabilities, but from the perception that it lacks crucial human elements. Specifically, AI is widely perceived as being emotionless and impartial, which undermines trust when decisions require any degree of subjectivity or moral reasoning ([38]). This reliance on affective consideration shows a significant gap: the technical focus on model explainability often fails to address the more tangible psychological barriers that prevent users from feeling confident in AI’s ultimate judgment. For instance, in healthcare, patients express deep concern over AI’s potential to create communication barriers

and reduce vital physician-patient personal interaction ([37]). Furthermore, AI’s reliance on massive datasets fuels privacy concerns about data collection and re-identification, and the lack of transparent regulatory standards generates anxiety over accountability and liability should a medical error occur ([37]). This shows that the public’s resistance is centered on the loss of human and social concepts, not simply a doubt about AI’s performance capabilities.

Recent empirical work in healthcare illustrates that richer or more numerous AI explanations do not always enhance trust or improve performance. In an experiment involving clinicians evaluating AI-assisted breast cancer diagnoses, higher levels of explanation sometimes failed to increase diagnostic accuracy or self-reported confidence, suggesting that explanations can create a fragile or illusory sense of understanding rather than robust comprehension [5]. This “illusion of explainability” occurs when users feel they understand an AI’s reasoning without fully grasping its mechanisms or limits. The present study builds upon this concept to examine how perceived clarity diverges from actual understanding and how such misalignments shape metacognition, heuristics, cognitive biases, and emotional reactions to AI in practice.

The legal profession, globally recognized as a major sector, has historically been cautious, adopting a traditional approach that leaves many operations under-digitized. While this sector is often slow to embrace technological change, the emergence of AI presents a significant opportunity to fundamentally transform the delivery of legal services, offering the potential to advance legal practice through enhanced efficiency, reduced error, and greater access to justice ([39]). Specifically, AI assists in critical tasks like contract drafting, legal research, and case prediction ([39]), enabling lawyers to work more strategically. However, this potential is met with significant concern: AI’s rise is viewed as a threat that could undermine client confidence, displace professionals in routine tasks, and disrupt the regulatory landscape ([39]). Addressing this tension, between technological opportunity and the structural challenges it poses, is crucial for successful integration.

The rapid advancement of AI presents a paradigm shift within the healthcare industry, promising improvements in diagnostic accuracy, workload reduction, and therapeutic efficiency through technologies like advanced imaging and robotic surgery. While the technical capabilities of AI are expanding and advancing more and more each day, the successful integration of these systems relies on the acceptance and preparedness of healthcare professionals. A study investigating medical doctors’ perceptions of AI in healthcare found a gap between opportunity and readiness: the practitioners reported a generally low degree of familiarity with core AI concepts, despite recognizing AI’s potential and expressing a strong interest in future education ([40]). The research also highlighted significant non-technical barriers to adoption, including the fear of losing essential clinical skills and the widespread concern over the lack of human supervision when AI is implemented in patient care ([40]). These findings depict that for AI to be effectively utilized to improve patient outcomes, medical institutions must first address these profound issues of professional understanding, ethical comfort, and the necessary balance between technological assistance and human judgment.

The deployment of AI in healthcare is met with distinct reservations from the patient population, particularly when comparing AI-driven care to traditional in-person encounters. Esmailzadeh et al.’s large-scale study on patient perceptions (N=634) revealed that while AI showed similar performance risk perception to physician care, seven of the nine measured outcome variables showed significant differences in patient acceptance. Specifically, patients with chronic illnesses expressed significantly lower trust in AI applications and, regardless of illness type, demonstrated significantly higher concerns about communication barriers and the transparency of regulatory standards. This data further solidifies the concept that the primary barriers to AI adoption are not performance-based, but center on ethical and social concerns like trust, human interaction, and accountability ([41]).

The adoption of AI-driven Robo-advisory services (RAs) is disrupting the wealth management industry by offering superior pricing and returns, causing a "threatening alarm" to traditional fund managers [42]. Despite this technical superiority, widespread investor acceptance is limited, with adoption rates as low as 3% in the USA [42]. This constraint is mostly non-technical, coming from a perceived deficit in AI's emotional capacity [42]. Investors require human consultation, especially during volatile or "bearish" market phases, to address their emotional and behavioral concerns [42]. They fear that the lack of human interaction will lead to insecurity and bias-driven decisions [42]. This evidence establishes that in high-stakes finance, trust is fundamentally, like the other high stakes fields, a cognitive-affective construct, where psychological comfort and the need for human accountability seems to be greater than AI's analytical efficiency. This study, therefore, explores how users negotiate the balance between AI's utility and the imperative for human judgment.

Additionally, anthropomorphizing AI, or attributing human-like traits to machines, can mitigate psychological barriers and enhance trust in high-stakes decision-making contexts [6]. While AI offers immense promise in domains such as healthcare, law, and finance, public perceptions remain cautious. People are generally open to AI assisting humans but express discomfort when AI replaces human judgment in critical decisions [7]. This emotional boundary reflects a broader moral intuition about the appropriate scope of automation in human-centered professions.

Despite growing interest, limited empirical research exists on how professionals and the public interact with AI in high-stakes contexts, particularly within Caribbean and small-island developing state contexts where technological infrastructures and cultural perceptions differ from larger economies. This study addresses that gap by combining quantitative survey data with qualitative expert interviews to explore how individuals interpret AI recommendations, assign responsibility, and navigate the balance between human judgment and machine assistance. Understanding these perceptions is particularly crucial in domains such as healthcare, law, and finance, where AI errors may carry severe consequences and erode institutional trust.

This mixed-methods study expands empirical understanding of algorithmic trust through quantitative modeling and qualitative discourse analysis. Two public surveys ($N = 69$ and $N = 44$) were analyzed alongside fifteen in-depth interviews with professionals across medicine, law, psychology, and technology. The surveys examined how comfort, transparency expectations, and accountability attributions shape willingness to trust AI in high-stakes domains. Logistic regression models tested predictive relationships between emotional comfort, perceived transparency, and trust behavior, while interviews explored how professionals negotiate human judgment, empathy, and explainability in their reasoning about AI-assisted work.

By integrating behavioral data with thematic narratives, this research constructs a comprehensive cognitive and affective profile of algorithmic trust. It situates trust not as a static trait but as a dynamic reasoning process ; a negotiation between human intuition and machine rationality that unfolds within ethical, emotional, and social contexts.

2 Methods

2.1 Ethical Considerations and Participant Identification

Participants were fully informed of the study's purpose, procedures, and their rights, including the right to withdraw at any time. Confidentiality and privacy were emphasized throughout the research process. All participant contributions are cited using professional roles or sectors (e.g., "Consultant Physician in Internal Medicine," "Technology Specialist," "Biomedical Engineer") to preserve anonymity while maintaining

professional credibility.

At the time of data collection, no local institutional review board was available in the island of St. Kitts and Nevis; therefore, ethical procedures were self-administered following the *Declaration of Helsinki*. All procedures adhered to widely recognized ethical research principles, including voluntary participation, informed consent, and protection of participants’ confidentiality.

Data were collected between **August 2025** and **September 2025**. Survey responses were obtained via *Google Forms*, and qualitative interviews were conducted through scheduled calls and online meetings, depending on participant preference.

2.2 Quantitative Data and Sampling

The quantitative dataset combines two separate convenience-based survey samples ($N_1 = 69$ and $N_2 = 44$), which limits generalizability (see Section 4, Limitations). Participants for both subsets were recruited from St. Kitts and Nevis and the St. Kitts and Nevis diaspora residing in the United States. Recruitment utilized a convenience sampling strategy, with outreach conducted through personal contacts, professional networks, email, and social-media messaging.

Survey data were cleaned and processed in *Python* (v3.11) using *pandas* (v2.2), *numpy* (v1.26), and *matplotlib* (v3.8), and then combined for descriptive analysis. Items measured comfort with AI across domains (administrative, diagnostic, legal, financial), accountability attribution, and demographic characteristics. Logistic regression was conducted to examine predictors of willingness to trust AI for decision advice. Two participant subsets ($N = 44$ and $N = 23$) were analyzed separately, with predictors including prior AI use, age group, comfort with AI, and perceived transparency. Odds ratios and 95% confidence intervals were estimated from logit standard errors using *statsmodels* (v0.14). Default parameters were used unless otherwise stated.

All Python scripts used for data cleaning and statistical analysis will be made available in an open-access *GitHub* repository upon official publication. Raw data are anonymized and can be accessed upon reasonable request to protect participant confidentiality.

2.3 Qualitative Data and Expert Recruitment

Qualitative data comprised fifteen in-depth, semi-structured interviews with domain experts representing healthcare, legal, psychology, and technology sectors. Thirteen professional participants were recruited from **St. Kitts and Nevis**, one from **Jamaica**, and one from **Trinidad**, ensuring a broader Caribbean professional perspective. Recruitment utilized a targeted convenience sampling approach via email and phone calls to personal and professional networks to ensure representation from high-stakes professional sectors within the region.

All participants provided informed consent prior to participation. Interviews were transcribed and coded thematically using a semantic approach. Coding reliability was strengthened through iterative review and cross-checking of themes. Key themes identified were: (1) emotional reasoning, (2) human accountability and judgment, (3) algorithmic bias, and (4) explainability and transparency. The semantic word cloud below visualizes linguistic salience within the corpus. Representative quotations are included to illustrate emerging patterns. The study is ongoing, and additional interviews may refine or expand these insights

2.4 Reproducibility and Data Availability

This study follows open research principles to support transparency and reproducibility. The complete analysis pipeline, including Python scripts and data-processing workflows, will be uploaded to an open-access *GitHub* repository upon official publication. Due to the small population size and potential identifiability of participants, de-identified raw data will be made available upon reasonable request to the corresponding author. Supplementary materials (e.g., survey instruments and interview guide) will also be provided to facilitate replication and secondary analysis.

3 Results

3.1 Overview of Findings

Results from surveys and interviews provide a multi-faceted view of AI trust in professional contexts. Survey responses indicate that perceived transparency and prior AI experience are associated with higher trust in AI-assisted decision-making, while accountability is consistently attributed to humans rather than AI systems. Expert interviews reinforce the importance of human judgment, ethical responsibility, and oversight, highlighting potential risks if AI is treated as a substitute for professional expertise.

3.2 Quantitative Data

Quantitative data were collected from two separate survey samples ($N_1 = 69$ and $N_2 = 44$) recruited from St. Kitts and Nevis and the St. Kitts and Nevis diaspora residing in the United States. Two datasets were collected to increase coverage across different participant pools and professional networks, allowing for a broader preliminary understanding of AI trust in diverse demographic and professional contexts. Participants reported their comfort with AI, perceived transparency, prior AI experience, and willingness to trust AI-generated advice across healthcare, legal, and financial domains. Data from both samples were combined for descriptive analyses, while smaller subsets ($N = 44$ and $N = 23$) were analyzed separately using logistic regression to examine predictors of AI trust and demand for transparency. Figures 1–6 and Tables 1–2 illustrate key patterns, including preferences for human oversight, accountability, and the influence of prior AI exposure on AI trust.

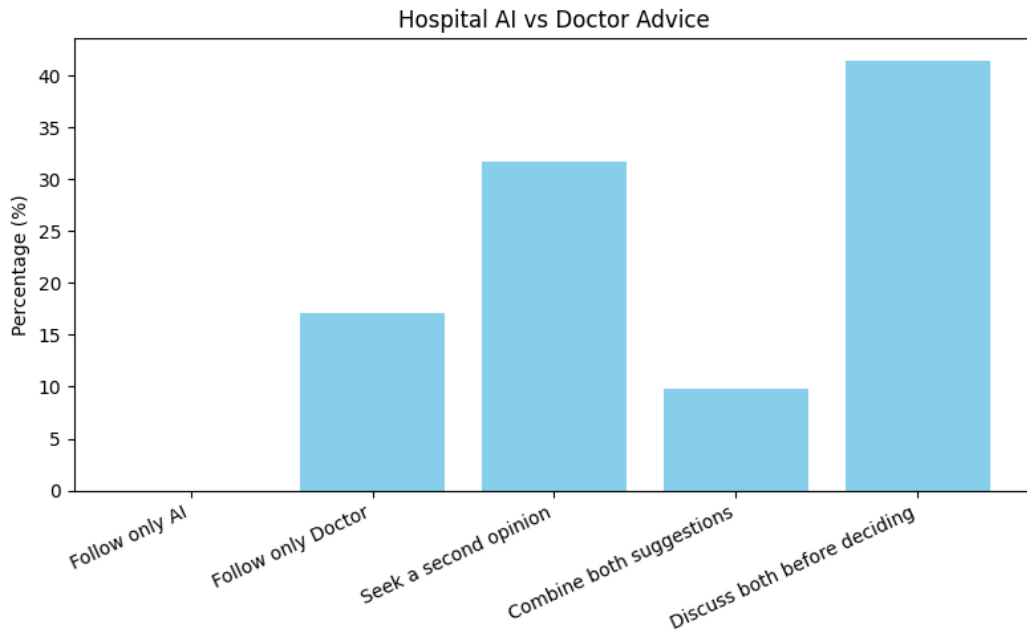


Figure 1: Respondent reactions to a hospital AI recommendation versus a doctor’s advice. A majority sought human confirmation before trusting AI-driven treatment suggestions. A small percentage would follow only the doctor’s advice.

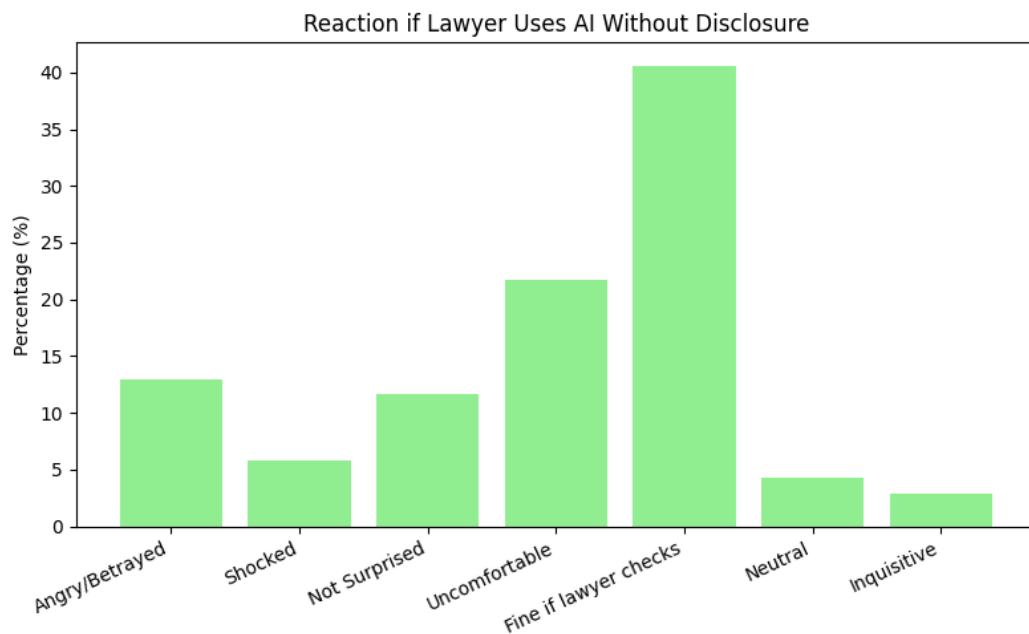


Figure 2: Perceived comfort with a lawyer using AI for case preparation or research tasks. Responses reflect higher acceptance for support roles, not autonomous legal judgments.

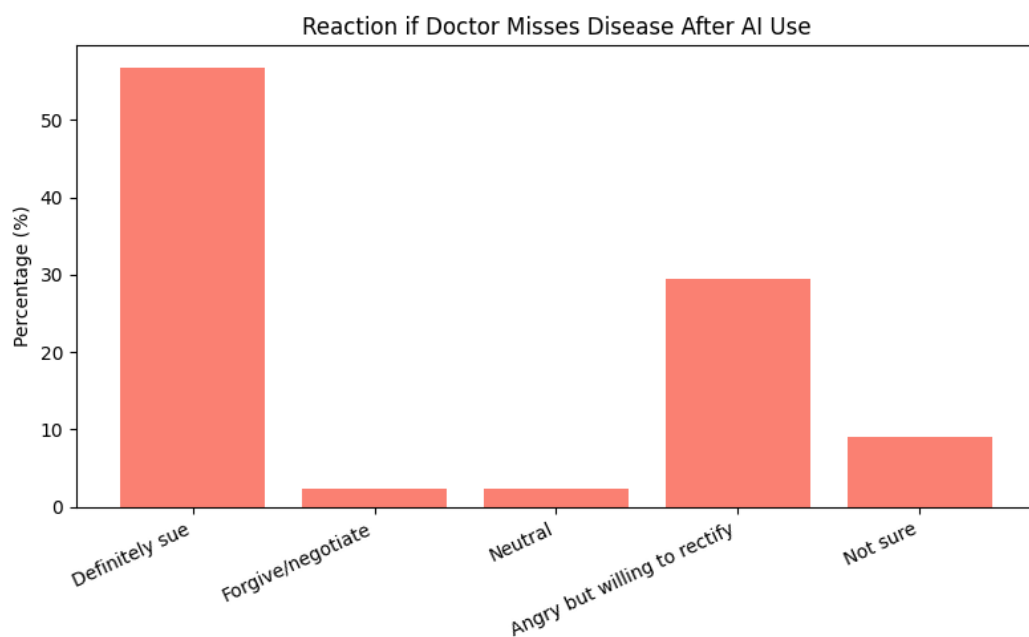


Figure 3: Hypothetical response if a doctor missed a diagnosis after using an AI tool. Most respondents held the doctor accountable, underscoring persistent human responsibility.

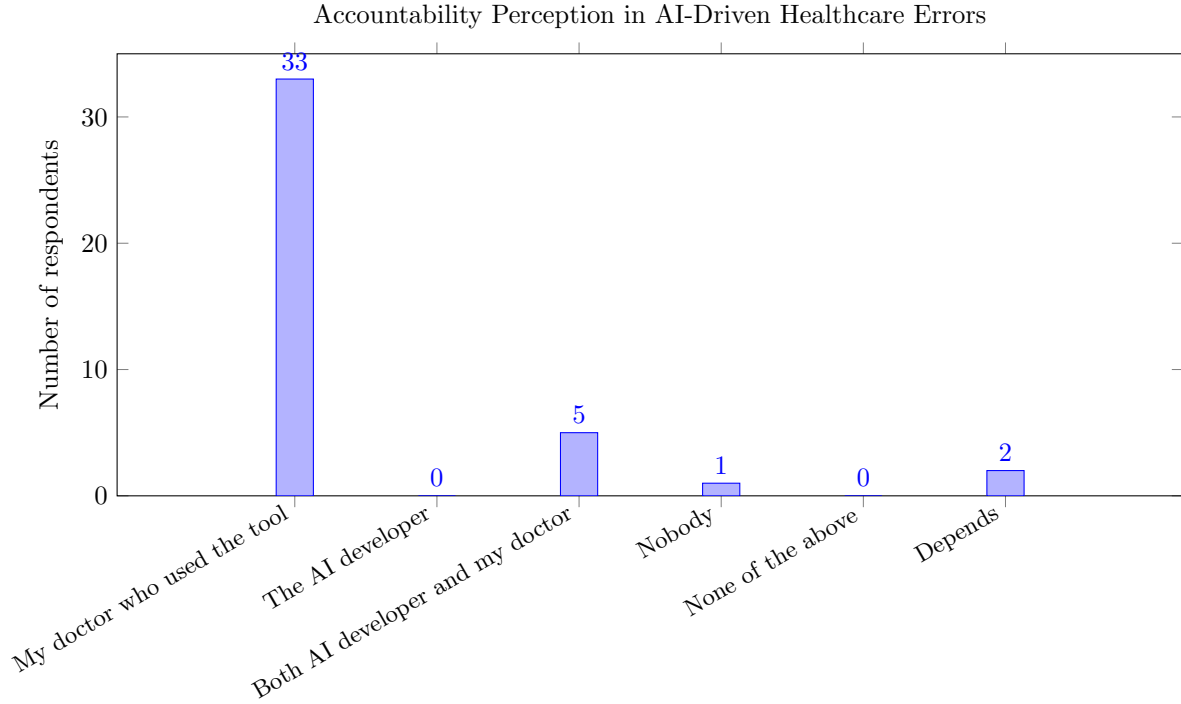


Figure 4: Survey responses to the question: “If your doctor used an AI healthcare tool that made a mistake and affected you, who would you hold responsible?” (N = 41). Most respondents held the doctor primarily accountable, indicating strong expectations for human oversight in AI-assisted medicine.

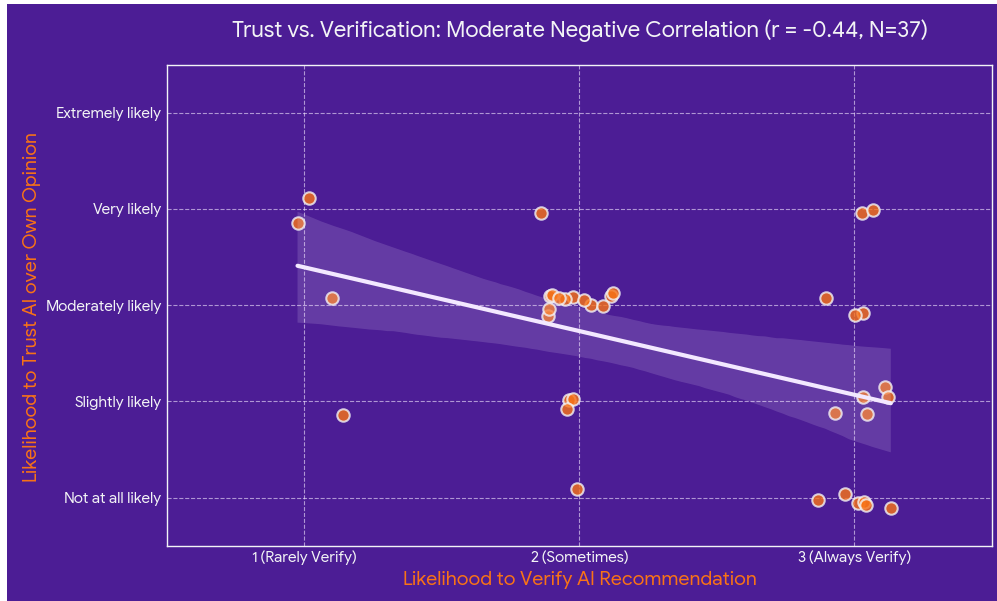


Figure 5: Trust vs. Verification: Moderate Negative Correlation ($r = -0.44, N = 37$). This scatter plot reveals a moderate negative correlation between a participant’s likelihood to verify an AI’s output and their likelihood to trust that AI over their own opinion. As the necessity of verification increases, epistemic trust decreases, demonstrating that the verification requirement acts as a cognitive tax on the system’s perceived authority.

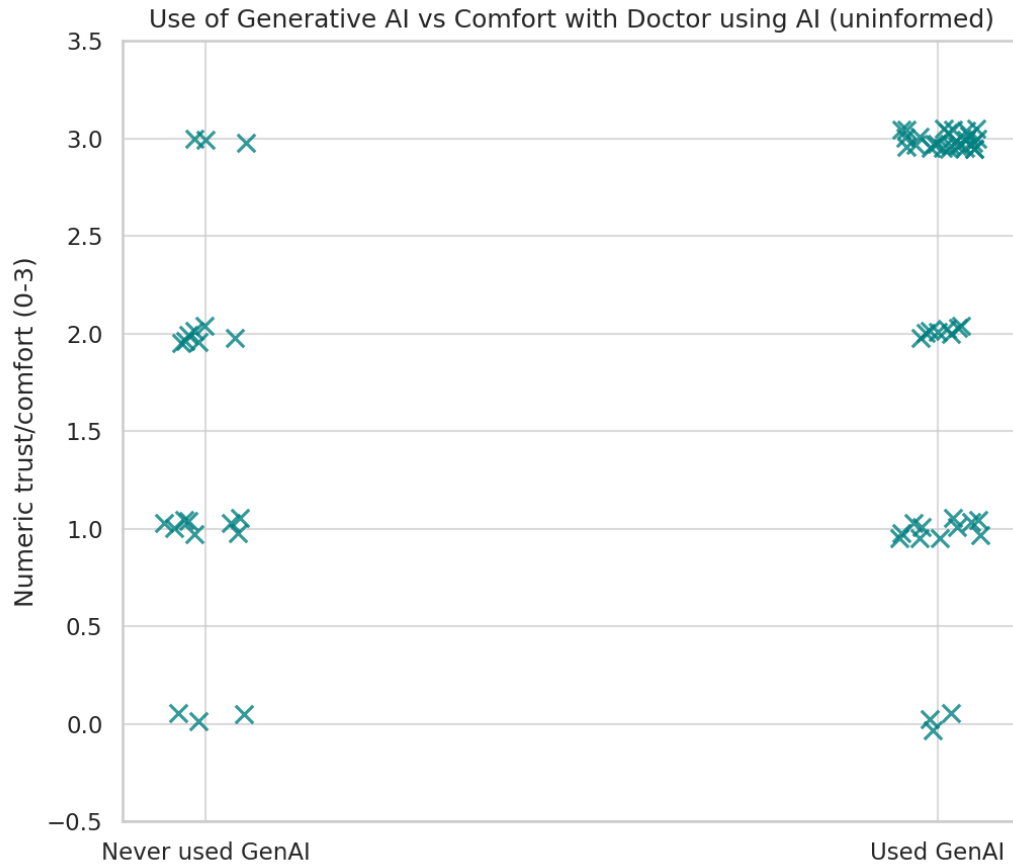


Figure 6: Relationship between prior generative AI use and comfort with a doctor using AI. Prior users of AI report higher comfort levels, supporting familiarity-based trust.

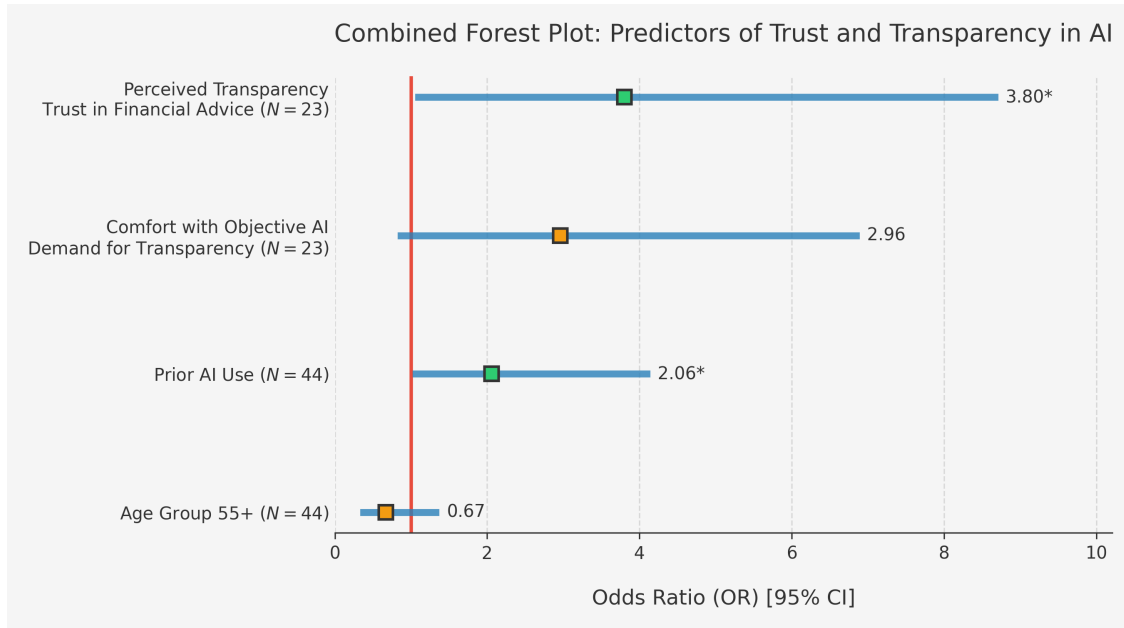


Figure 7: Forest Plot: Combined Predictors of AI Attitudes. This plot summarizes the binary logistic regression models (Model 1: $N = 44$; Model 2: $N = 23$), showing Odds Ratios (OR) and 95% Confidence Intervals (CI) for key predictors of AI trust and transparency demands. The vertical dashed line shows the null effect ($OR = 1.0$). Statistically significant positive effects were found for Perceived Transparency ($OR = 3.80^*$) and Prior AI Use ($OR = 2.06^*$), depicting that both rational clarity and familiarity are important drivers of trust in high stakes fields. Comfort with Objective AI ($OR = 2.96$) and Age Group 55+ ($OR = 0.67$) were not statistically significant predictors.

3.3 Logistic Regression: AI Attitude Predictors

Table 1: Model Overview: Predictors of AI Attitudes (N = 44)

Predictive Factor (X)	Outcome (Y)	Odds Ratio (OR)	95% CI	p-value
Comfort with Objective AI	Demand for Transparency (loan decision scenario)	2.96	[0.82, 6.89]	0.102
Perceived Transparency	Trust in AI for Financial Advice	3.80	[1.05, 8.71]	0.038*

*Significant at $p < .05$. Confidence intervals estimated from logit standard errors.

Interpretation. The first model suggests that individuals who feel emotionally comfortable with AI are almost three times more likely to demand transparent explanations when automation fails, though this effect only approached significance ($p = .102$). The second model achieved significance ($p = .038$), indicating that perceived transparency substantially increases trust in AI-driven financial advice. Together, these results highlight the cognitive-affective consistency underlying algorithmic trust: users who value rational clarity also exhibit stronger comfort with machine-supported reasoning.

3.4 Logistic Regression: Inferential Analysis:

Table 2: Binary Logistic Regression Predicting Trust in AI (N = 23)

Predictor	Coefficient	Odds Ratio (OR)	95% CI	p-value
Intercept	-0.5526	0.5755	[0.18, 1.78]	0.273
Prior AI Use	0.7208	2.0561	[1.02, 4.14]	0.043*
Age Group 55+	-0.3963	0.6728	[0.33, 1.37]	0.225

*Significant at $p < .05$ (two-tailed). Confidence intervals estimated from logit standard errors.

Interpretation. Individuals with prior AI experience were approximately twice as likely to trust AI advice ($OR = 2.06$, $p = .043$), while participants aged 55 and older were about 33% less likely ($OR = 0.67$, $p = .225$). Although the age effect was not statistically significant, the directionality suggests greater caution among older participants. The model supports the hypothesis that familiarity fosters calibrated trust, aligning with broader findings on exposure-driven confidence in AI systems.

3.5 Qualitative data

Fifteen expert interviews (healthcare, legal, psychology, tech) were transcribed and coded thematically (semantic approach). Themes identified in the qualitative analysis include emotional reasoning, human accountability, algorithmic bias, the illusion of explainability, AI-augmented creativity and innovation, metacognition, and cognitive and decision-making biases (including heuristics, status quo bias, and automation bias). The semantic word cloud below visualizes linguistic salience in the corpus.

Theme 5: The Therapeutic Threshold - Limits of Machine Empathy

“If data is biased, then it cannot be fair, can it? . . . There is no practical reason why AI cannot give samples of a range of views and explanations. When you train as a psychologist, you learn to draw from multiple theories. AI could do the same if designed with integrity.” – **Clinical Psychologist**

Theme 6: Explainability and transparency

“Simpler is better. Before committing to any platform, have AI show similar cases and mistakes.” – **Technology Expert**

Theme 7: AI-Augmented Creativity and Innovation

“I think AI will augment natural creativity and expand how tissue-engineered projects are used clinically. We will rely heavily on experimentalists to drive the field forward.” – **Biomedical Engineer**

Integration of Quantitative and Qualitative Data Taken together, survey and interview data suggest that trust in AI is shaped both by individual experience and professional norms. While transparency and prior exposure to AI increase user confidence, experts consistently emphasize that accountability and ethical judgment cannot be outsourced to machines. These preliminary findings highlight the cognitive-affective and professional dimensions of AI trust across domains. Quotes included in the qualitative analysis are illustrative and reflect preliminary results; additional interviews may refine or expand these themes.

4 Discussion

4.1 Cognitive, Professional, and Ethical Dimensions of Trust in AI

Trust in AI is multidimensional, encompassing cognitive understanding, affective comfort, ethical evaluation, and alignment with professional norms. Users rely on instrumental trust, based on competence and performance, and moral trust, based on fairness, transparency, and ethical alignment. In healthcare, trust also depends on perceived support of clinical expertise, patient safety, and workflow integration. In law, procedural fairness, auditability, and explainability dominate, while in finance, predictive reliability and ethical handling of information are prioritized. Across sectors, adoption is influenced not only by technical performance but also by professional legitimacy, perceived competence, and the assurance that AI will augment rather than replace human judgment.

Our findings reinforce these dimensions: survey respondents largely sought human confirmation before trusting AI-driven treatment suggestions, emphasizing that cognitive trust alone is insufficient without human oversight (Figure 1). Qualitative interviews highlighted emotional reasoning as a key factor in AI acceptance, demonstrating that affective comfort and instinctive oversight strongly influence adoption even among technically literate professionals. Across domains, human accountability remains central, with most participants holding doctors responsible for errors despite AI assistance (Figures 3–4; Themes 1 & 2).

This necessity of verification comes at a psychological cost. We found a moderate negative correlation ($r = -0.44$) between how likely a user was to verify the AI’s output and their willingness to trust the AI over their own opinion. This suggests that requiring verification acts as a cognitive tax on trust. It emphasizes

an important proposition: while shallow transparency encourages over-reliance, the deeper demand for traceability and human effort actually diminishes the AI’s perceived authority.

4.2 Healthcare: Knowledge, Attitudes, Acceptance, and Application

Healthcare represents a highly complex domain for AI integration. Namdar Areshtanab et al. (2025) found that 41.1% of nurses had low knowledge of AI, yet 65.8% reported positive attitudes toward its use, 74.6% demonstrated moderate acceptance, and 55.8% applied AI at a high level. Correlations between knowledge, attitude, acceptance, and application ($r = 0.31\text{--}0.74$, $p < 0.001$) suggest that increasing AI literacy strengthens engagement, acceptance, and practical application.

Survey data corroborate these findings: prior AI experience predicted higher comfort with clinician AI use ($OR = 2.06$, $p = 0.043$), highlighting that familiarity breeds calibrated trust (Figure 5; Logistic Regression, Table 2). Older participants exhibited slightly lower trust ($OR = 0.67$, $p = 0.225$), suggesting demographic differences in adoption risk. AI technologies, including predictive analytics, robotic tools, and virtual assistants, improve operational efficiency, allowing clinicians to focus on patient care. Despite these benefits, ethical challenges persist, including concerns over algorithmic bias, data privacy, and the potential for over-reliance, which can compromise clinical oversight.

The Necessary Link Between Exposure and Trust

Building on these findings, it is clear that familiarity is a critical driver of integration, not merely a secondary factor. The survey data strongly suggests that increased clinician experience with AI correlates positively with enhanced comfort and trust. Crucially, when participants observe the direct demonstrated benefit of AI in their daily workflow, for example, predictive analytics instantly flagging a high-risk patient or a virtual assistant handling routine paperwork, the perceived utility shifts AI from an abstract threat to a tangible, beneficial tool. This mechanism facilitates greater openness and accelerated adoption.

However, this progress is threatened by a growing and dangerous trend: radical skepticism. When a generative AI model, for instance, produces a single, high-profile error (a "hallucination"), the immediate and widespread reaction is often the rejection of the entire AI concept. This shows a fundamental cognitive error and a failure to distinguish between a specific tool’s flaw and the potential of the broader field. This overreaction risks stalling the adoption of countless proven, beneficial applications. Therefore, empirical research is required to develop and validate intervention strategies that prevent this error generalization from becoming the dominant driver of AI skepticism and subsequent rejection in clinical settings.

4.3 Generative AI in Medical Decision-Making and Peer Perception

Yang et al. (2025) examined how the use of generative AI (GenAI) in clinical decision-making affects peer evaluations. Physicians who used GenAI as a primary decision-making tool were rated significantly lower in clinical skill (mean = 3.79) than controls (mean = 5.93, $p < 0.001$). Framing GenAI as a verification tool partially mitigated this effect (mean = 4.99, $p < 0.001$). While GenAI improved clinical assessment accuracy (mean = 4.30, $p < 0.002$), its visible use carries reputational risks, aligning with attributional discounting theory.

Our findings complement this: participants emphasized non-delegable human judgment (Theme 1) and therapeutic limits of AI empathy (Theme 4). Emotional reasoning and professional intuition strongly in-

fluenced comfort with AI, suggesting that adoption depends not only on performance metrics but also on alignment with professional norms and peer perceptions.

The Comfort of AI and the Risk of Bias

The observation that doctors prefer using generative AI for verification rather than as a primary tool suggests its adoption begins as a form of comfort and security. Even if practitioners state they won't rely completely on the tool, this setup creates a high risk of automation bias. This happens especially when the AI provides data based on analysis that would take humans months to complete. When a machine offers information with such immense computational authority, it becomes very difficult, psychologically, for a person to challenge or ignore it even if their professional intuition suggests they should. This tension between trusting one's own expert judgment and believing the AI must be correct is a fundamental human challenge in high-stakes fields. Therefore, we need future research to identify the point where the computational power of AI starts to override human judgment and to determine how to prevent this over-reliance in critical, clinical moments.

4.4 Primary Care and Patient Portal Messaging

Biro et al. (2025) studied the use of generative AI in drafting patient portal messages. AI drafts significantly improved efficiency, reducing cognitive workload in 80% of cases and being perceived as safe by 75% of participants. However, this efficiency came with a critical caveat regarding safety: uncorrected errors were common. Specifically, 65% of participants missed at least one error, and 35% of erroneous drafts were submitted without any edits. Only one participant successfully corrected all four errors introduced into the sample drafts. This outcome demonstrates that a user's positive attitude toward AI does not guarantee safe outcomes.

The persistence of errors despite human review highlights a crucial failure in metacognition, the user's ability to assess their own comprehension and performance, and is directly relevant to our study's findings on algorithmic trust. This phenomenon is a textbook case of automation complacency, where users over-rely on a system they perceive as competent, leading to reduced vigilance. Furthermore, the 35% rate of submitting erroneous drafts suggests that the AI may create an illusion of explainability, leading users to assume the output is correct without fully engaging in critical analysis.

Cognitive biases such as functional fixedness, which is described as being unable to see alternative solutions; confirmation bias, which is seeking evidence that confirms the AI is right; and automation complacency explain why errors persist despite positive attitudes, depicting that the expert theme that human judgment and oversight are non-delegable. Our qualitative and survey data further illuminate this risk, as users demanded transparency when automation failed, and interviews underscored the need for explainability and error visibility (Theme 5) to interrupt this complacency loop.

4.5 Integration of Knowledge, Attitude, and Peer Perception

Integrating Namdar Areshtanab et al. (2025), Biro et al. (2025), Yang et al. (2025), and our findings reveals a convergent pattern: The integration of AI into professional practice is driven by a complex interplay of experience and perception. Research consistently shows that familiarity breeds calibrated trust: prior exposure to AI tools is a strong predictor of both comfort and subsequent acceptance among users. Beyond mere exposure, emotional reasoning shapes engagement, with instinctive oversight and affective comfort playing a significant role in guiding how and when AI is incorporated into high-stakes decision-making. Critically,

these factors operate within the framework of the existing professional structure, as professional identity anchors accountability. This means that regardless of AI performance, ultimate responsibility and judgment remain human-centered, and peer perception significantly influences the rate and method of technological adoption.

Structured AI education, professional development, and integration into curricula are essential. Visible adoption of GenAI should be framed as verification to maintain credibility while leveraging AI benefits.

4.6 Law: Legitimacy, Accountability, and Trust Calibration

In legal practice, AI adoption focuses on procedural fairness, explainability, and accountability. Judges and lawyers interact with AI in tasks such as predictive policing, case research, document review, and risk assessment. Trust is highly context-dependent, and our findings indicate that the public distinguishes between AI used as a support tool versus AI making autonomous legal judgments. Survey responses show higher acceptance for AI performing research or preparatory functions, whereas AI making independent legal decisions is met with skepticism (Figure 2).

Fine et al. (2025) show that judges using AI are generally rated lower than those relying solely on expertise, although confidence in AI by the judge mitigates skepticism. Racial differences in perception suggest that AI may be seen as a bias-mitigating tool in some contexts, with Black participants rating AI-assisted judges more favorably than White or Hispanic participants.

Qualitative interviews reinforce these patterns. Legal practitioners emphasized non-delegable human judgment (Theme 1), cautioning against treating AI as a substitute rather than an assistant. Emotional reasoning was prominent, with intuition and professional norms remaining critical for interpreting AI outputs. Experts repeatedly stressed the importance of transparency, explainability, and auditability, noting that AI systems must clearly show the reasoning behind suggestions and provide examples of prior cases to maintain trust (Theme 5). Survey data also indicated that emotional comfort and familiarity with AI increased acceptance for support roles, highlighting the cognitive and affective mechanisms of trust.

These findings converge to indicate that AI legitimacy in law is contingent on human endorsement, visibility of oversight, and alignment with ethical norms. AI adoption can enhance efficiency and procedural justice but requires careful framing to maintain public trust and professional accountability.

4.7 Finance: Predictive Reliability, Efficiency, and Ethical Risk

The Illusion of Objective Data and the Role of Moral Reasoning While AI adoption in finance heavily emphasizes technical tasks like predictive analytics, fraud detection, compliance monitoring, and trading optimization, our findings reveal that high technical trust does not eliminate human factors. There is a common perception that finance primarily deals with objective numbers and transactions, making AI seem like a natural, purely logical fit, a view often contrasted with the deeply nuanced, subjective moral reasoning required in fields like medicine or law. However, respondents' trust is clearly shaped by persistent moral and affective considerations, not just performance. Transparency, perceived fairness, and the assurance of human oversight are essential, with clarity strongly increasing trust in AI-driven advice (OR = 3.80). Financial professionals' own experiences and professional judgment are essential for properly contextualizing client problems and determining ethical action. The data shows that even when handling money, the need for rational clarity is tightly linked to emotional comfort, confirming that the professional's subjective view remains vital for calibrated trust and responsible AI integration.

Qualitative interviews highlight the risks of ethical deskilling, over-reliance, and diminished professional judgment. Experts stressed the need for transparent, auditable models to allow human professionals to understand and challenge AI recommendations. Emotional reasoning plays a role, as users' comfort with AI is linked to perceived empathy and fairness of algorithmic outputs.

These findings suggest that financial AI should augment professional decision-making, delivering efficiency and predictive reliability while robust governance mechanisms maintain accountability and ethical standards. Transparency, explainability, prior experience, and demographic considerations are all critical in trust calibration.

4.8 Cross-Domain Synthesis: Toward Context-Specific Governance

Across healthcare, law, and finance, AI trust and adoption are context-dependent. Positive perceptions increase adoption, but cognitive biases, knowledge gaps, workflow pressures, reputational concerns, demographic differences, and emotional reasoning create risks. AI should augment human expertise, not replace it. Structured training, context-aware system design, transparent decision support, and policies supporting human oversight are essential. Framing and integration strategies should account for professional norms, peer evaluations, workflow realities, and public expectations to maximize efficiency, safety, legitimacy, and ethical practice.

5 Limitations

This study has several limitations. First, sample sizes were modest ($N = 69$, 44 , and regression $N = 23$), which limits the statistical power of quantitative analyses and the generalizability of findings. Participants were restricted to the Caribbean and the Caribbean diaspora, making results context-specific and less applicable to other regions or populations. All data coding and analysis were conducted by a single independent researcher, which may introduce bias despite systematic procedures and reflective practices.

The qualitative interviews were coded by only one researcher, and while coding was systematic, the absence of multiple coders or external validation may reduce inter-rater reliability. Survey measures relied on self-report, which may be subject to social desirability or recall bias. Additionally, the study focused on selected domains of AI use (administrative, diagnostic, legal, and financial), and other high-stakes domains were not explored. Finally, the cross-sectional design of the survey prevents causal inference, and findings should be interpreted as indicative rather than definitive.

6 Conclusion

The findings from this mixed-methods study indicate that algorithmic trust is not a purely rational process but a dynamic negotiation between cognitive reasoning and affective response. Trust in artificial intelligence (AI) emerges at the intersection of two systems of human thought ; System 1 (intuitive and affective) and System 2 (deliberative and metacognitive). In high-stakes professional domains such as healthcare, law, and finance, these dual processes collectively determine whether AI is accepted as a supportive collaborator or resisted as a threat to human agency.

Our results demonstrate that adoption decisions are constrained by enduring cognitive biases and heuristics, including status quo bias (a preference for human judgment) and automation bias (over-reliance on machine output). Professionals actively employ metacognition to balance these tendencies, positioning AI

as a subordinate decision-support tool that preserves human accountability and moral responsibility. This reflects a broader epistemic stance: AI can inform human reasoning, but the legitimacy of judgment remains human-centered.

Sustainable AI integration therefore requires moving beyond narrow notions of technical explainability toward achieving cognitive and ethical compatibility. Future AI design and policy efforts should explicitly address these psychological and normative dimensions by:

1. Framing AI transparently to support users' metacognitive capacity to evaluate the quality and limits of algorithmic reasoning.
2. Mitigating affective resistance through alignment with human ethical norms, perceived fairness, and emotional intelligibility.
3. Embedding human oversight within all high-stakes decision systems to ensure accountability, empathy, and trust remain central.

In sum, trust in AI is best understood as a cognitive-affective equilibrium; one that develops through iterative exposure, ethical design, and institutional transparency. The long-term success of AI in human-centered professions will depend not only on improving model performance but also on cultivating a psychologically compatible partnership between humans and intelligent systems.

7 Acknowledgments

The author thanks colleagues from the University of the West Indies for their valuable feedback on earlier drafts of this manuscript and assumes full intellectual responsibility for the content, integrity, and scientific conclusions of this manuscript. **The entire research process, including the conceptualization, data collection, statistical analysis, and interpretation of all results, was conducted solely by the human author.**

Generative Artificial Intelligence (GenAI) and AI-assisted tools (Gemini, ChatGPT-4.0, and Copilot) were utilized **only for non-substantive tasks, specifically: language refinement, manuscript structure and flow enhancement, grammar checks, and proofreading**. AI tools were also used to provide coding suggestions and debugging support for the Python and Jupyter Notebook scripts employed to execute the statistical analyses and generate figures (including the logistic regression, pie charts, and scatter plots).

The use of AI tools did not influence the scientific design, statistical outcomes, or integrity of the data presented. All content was thoroughly reviewed and verified by the author, who retain full accountability for the work. The underlying data analyses (performed using Google Sheets and Microsoft Excel), along with all code and data, will be made publicly available in a GitHub repository upon publication.

References

- [1] Frank, D.-A., Chrysochou, P., Mitkidis, P., Otterbring, T., & Ariely, D. (2024). *Navigating uncertainty: Exploring consumer acceptance of artificial intelligence under self-threats and high-stakes decisions. Technology in Society*, 79, 102732. <https://doi.org/10.1016/j.techsoc.2024.102732>

- [2] Witkowski, K., Dougherty, R. B., & Neely, S. R. (2024). *Public perceptions of artificial intelligence in healthcare: Ethical concerns and opportunities for patient-centered care*. *BMC Medical Ethics*, 25, 74. <https://doi.org/10.1186/s12910-024-01056-7>
- [3] Naiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2023). *How the different explanation classes impact trust calibration: The case of clinical decision support systems*. *International Journal of Human-Computer Studies*, 169, 102941. <https://doi.org/10.1016/j.ijhcs.2022.102941>
- [4] Fine, J. (2025). *Algorithmic justice: Perceptions of legitimacy and fairness in AI-assisted judicial decision-making*. *Journal of Empirical Legal Studies*, 22(1), 45–67. <https://doi.org/10.1111/jels.12345>
- [5] Rezaeian, M., Sharma, S., & Patel, N. (2024). *The impact of AI explanations on clinical decision confidence and diagnostic accuracy: A study on breast cancer imaging*. *Frontiers in Artificial Intelligence*, 7, 1562. <https://doi.org/10.3389/frai.2024.01562>
- [6] Saffarini, A. (2023). *Trusting AI in high-stake decision making: Anthropomorphism, moral judgment, and affective reasoning*. *Proceedings of the 2023 IEEE International Conference on Humanized AI Systems*, 1–8. <https://doi.org/10.1109/HumanAI.2023.00015>
- [7] Ipsos Global. (2021). *How do people feel about artificial intelligence? Global survey report*. Ipsos MORI Public Affairs. <https://www.ipsos.com/en/global-advisor-ai-attitudes-2021>
- [8] Naeem, S., Abbas, H., & Hamna, S. (2025). *Knowledge and perception of practicing doctors on AI in medicine*. *JAIMC Journal*, 23(2).
- [9] Mahadik, S. S. (2024). *Perceptions of lawyers on the impact of AI in legal practice*. *International Journal for Multidisciplinary Research*, 6(3).
- [10] Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- [11] Wang, L. (2025). *Machine translation and moral responsibility: The ethics of automated communication*. *Journal of Applied Ethics in Technology*, 12(1), 77–94.
- [12] Robinson, O. (2020). *Artificial intelligence awareness and acceptance among clinicians in Nigeria*. *African Journal of Medical Informatics*, 6(2), 45–58.
- [13] Basri, W. S., & Almutairi, A. (2023). *Enhancing financial self-efficacy through artificial intelligence in the banking sector*. *International Journal of Cyber Criminology*, 17(2), 284–311. <https://doi.org/10.5281/zenodo.4766717>
- [14] Ursachi, G. (2019). *Applications of natural language processing in cybersecurity: Insurance sector use cases*. *Cybersecurity Review*, 8(1), 45–52.
- [15] Mansour, A. (2020). *Phishing detection through AI-driven NLP models*. *Information Security Journal*, 29(4), 322–336.
- [16] Akyuz, M., & Mavnacio, E. (2021). *Adoption of AI in financial services: Drivers, barriers, and ethical implications*. *Journal of Financial Innovation*, 9(3), 211–229.
- [17] Jan, M. (2021). *Information asymmetry and algorithmic decision-making in financial engineering*. *Finance and Data Ethics*, 5(2), 77–89.

- [18] Vandrangi, R. (2022). *Machine learning applications in insurance analytics: Predicting claims and risk*. *Journal of Risk and Data Science*, 4(1), 55–70.
- [19] Piriälä, M., Rahman, A., & Li, Y. (2022). *Artificial intelligence for fraud detection and cost control in insurance systems*. *Computational Finance Letters*, 12(4), 118–132.
- [20] Maple, C., Szpruch, L., Epiphaniou, G., Staykova, K., Singh, S., Penwarden, W., Wen, Y., Wang, Z., Hariharan, J., & Avramovic, P. (2023). *The AI revolution: Opportunities and challenges for the finance sector*. *The Alan Turing Institute Report*.
- [21] Buckley, R., Milana, F., & Ashta, A. (2021). *Transparency in AI-based decision models*. *AI and Society*, 36(4), 1153–1171.
- [22] Ryll, M., Fabri, C., & Herrmann, P. (2020). *Black-box challenges in financial AI models*. *Computational Finance Review*, 10(2), 44–59.
- [23] Danielsson, J., Macrae, R., & Zigrand, J.-P. (2022). *Systemic risk and AI in finance*. *Financial Stability Studies*, 14(3), 99–121.
- [24] Ashta, A., & Herrmann, P. (2021). *Accountability dilemmas in AI-driven financial decision-making*. *Journal of Business Ethics*, 178(5), 987–1002.
- [25] Digital Regulation Cooperation Forum (DRCF). (2022). *Auditing AI: The role of oversight bodies in algorithmic accountability*. UK Government White Paper.
- [26] Kruse, T., Milana, F., & Ashta, A. (2021). *AI adoption and skill transformation in financial organizations*. *Journal of Economic Policy and Technology*, 15(2), 145–163.
- [27] Shuaib, A. (2024). *Transforming healthcare with AI: Promises, pitfalls, and pathways forward*. *International Journal of General Medicine*, 17, 1765–1771. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11070153/>
- [28] Lee, J., & Shin, D. (2020). *Data governance and privacy in AI finance*. *Information Systems Frontiers*, 22(3), 641–655.
- [29] Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). *How AI will change the future of marketing and finance*. *Journal of the Academy of Marketing Science*, 48(1), 24–42.
- [30] Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähtenmäki, I. (2018). *Using consumer data in financial services: Ethical considerations*. *Journal of Service Management*, 29(5), 821–845.
- [31] Yang, H., Dai, T., Mathioudakis, N., Knight, A. M., Nakayasu, Y., & Wolf, R. M. (2025). *Peer perceptions of clinicians using generative AI in medical decision-making*. *npj Digital Medicine*, 8(530). <https://doi.org/10.1038/s41746-025-01901-x>
- [32] Namdar Areshtanab, H., Rahmani, F., Vahidi, M., Saadati, S. Z., & Pourmahmood, A. (2025). *Nurses’ knowledge, attitudes, acceptance, and application of AI: Cross-sectional study in healthcare*. *Scientific Reports*, 15(1), 11002. <https://doi.org/10.1038/s41598-025-11002-0>
- [33] Biro, J. M., Handley, J. L., McCurry, J. M., Visconti, A., & Weinfeld, J. (2025). *Opportunities and risks of artificial intelligence in patient portal messaging in primary care*. *npj Digital Medicine*, 8(35). <https://doi.org/10.1038/s41746-025-01586-2>

- [34] Citron, D. K., & Pasquale, F. (2014). *The scored society: Due process for automated predictions*. *Washington Law Review*, 89(1), 1–33. <https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2/>
- [35] Fine, A., Berthelot, E. R., & Marsh, S. (2025). *Public perceptions of judges’ use of AI tools in courtroom decision-making: An examination of legitimacy, fairness, trust, and procedural justice*. *Behavioral Sciences*, 15(4), 476. <https://doi.org/10.3390/bs15040476>
- [36] Yigitcanlar, T., Degirmenci, K., & Inkinen, T. (2024). *Drivers behind the public perception of artificial intelligence: Insights from major Australian cities*. *AI & Society*, 39, 833–853. <https://doi.org/10.1007/s00146-022-01691-4>
- [37] Esmaeilzadeh, P., Mirzaei, T., & Dharanikota, S. (2021). *Patients’ Perceptions Toward Human–Artificial Intelligence Interaction in Health Care: Experimental Study*. *J Med Internet Res*, 23(11):e25856. <https://doi.org/10.2196/25856>
- [38] Saffarini, A. (2023). *Trusting AI in High-stake Decision Making*. arXiv preprint arXiv : 2401.13689. <https://doi.org/10.48550/arXiv.2401.13689>
- [39] Mahadik, S. S. (2024). *A Study on Perception of Lawyers about the Impact of Artificial Intelligence on the Legal Profession*. *International Journal for Multidisciplinary Research (IJFMR)*, 6(3), 1–9.
- [40] Banerjee, A., Sarangi, P. K., & Kumar, S. (2024). *Medical Doctors’ Perceptions of Artificial Intelligence (AI) in Healthcare*. *Cureus*, 16(9): e70508. DOI: 10.7759/cureus.70508.
- [41] Esmaeilzadeh, M., Treiblmaier, H., & Al-Jubari, A. M. (2024). *Patients’ Perceptions of AI in Healthcare: An Experimental Study of the Effects of Illness Type and Service Encounter Type on Risk and Benefit Perceptions*. *JMIR Preprints*, 25856.
- [42] Rasiwala, F. S., & Kohli, B. (2021). *Artificial Intelligence in FinTech: Understanding Stakeholders Perception on Innovation, Disruption, and Transformation in Finance*. *International Journal of Business Intelligence Research*, 12(1), 48–57. DOI: 10.4018/IJBIR.20210101.0a3.