

When AI Enters the Decision-Making Room: How Artificial Intelligence Shapes Human Judgment, Behavior, and Institutions

Research by: Parth Parnami, Student at Kunskapsskolan International School

Research supervised by: Rajiv Malik, Legal Leader at LG Electronics India

1. Abstract

The influence of artificial intelligence on human decision-making cannot be understood through technical performance alone. AI enters environments already shaped by human psychology, social expectations, institutional practices, and competing interests. This paper develops an exploratory thematic synthesis of a virtual roundtable and a conference panel, both concerning AI and its implications for human decision-making and institutional behavior. The analysis examines how these influences intersect across six themes: reliance, agency, accountability, governance, equity, and humanity. Its central argument is that the consequences of AI adoption depend on the interaction of multiple factors rather than any single characteristic of the technology or its users. Approaches that focus exclusively on accuracy, trust, or formal oversight may therefore overlook the wider conditions shaping how AI is accepted, challenged, and incorporated into decisions. By bringing diverse academic and practitioner perspectives into a shared analytical framework, this paper identifies tensions and questions for further investigation. It is an exploratory synthesis, not an empirical study, and does not independently establish the causal relationships or generalizability of the ideas discussed.

2. Introduction

Artificial intelligence is increasingly being integrated into decisions that shape education, professional work, organizational practices, and institutional governance. By processing information, generating recommendations, and automating tasks, AI systems can improve efficiency and expand the range of decisions people can make. However, their growing role raises a psychological question that extends beyond their technical capabilities: why do people follow AI recommendations, and under what circumstances do they question them?

Accuracy is an important part of the answer, but human reliance on AI may also be shaped by perceived credibility, confidence, convenience, workload, professional norms, institutional expectations, and the consequences of disagreement. An AI recommendation does not enter a neutral environment; it becomes part of an existing social and institutional structure in which authority, relationships, responsibilities, and expectations already influence human behavior. Consequently, the influence of AI may depend not only on the content of its recommendations, but also on how people interpret its role and understand their own responsibility for decisions.

This distinction becomes particularly important when AI shifts from serving as an additional source of information to becoming a source of authority that people defer to. Although delegation can improve efficiency, excessive reliance may weaken independent scrutiny, complicate accountability, and create challenges for institutions seeking to govern AI-assisted decisions. Understanding these consequences therefore requires examining AI not simply as a technological system, but as part of a broader behavioral and institutional environment.

3. Methodology

The above questions informed the virtual roundtable *When AI Enters the Room: Human Judgment, Responsibility, and Institutional Behaviour*, held on 30 July 2026 (GeneralCounsel360, n.d.), and the panel *When AI Makes the Decision* at the Privacy Decision Forum – India 2026 (Events 4 Sure, n.d.), held on 10 September 2026, where insights from the roundtable were presented in a practitioner-oriented context. Together, these discussions provided the foundation for examining recurring questions about human-AI interaction, responsibility, governance, and equity.

The materials from both settings were reviewed to identify recurring concepts, concerns, and tensions relating to human-AI interaction. These ideas were organized into six overarching themes: Reliance, Agency, Accountability, Governance, Equity, and Humanity. A cross-theme analysis was then developed to examine connections between these themes and identify broader implications and directions for future research. The resulting synthesis interprets ideas raised in the discussions rather than independently testing their validity; its conclusions should therefore be understood as exploratory rather than empirical findings.

4. Thematic Analysis

4.1 Reliance

4.1.1 AI as a New Authority

The discussions invite a comparison between deference to traditional authorities and reliance on AI-generated recommendations. Milgram's obedience experiments (Milgram, 1963) illustrate how people may follow instructions when they come from a figure presented as legitimate within an authoritative setting. The laboratory environment, the researcher's role, and other authority cues can make instructions seem legitimate. A related question is whether similar tendencies arise when AI systems are treated as authoritative sources in decision-making. As AI becomes increasingly integrated into educational, professional, and institutional settings, its recommendations may acquire influence beyond the information they provide. Examining AI as a potential source of authority therefore requires investigating how the circumstances in which people encounter and use these systems shape their willingness to defer to their recommendations.

4.1.2 The Social Psychology of Compliance

Compliance with authority can have several psychological and institutional sources. Organizations may reward compliance and penalize defiance, meaning obedience can reflect the environment people operate in rather than an inherent tendency to follow orders. Yet incentives alone may not explain compliance: belonging and group identity can also motivate people to conform to institutional expectations. Procedural fairness (Tyler, 1990) matters as well. People may be more willing to accept decisions when they feel heard, respected, given explanations, and treated impartially, even when the outcome is unfavourable. At the same time, authority need not imply blind obedience; institutions can retain leadership while allowing people to question decisions. Together, these perspectives suggest that compliance may arise through external incentives, the desire to belong, perceived fairness, and confidence in authority. For AI use, the key question is whether people accept a recommendation because it is justified, because the institution endorses it, or because challenging it feels socially or professionally costly.

4.1.3 Credibility and Human-Like Interaction

Acceptance of AI may be shaped both by people's general views of the technology and by their experiences with particular systems. The discussions raised the possibility that conversational AI's human-like qualities may influence how users approach it as a source of advice. Seemingly superficial features, such as a system's name or conversational style, may also affect acceptance. These responses may reflect familiar judgments of credibility and trustworthiness, as people often accept information from sources they regard as reliable. Similarly, people may apply the criteria they use to evaluate human decision-makers, including fairness, respect, and impartiality, when evaluating AI decisions. This raises an important distinction between a system actually being reliable and a user perceiving it as credible. A human-like interface cannot by itself establish the accuracy of information. Research could examine whether users' perceptions of an AI system's credibility accurately reflect the reliability of its recommendations, particularly when the system's human-like characteristics influence their judgments.

4.1.4 The Appeal of Agreeable AI

AI's agreeableness presents a tension between user engagement and perceived credibility. Accessible, accommodating chatbots can make interaction appealing, while sycophancy, excessive agreement with or accommodation of a user, may encourage repeated use. The discussions raised the possibility that commercial incentives could reinforce this tendency toward agreeable behavior if users engage less with systems that become less accommodating. Yet agreeableness may have the opposite effect for some users: excessive flattery or eagerness to please can provoke distrust, much as it might in a human leader. The relationship between agreeableness and trust is therefore unlikely to be uniform; it may depend on the user, the context, and the degree of accommodation. Pleasant interaction can make a system appealing while also obscuring whether it is evaluating a claim independently. A useful question is when

warmth and reassurance support effective communication, and when they become signals that users should scrutinize rather than accept automatically.

4.1.5 Automation Bias and Algorithmic Aversion

Two opposing responses can shape how people evaluate AI recommendations. Automation bias (Skitka et al., 1999; Lyell & Coiera, 2017) occurs when people place excessive confidence in AI and rely on it more than its actual reliability warrants. Algorithmic aversion (Dietvorst et al., 2015), by contrast, occurs when people reject a recommendation because it comes from a machine, potentially dismissing useful information without adequately evaluating it. These tendencies point in opposite directions, but either can undermine sound decision-making: one encourages excessive acceptance, while the other can produce unjustified rejection. This complicates the broad question of whether people trust AI. A person may accept one system and reject another, or trust AI for some tasks while resisting it for others. The aim should not simply be to maximize trust or distrust. Instead, people need to evaluate the particular task, source, and evidence. Further research could investigate the circumstances under which people become overly reliant on AI recommendations or reject them despite their potential usefulness.

4.1.6 Calibrating Trust in Inconsistent Systems

Several examples highlighted how difficult it can be to judge AI reliability when outputs vary. AI can generate multiple-choice questions or polished-looking feedback efficiently, but the apparent quality of an answer does not guarantee correctness. Checking every output may reduce some of the time saved, creating a practical tension between efficiency and verification. Different AI tools and paid or free versions can also produce different outputs, making a single standard of trust difficult to establish. At the conference, participants noted that different applications can answer the same question differently and recommended checking outputs against relevant regulations and authoritative sources. One example involved markedly different marks for the same schoolwork feedback, illustrating how AI evaluations may vary across tools or configurations. These examples suggest that trust must be calibrated to the specific system and task. Users need enough foundational knowledge to spot problems, and organizations may need verification standards that evolve as tools change.

4.2. Agency

4.2.1 The Cost of Cognitive Ease

Following another source's direction can be easier than independently evaluating a complex issue. Research on authority and leadership raises questions about how much effort people are willing to invest in resisting direction, while concepts such as ego depletion (Baumeister et al., 1998) have been proposed as one way of thinking about the difficulty of sustained self-control. People may also develop habits and may prefer not to engage in effortful deliberation when an easier alternative is available. AI can make that shortcut especially tempting: generating a large set of multiple-choice questions, for example, can be quick, and the results may look convincing

enough to accept without checking. This creates a practical tension. AI can reduce repetitive effort, but careful verification also requires attention and time. The issue is not that all cognitive offloading is harmful; delegating routine steps can be useful. The risk arises when the convenience of an answer leads users to stop considering whether it is accurate, suitable, or complete. A central question is how people can preserve meaningful evaluation while still benefiting from automation.

4.2.2 Preserving Independent Judgment

Across the discussions, a recurring distinction was between receiving assistance and surrendering judgment. AI should remain a tool rather than become a substitute for human decision-making. Workplace examples illustrated how interns may return AI-generated work quickly without necessarily understanding the task's underlying logic. Even when AI output looks more capable than a user's own work, the user still needs to evaluate what they submit or act upon. AI can also help people complete analyses faster, but the time saved should be used for work requiring judgment rather than passive acceptance of automated output. Inconsistent AI evaluations further highlight the need for foundational knowledge to assess results. Agency therefore involves more than retaining the final decision: users must understand enough about a task to recognize when AI may be wrong, decide which evidence matters, and take responsibility for how the output is used.

4.2.3 Learning Versus Delegation

A concern raised in the discussions was that people may gradually delegate more to AI, moving from using it to edit their own writing to outsourcing larger tasks they previously enjoyed doing themselves. This prompts a distinction between using AI to improve a skill and using it instead of developing that skill in the first place. Over time, this delegation may become habitual, with people automatically turning to AI even for activities they previously enjoyed doing independently. The effects of AI use may therefore depend on what the system replaces. Using AI to receive feedback or refine an existing ability may support learning, whereas using it to bypass the effort through which an ability develops may reduce opportunities for practice. These concerns do not prove that AI necessarily weakens skills. They nevertheless raise important questions for education and professional training: which tasks should learners complete independently, when should AI support the learning process, and how can we distinguish AI use that helps people develop skills from use that allows them to avoid developing those skills?

4.2.4 Productivity and the Changing Nature of Work

The conference discussions presented AI as a way to accelerate work while also raising questions about job design and staffing. One perspective was that the goal should be to increase productivity rather than simply replace jobs. AI can automate tasks such as spreadsheet analysis, potentially reducing work that previously took considerably longer; the time saved could then be directed toward additional work and decisions requiring executive judgment. Another example described AI agents being used across the software development

life cycle, including tasks previously performed by junior developers. Together, these accounts suggest that AI may change not only how quickly individual tasks are completed but also which tasks people perform, how junior employees develop practical skills, and how organizations allocate staff. The unresolved question is how to capture efficiency gains while preserving human expertise, meaningful work, and opportunities for professional development.

4.3. Accountability

4.3.1 The Diffusion of Responsibility

AI involvement may make it easier for people to distance themselves from the consequences of their decisions. Milgram's obedience experiments (Milgram, 1963) illustrate how people may perceive an authority figure as responsible when they are following instructions. When a decision made with AI assistance causes harm, a person or organization may similarly blame the algorithm rather than accept responsibility. Research on diffusion of responsibility (Darley & Latané, 1968) offers a way to examine whether people shift responsibility to AI, and to what extent. The central issue is a testable question: does the presence of AI change how people perceive their own responsibility, the system's contribution, or the obligations of others involved? The answer may depend on the task, the user's role, and the way responsibility is assigned within an institution.

4.3.2 Who Should Be Held Responsible?

It is important to distinguish descriptive questions about how people actually assign responsibility from normative questions about who should be responsible. Discovering that people blame AI, for example, would not establish that they ought to do so. The philosophical question of whether AI can have the normative status of an accountable agent also remains distinct from the practical question of how people behave when AI is involved. Accountability matters partly because people live together and must respond to the consequences of one another's actions. An important institutional question raised by these discussions is how responsibility should be allocated when AI contributes to a decision. AI involvement does not, by itself, resolve whether responsibility should rest with the user, developer, deployer, or organization. These perspectives converge on the need to clarify accountability while leaving open difficult questions about the responsibilities that might appropriately be assigned to AI developers, deployers, users, and institutions. A rigorous analysis should keep observed blame attribution separate from the ethical and institutional principles used to allocate responsibility.

4.3.3 Human Oversight and AI Self-Correction

Human responsibility remains important even when AI contributes substantially to a decision. In an AI-assisted contract review, for example, a reviewer cannot simply blame the tool if it misses a risky clause. An organization would similarly need to answer for a failure in a decision such as loan approval, even if AI contributed to the process. A related issue concerns whether AI systems should proactively acknowledge errors they discover in their previous outputs. In one

example, a newer AI model identified omissions and a possible misinterpretation in an earlier literature review, but only after being prompted to reassess the work. This prompted a comparison with the expectations placed on a graduate student, who would be expected to communicate relevant omissions discovered after completing a literature search. The comparison raises a normative question: should AI systems also be expected to disclose and correct errors they identify, even when users have not explicitly asked them to do so? This goes beyond whether AI can detect mistakes to consider whether systems should be designed to communicate discovered errors proactively.

4.4. Governance

4.4.1 Culture Over Formal Rules

Organizational culture can matter as much as, or more than, written policies in determining whether employees raise concerns. When leaders respond constructively to questions, employees may feel able to flag risks; when questioning is repeatedly criticized, people can learn to remain silent even when they privately disagree. One example from a legal department showed how open discussions allowed a business team to modify a proposed approach rather than overlook legal exposure. Institutions are also shaped by unspoken social cues: written policies establish expectations, but people learn by observing what colleagues and leaders actually do. A policy's existence therefore does not guarantee its effectiveness. Governance depends partly on whether everyday interactions reinforce its stated principles. For AI, the relevant question is whether employees can realistically challenge an output that conflicts with their expertise or raises ethical concerns, or whether workplace incentives encourage them to accept it in order to appear efficient, compliant, or loyal.

4.4.2 Making Dissent Safe

The ability to question AI and institutional decisions depends on whether dissent is safe. Employees may publicly agree despite privately disagreeing when a workplace discourages challenge. One proposed response is to make questioning a normal institutional practice by explicitly incorporating dissent mechanisms into governance structures, rather than relying solely on individual courage. People should not face social, political, or career-related penalties for raising concerns, and constructive dissent should be treated as valuable rather than disloyal. The costs of dissent are not distributed equally: people from marginalized groups may face greater professional, social, or personal consequences, especially when institutional norms do not account for intersecting identities. An effective dissent process must therefore do more than invite feedback. It must protect those who speak up, make challenges legitimate, and consider whose perspectives are more likely to be suppressed. The design of such mechanisms is both a governance issue and an equity issue.

4.4.3 Training for Responsible AI Use

Education and training can help people question AI without abandoning its benefits. People need to learn how to challenge AI and institutional decisions, while organizations should examine whether employee training promotes active scrutiny or passive acceptance. Training should not merely demonstrate how to operate a tool or produce outputs more quickly. It should also help users identify system limitations, verify important claims, recognize when foundational expertise is needed, and understand that responsibility remains with the person making or approving a decision. Efficiency does not automatically translate into competence: an employee can complete a task quickly without understanding its underlying logic. This creates a challenge for institutions, because training must develop practical AI literacy and independent professional judgment together. The discussions do not establish a single effective training model, but they suggest that responsible use requires more than access to technology or written rules. It requires repeated practice in questioning, checking, and making informed decisions.

4.4.4 Adaptive Governance and Verification Standards

AI policies must operate in an environment where tools and capabilities continue to change. Different systems, subscription tiers, and prompting practices can produce different outputs, making it difficult to define one permanent threshold for trust. Formal governance documentation provides a starting point, but spreading norms through a large institution or across a country may take considerable time and resources. This creates a tension between the need for clear standards and the need for policies that can adapt. Institutions need usable expectations about when AI may be used, what must be verified, and who remains responsible, but rules can become outdated as systems change. Governance therefore needs both formal structure and ongoing review. The practical question is how organizations can establish stable principles, such as transparency, verification, and human accountability, while updating their implementation as technologies and risks evolve.

4.5. Equity

4.5.1 AI and Structural Inequality

The discussions raised the concern that AI could magnify existing social inequalities, particularly when systems inadequately represent the experiences of marginalized communities. Because systems are trained on historical data, they may reproduce prejudices, structural inequalities, and power imbalances already present in society. Marginalized communities may bear disproportionate consequences when their experiences are inadequately represented or understood. AI-generated judgments can also acquire a modern appearance of authority, potentially making inherited biases seem more legitimate or neutral than they are. This connects to intersectionality (Crenshaw, 1989): people's experiences are shaped by the interaction of different social identities, and a system that fails to account for these differences may overlook important sources of disadvantage. These concerns do not imply that every AI system necessarily reproduces every historical bias, but they identify a governance risk that deserves scrutiny. Evaluating AI fairly requires asking not only whether an output appears consistent, but

whose experiences are represented, which groups may be harmed, and whether apparent neutrality hides unequal effects.

4.5.2 Unequal Access to AI

AI adoption varies across countries and institutions partly because resources, expertise, and infrastructure are unevenly distributed. In education, faculty members may avoid AI, pay for subscriptions themselves, or use free versions, creating different experiences and levels of trust. Students may likewise use tools with different capabilities without fully understanding how those differences affect outputs. This raises a concern for institutions evaluating AI-assisted work: two people may have unequal access to systems even when they are expected to meet the same standards. Unequal access may also affect who can benefit from more capable tools and who has the resources to learn how to use them well. The broader issue is how institutions can respond to differences in access and capability without assuming that everyone is working with the same technology or has equal opportunities to develop AI literacy.

4.5.3 Commercial Influence and Consumer Behavior

AI platforms may face commercial incentives similar to those of social media companies, encouraging users to spend more time on their services. Advertising through AI platforms could change search engine optimization and allow businesses to target products through specific systems. There is also a concern that tools optimized to retain users might become increasingly agreeable and influential, shaping what consumers consider or want before they make an explicit purchase. These possibilities connect commercial incentives with the design of AI interactions: a platform's objectives may not always align with a user's interest in receiving balanced, independent advice. The central question is how advertising, recommendations, and engagement incentives should be disclosed and governed when AI becomes an intermediary between consumers and information. This issue is distinct from algorithmic bias alone because even a system without an obvious demographic bias could still be shaped by commercial priorities.

4.6. Humanity

4.6.1 The Illusion of Empathy

AI can provide empathetic-sounding support without genuinely experiencing empathy. One example involved a virtual assistant modeled on a faculty member that offered around-the-clock support to students who felt unsuccessful or unsure they belonged, while directing them toward university support services. The system could reproduce reassuring responses without possessing the underlying emotional commitment or the full capacity to help that a human might provide. A related psychological distinction is that perceived respect and concern can still influence a person's experience even when the source does not actually feel those emotions. The important distinction is between a system's internal emotional state and the effect of its outward communication on a person. These discussions do not settle whether simulated

empathy is beneficial or harmful. They suggest instead that the consequences depend on how people interpret the interaction, what assistance the system can actually provide, and whether users understand the limits of the support being offered.

4.6.2 AI and Human Connection

AI may influence human relationships in opposing ways. Consistently reassuring AI companions might become more attractive than human interaction, which can involve disagreement, uncertainty, and the risk of rejection. This raises concerns about social anxiety and reduced willingness to interact with other people. People accustomed to agreeable AI responses may also become less willing to engage with humans who challenge their assumptions. Yet another possibility is that people could learn from AI's supportive language and become more considerate, empathetic communicators themselves. Technology should not distance people from collective responsibility and authentic human connection. These are competing possibilities rather than established outcomes. AI may provide useful support, particularly when human help is unavailable, while also changing expectations about what comfortable interaction feels like. The important question is how systems can support people without encouraging them to substitute frictionless simulated relationships for the reciprocal, sometimes difficult relationships through which human connection develops.

4.6.3 Creativity and the Future of Human Contribution

A concern raised in the discussions was that people may gradually delegate activities to AI that they previously found rewarding, including writing and other creative tasks. This prompts a distinction between using AI to improve an existing skill and using it instead of developing or exercising that skill. A related concern is that if people increasingly rely on AI-generated content, the creation of new human-produced material could decline, potentially affecting systems that depend on existing data. This was raised as a prospective concern about model collapse (Shumailov et al., 2024), not an established outcome. These ideas connect individual behavior with a wider information ecosystem: if people outsource more creative work, they may lose opportunities to practice it, and the amount or character of new human-created material could change. Conversely, AI may also support creativity by helping people develop ideas or complete routine steps. The key distinction is whether the technology expands human creative capacity or gradually displaces the independent contribution on which learning and future innovation may depend.

4.6.4 Uncertainty About AI's Future

AI's future remains highly uncertain: capabilities that once seemed difficult to anticipate have already emerged, and the direction of future development is not known. This uncertainty runs through concerns raised in the discussions, including possible changes to employment, creativity, relationships, model collapse (Shumailov et al., 2024), and the reliability of governance rules as systems evolve. For example, the possibility that AI agents may replace some junior-level tasks was described through an organizational example, while model collapse

(Shumailov et al., 2024) was raised as a potential future risk rather than a demonstrated inevitability. Recognizing uncertainty does not mean institutions should avoid making decisions. Rather, they should distinguish observed experiences from forecasts, monitor consequences, and revise policies as evidence changes. Together, these examples underscore the need for open, evidence-sensitive inquiry as AI becomes more embedded in social and institutional life.

5. Cross-theme analysis

5.1 From trust to dependence

Trust in AI extends beyond whether users consider its outputs credible; it also concerns how they decide when to rely on them and when to exercise independent judgment. The discussions highlighted a tension between trust and verification: AI can provide useful answers efficiently, yet inconsistent outputs and the possibility of error make scrutiny necessary. Verification, however, introduces additional effort into the process that makes AI attractive. A related tension emerges between efficiency and competence. Completing a task more quickly does not necessarily mean understanding the reasoning behind it, particularly when AI performs substantial portions of the work. Taken together, these tensions suggest that the value of AI reliance cannot be assessed through convenience or output quality alone. It also depends on whether users retain the knowledge and judgment needed to evaluate the results. The central challenge is therefore to use AI in ways that improve efficiency without allowing the convenience of delegation to undermine independent reasoning.

5.2 From productivity to the distribution of benefits

The integration of AI into professional environments raises questions not only about how work is performed, but also about how its benefits and costs are distributed. The discussions presented productivity as an opportunity for workers to move beyond repetitive tasks and devote more time to higher-level judgment. However, examples of AI agents performing work previously assigned to junior developers raised the possibility that automation could also reduce demand for certain roles. This creates a tension between productivity and employment: improvements in organizational efficiency do not necessarily translate into benefits for every worker. The consequences may depend on how organizations redistribute tasks, restructure teams, and create opportunities for employees to develop more advanced skills. In particular, automating entry-level work may have implications for how newcomers acquire the experience needed to progress professionally. These considerations suggest that the impact of AI should not be evaluated solely through the speed or volume of work produced, but also through its effects on employment opportunities and the distribution of productivity gains.

5.3 From Individual Judgment to Institutional Responsibility

Responsible AI use depends on both the judgment of individuals and the institutions in which decisions are made. Individuals must be able to evaluate AI-generated recommendations critically, recognize potential errors, and understand the decisions they make using these

outputs. However, this responsibility cannot be separated from the organizational environments that shape their behavior. Workplace expectations, limited training, pressure to work efficiently, and the social costs of challenging authority may discourage employees from questioning AI, even when they have legitimate concerns. This creates a connection between individual judgment and institutional responsibility: employees need the knowledge and freedom to scrutinize AI outputs, while organizations must establish clear accountability, provide appropriate training, and create conditions in which questioning recommendations is encouraged rather than penalized. Institutional policies alone may be insufficient if everyday workplace practices reward unquestioning compliance. Ultimately, responsible AI integration requires more than expecting individuals to make sound decisions; it also requires institutions to create the conditions in which independent judgment, meaningful oversight, and accountability can be exercised.

5.4 From formal rules to the social conditions for dissent

Effective AI governance depends not only on the existence of formal policies, but also on whether people can meaningfully challenge decisions made with AI assistance. The discussions highlighted a tension between authority and dissent: institutions need clear decision-making structures, yet employees may hesitate to question AI-supported recommendations when disagreement carries social or professional costs. Even well-designed oversight procedures may therefore prove ineffective if organizational culture discourages scrutiny. At the same time, governance faces a tension between policy and adaptability. As AI systems and their capabilities evolve, rules that once provided useful guidance may become outdated or insufficient. These challenges are interconnected because institutions must establish clear expectations while remaining responsive to changing technologies and workplace practices. The discussions consequently point toward governance that combines explicit standards with regular review and an organizational culture that treats constructive questioning as a legitimate part of decision-making. Without both, policies may exist on paper while failing to ensure that AI-assisted decisions receive meaningful scrutiny in practice.

5.5 From technical capability to unequal consequences

The benefits of AI may be distributed unevenly because access to its capabilities depends partly on the resources, infrastructure, and expertise available to different individuals and institutions. The discussions raised concerns that differences in access to AI systems, including differences between free and paid tools, could create unequal opportunities to benefit from their capabilities. This creates a tension between access and fairness: although AI may expand the availability of information and assistance, unequal access may reinforce existing disparities rather than eliminate them. Differences in expertise may further affect whether individuals can use available tools effectively and evaluate their outputs critically. These concerns suggest that fairness in AI adoption cannot be assessed simply by asking whether a technology is available or whether it performs a task successfully. It also requires examining who can access its capabilities, under what conditions, and with what opportunities to benefit. The broader implication is that responsible AI adoption must consider the distribution of resources and opportunities alongside

the technology's potential advantages, rather than assuming that greater availability will automatically produce more equitable outcomes.

6. Implications and Future Research

The discussions identified several directions for future research into how people evaluate, rely on, and interact with AI within decision-making environments. One area concerns the criteria people use to accept or doubt AI-generated information, including the influence of AI's apparent empathy and capacity for connection. The discussions also raised the possibility that people may increasingly delegate activities they previously enjoyed because AI makes them easier. A controlled experiment could investigate this possibility, while comparisons between human management, interaction, and leadership and their AI-mediated counterparts could clarify how behavior changes when AI takes on functions traditionally performed by people. These questions suggest the need to examine not only whether people accept AI, but also how its use may change their engagement with activities and their relationships with others.

A second direction concerns the relationship between AI reliance, critical thinking, and the development of expertise. The discussions raised questions about whether extensive reliance on AI might weaken independent critical thinking or hinder younger professionals from developing the expertise required in their fields. They also highlighted the importance of examining how organizations train employees to use AI while continuing to question its recommendations. Future research could investigate whether completing tasks with AI translates into independent competence and how different forms of training influence users' willingness to scrutinize AI-generated outputs. In education, the discussions of AI-based oral assessment also raise questions about whether alternative assessment methods can provide meaningful evidence of student learning while making effective use of AI.

A third direction concerns responsibility, dissent, and institutional governance. The discussions raised questions about how responsibility is understood when AI contributes to a decision, whether people are willing to challenge AI-supported recommendations, and how organizational environments influence these responses. Research could compare reactions to equivalent recommendations from human and AI sources, examining acceptance, willingness to disagree, and perceived responsibility when outcomes are uncertain or incorrect. The proposal to develop a shared charter for AI use also suggests investigating whether collaboratively established principles influence how people use, question, and oversee AI within institutions. Such research could help clarify how formal governance arrangements interact with workplace norms and individual behavior.

Finally, the discussions highlighted potential bias and unequal access to AI capabilities, including the possibility that greater financial resources may provide access to more capable or less biased systems. This raises questions about how differences in system capability, resources, and access affect decision quality and the distribution of AI's benefits. Future studies could compare system performance and examine whether disparities in access translate into unequal opportunities or outcomes, rather than assuming that cost alone determines reliability

or fairness. Taken together, these research directions point toward a central question: how can AI be integrated into educational, professional, and institutional settings while preserving critical thinking, meaningful human oversight, and equitable access to its benefits? The discussions did not resolve these questions; instead, it identified avenues for future investigation that require systematic empirical study.

7. Scope of Interpretation

This paper draws on discussions from two connected settings: a virtual roundtable and a practitioner-oriented conference. The themes identified are therefore shaped by the subjects raised and perspectives represented in these discussions. Other relevant dimensions of human-AI interaction may not have been addressed.

The organization of discussion material into six themes and the subsequent identification of cross-theme tensions involved analytical judgment. Although the framework aims to capture recurring concepts in the discussions, alternative interpretations or thematic groupings may also be possible.

The paper synthesizes observations, arguments, and questions raised during the discussions but does not independently test them through experiments, surveys, or other systematic empirical methods. Consequently, it cannot establish causal relationships or determine whether the proposed behavioral patterns occur as described.

The discussions do not establish how widely the ideas examined apply across different populations, cultures, professions, or institutional environments. The extent to which the identified themes reflect broader patterns of human-AI interaction remains an open question requiring further investigation.

8. Conclusion

Understanding human-AI interaction requires moving beyond explanations based on the technology alone and recognizing the many psychological, social, institutional, and economic factors that shape how people respond to it. The discussions reveal that these factors do not operate independently: their influence may vary across individuals, tasks, systems, and organizational environments, while sometimes reinforcing or contradicting one another. Consequently, responsible AI integration cannot be reduced to maximizing trust, improving technical performance, or establishing formal policies in isolation. It requires considering how these different influences interact and how their consequences vary across contexts. Although this synthesis does not establish definitive causal relationships, it highlights the importance of examining human-AI interaction from multiple perspectives and investigating how their combined effects shape the role of AI in human decision-making.

9. References

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5), 1252–1265. <https://doi.org/10.1037/0022-3514.74.5.1252>

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1), Article 8. <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8(4, Pt. 1), 377–383. <https://doi.org/10.1037/h0025589>

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>

Events 4 Sure. (n.d.). Privacy Decision Forum – India 2026. <https://www.events4sure.com/privacy-decision-forum-india2026>

GeneralCounsel360. (n.d.). When AI enters the room: Human judgment, responsibility, and institutional behaviour. <https://www.generalcounsel360.com/when-ai-enters-the-room>

Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association*, 24(2), 423–431. <https://doi.org/10.1093/jamia/ocw105>

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371–378. <https://doi.org/10.1037/h0040525>

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631, 755–759. <https://doi.org/10.1038/s41586-024-07566-y>

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51(5), 991–1006. <https://doi.org/10.1006/ijhc.1999.0252>

Tyler, T. R. (1990). *Why people obey the law*. Yale University Press.

10. Author's Note

This paper represents independent research conceived, developed, and written by Parth Parnami. It reflects his independent analysis and learning from discussions held during the

When AI Enters the Room: Human Judgment, Responsibility, and Institutional Behaviour virtual roundtable on 30 July 2026 and the panel When AI Makes the Decision at the Privacy Decision Forum – India 2026 on 10 September 2026. The paper presents an exploratory thematic synthesis of perspectives shared by academics, professionals, and industry practitioners. It is not an empirical study; the themes and questions presented are interpretive and have not been independently tested through systematic empirical methods.