

## Trust and Continuance Intention Toward AI-Powered Research Tools: The Moderating Role of Transparency Framing Style.

Mohammad Dalvi-Esfahani <sup>1,2\*</sup>, Sharanjit Kaur Bhathal Singh <sup>1</sup>, Mohammad Falahat <sup>2</sup>, Mahdi Mohamed Omar <sup>3</sup>, Ali Najeeb <sup>4</sup>

<sup>1</sup> Department of Information Systems, Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Kampar Campus, Perak, Malaysia.

<sup>2</sup> Strategic Research Institute, Asia Pacific University of Technology and Innovation, Kuala Lumpur, Malaysia.

<sup>3</sup> Faculty of Economics and Management, Jamhuriya University of Science and Technology, Mogadishu, Somalia.

<sup>4</sup> Qasim Ibrahim School of Business, Villa College, QI Campus, Rahdhebai Magu, Male, Maldives.

**\*Corresponding Author:** [dalvi@utar.edu.my](mailto:dalvi@utar.edu.my)

Received: 2026-09-11; Accepted: 2026-09-27; Published: 2026-09-30

### Abstract

As AI becomes more deeply embedded in scientific workflows, understanding what drives user trust has become central to the sustainable adoption of these tools. This study draws on the Ability-Benevolence-Integrity (ABI) model to examine what shapes trust in AI-powered research tools and how that trust relates to users' intention to continue using them. A key contribution is the introduction of Transparency Framing Style (TFS) as a novel construct. Whereas transparency refers broadly to how much information an AI system discloses about its data, reasoning and limitations, TFS concerns how that information is structured, emphasised and worded when presented to users. Data were collected from 274 academics and postgraduate researchers across 10 Malaysian universities using purposive sampling and analysed through partial least squares structural equation modelling (PLS-SEM). The structural model explains 68.6% of the variance in trust and 39.8% in continuance intention. Among the ABI dimensions, integrity showed a larger estimated association with trust ( $\beta = 0.251$ ,  $p < 0.001$ ) than benevolence ( $\beta = 0.081$ ,  $p = 0.047$ ). TFS was the strongest direct predictor of trust ( $\beta = 0.531$ ,  $p < 0.001$ ) and positively moderated the relationship between trust and continuance intention ( $\beta = 0.180$ ,  $p < 0.001$ ). These findings extend the human-AI interaction literature by highlighting the importance of interpretive compatibility alongside technical competence in sustaining engagement with AI-powered research tools.

**Keywords:** *AI-Powered Research Tools; User Trust; Continuance Intention; Transparency Framing Style; ABI Model.*

### 1. Introduction

In recent years, scientific research workflows have been significantly transformed because of advancements in artificial intelligence (AI), particularly large language models (LLMs) and generative AI (Ding et al., 2025). A wide range of research-related tasks, such as conducting literature searches, summarizing articles, developing research ideas, and analyzing data, as well as research paper drafting, can be supported by AI-powered research tools (Marabelli et al., 2026). These AI tools are designed to

increase the performance of researchers and further facilitate their access to expanding knowledge, which is now integrated into their everyday workflow. In an environment where the volume of research knowledge production is increasing with unprecedented speed, AI-powered research tools can play a significant role in managing information complexity and further accelerating the process of knowledge development (Andersen et al., 2025). However, the success and sustainable integration of these tools into academic research activities depend not only on their technical capabilities but also on the degree to which researchers trust these AI-powered tools.

Trust is a fundamental concept in the literature on technology adoption and human interaction with intelligent systems, as users often decide to use a system in situations associated with uncertainty and risk (Lee & See, 2004; McKnight et al., 2002). In the context of AI, this uncertainty is intensified due to the complex nature of AI systems and their opaque algorithms (Burrell, 2016). In recent years, the use of AI tools in academic environments has rapidly increased (Zou & Huang, 2023); however, several significant concerns have been raised regarding their reliability, accuracy, and transparency of their outputs (Bender et al., 2021). For example, recent studies have reported that generative AI systems may produce false information (Bubeck et al., 2023), fabricated references (Walters & Wilder, 2023), or make arguments that are not scientifically reliable. These challenges have made researchers use AI tools cautiously, which further makes them skeptical about the reliability and trustworthiness of their outputs. Hence, the current situation presents a paradox in the academic context: on the one hand, the use of AI-powered tools is expanding; on the other hand, researchers' level of trust is not yet sufficient to rely on them completely. However, as AI tools transition from automated back-office assistants to proactive co-creators in knowledge-intensive domains, initial adoption metrics fail to capture the nuances of long-term technology retention. In high-stakes academic workflows, users face distinct "trust tensions" characterized by system opacity versus output performance, exposing researchers to unprecedented epistemic and professional risks (Hanelt et al., 2026). Consequently, sustaining long-term user engagement requires moving beyond basic cognitive utility to examine the structural formation of continuous trust (Kang et al., 2024; Khan et al., 2026).

To explain the formation of trust in interactions with AI systems, researchers often draw on established and empirically validated theoretical frameworks from the trust literature. One of the well-established and highly cited trust frameworks is the Ability-Benevolence-Integrity (ABI) model proposed by Mayer et al. (1995). This model hypothesises that trust is shaped when users evaluate an agent or system as reliable in terms of its competence in performing tasks, its benevolent intentions toward users' interests, and its adherence to principles of honesty and fairness. These three dimensions; ability, benevolence, and integrity; collectively shape users' perceptions of trustworthiness, which ultimately determine whether they trust a system. The ABI trust model has been used widely in various contexts to study trust formation, such as e-commerce (Chen & Dhillon, 2003), automation systems (Hoff & Bashir, 2015), and AI (Toreini et al., 2020), and is known as one of the valid theoretical models for explaining the formation of trust in human interactions with non-human agents.

However, trust formation in interactions with AI tools is not solely determined by the internal technical parameters of these systems (Lee & See, 2004). The way information about the performance of the system is presented can also play a crucial role in forming users' perceptions. Accordingly, transparency has been raised as one of the key determinants of developing trustworthy AI systems (Doshi-Velez & Kim, 2017). Transparency is the degree to which an AI system discloses information about its operations, decision-making logic, and limitations, helping users understand how its outputs are produced (Durán & Pozzi, 2025). Several studies have reported that increasing transparency of AI systems can reduce users' uncertainty and further enhance their trust in AI systems (e.g., Fehr et al., 2024; Glassberg et al., 2025; Radanliev, 2025). The main argument of the current study is that how transparency information is framed can determine how users interpret the information provided and, consequently, how much they trust the system and how willing they are to continue using it.

Studies of AI transparency have so far concentrated on the amount of information a system discloses. Communication research, however, has long shown that identical information can be judged differently

depending on how it is presented (Entman, 1993; Tversky & Kahneman, 1981). An AI research tool that reports its limitations, for example, may present them as a routine feature of responsible use or as a defensive disclaimer, and a researcher deciding whether to rely on the output is unlikely to read these two presentations in the same way. It is this presentational layer, rather than disclosure as such, that the present study addresses through the construct of Transparency Framing Style (TFS).

Three gaps motivate the present study. First, although research on trust in AI has grown considerably, most of it concerns autonomous systems (Hatherall & Sethi, 2025; Szatmary, 2025), healthcare (Alharbi et al., 2026; Alonso et al., 2025) or general human–AI interaction (McGrath et al., 2025; Ueno et al., 2022), and few studies have examined how academics come to trust AI-powered research tools (Bach et al., 2024). Second, existing work concentrates mainly on initial adoption, whereas continuance intention offers a better indication of whether a technology will be sustained in practice (Bhattacharjee, 2001); how trust in these tools carries forward into post-adoption use therefore remains unclear. Third, transparency has largely been studied as a system property that is either present or absent, and the way transparency information is framed has received little empirical attention (Fisher & Hopp, 2020), particularly as a condition that may strengthen or weaken the link between trust and continued use. Accordingly, the current study seeks to understand trust formation and continuance intention toward AI-powered research tools among academics through the lens of the ABI trust model, together with the direct and moderating roles of TFS.

The aim of this study is to explain how academics come to trust AI-powered research tools and how that trust relates to their intention to continue using them. Three objectives follow from this aim: (i) to examine how the perceived ability, benevolence and integrity of AI-powered research tools relate to academics' trust in these tools; (ii) to examine how trust relates to continuance intention, and whether trust carries the effects of the ABI dimensions through to continuance intention; and (iii) to conceptualise TFS and to test its direct relationship with trust and its moderating role in the relationship between trust and continuance intention. Accordingly, the study addresses three research questions:

**RQ1:** *To what extent do the perceived ability, benevolence and integrity of AI-powered research tools explain academics' trust in these tools?*

**RQ2:** *To what extent does trust explain academics' continuance intention toward AI-powered research tools, and does trust transmit the effects of ability, benevolence and integrity to continuance intention?*

**RQ3:** *How is TFS related to trust, and does it moderate the relationship between trust and continuance intention?*

By answering the research questions, the current study contributes theoretically and practically to the current body of knowledge. First, by focusing on the academic context, this study aims to expand research on trust in AI-human interactions to an area of growing importance that has not yet been sufficiently investigated. Second, by applying the ABI trust model, this research makes a theoretical contribution by elucidating the formation of trust in AI-powered research tools. Third, by introducing TFS as a moderating variable, this study provides new insights into the role of communication strategies in designing trustworthy AI systems. Finally, by focusing on continuance intention rather than initial intention or willingness, this study contributes to understanding the factors that sustain long-term engagement with AI-powered research tools, addressing a gap in the current literature on AI-human interaction.

## 2. Literature Review

### 2.1. Prior Studies

Research on trust and continuance intention toward AI-powered systems has expanded rapidly, yet it has not converged on a settled account of how transparency, trust and continued use relate to one another. This section organised prior studies around three questions on which the evidence remains divided: (i) whether transparency raises or lowers trust in AI, (ii) through which routes trust in AI is formed, and (iii)

under what conditions trust carries over into continued use. Table 1 summarises the studies reviewed and the question to which each contributes.

### 2.1.1. Does transparency raise or lower trust?

Evidence on whether transparency builds trust in AI is divided. A substantial body of work reports that explanations and disclosures reduce uncertainty and raise trust. Wang and Ding (2024) found that AI explanations improved behavioural trust and decision accuracy, and Lee and Cha (2025), separating transparency into accuracy and clarity components, showed that predictability in particular reduced systemic mistrust among ChatGPT users. In workplace settings, Glassberg et al. (2025) observed that systematic transparency still acted as a meaningful mediator of trust, even though users' pre-existing perceptions were resistant to change. Silalahi (2025) reached a similar conclusion for sustainability-oriented generative AI use, where interaction transparency and data clarity strengthened trust, although structural complexity was found to weaken it. Other studies point in the opposite direction. Across thirteen experimental designs, Schilke and Reimann (2025) identified a transparency dilemma in which openly disclosing AI involvement lowered trust by undermining perceived legitimacy, irrespective of task type. Wang and Ding (2024) themselves detected a placebo effect, with less technically capable users responding with greater trust even when the explanations they received carried no genuine information.

These results are difficult to reconcile as long as transparency is treated as a quantity that is either present or absent. They become more coherent once the presentation of transparency is taken into account. Using eye-tracking, Varnum et al. (2025) showed that the position and contrast of transparency disclosures on an interface shaped both cognitive attention and ethical perception. Lee et al. (2025) found that the way AI systems framed their own errors affected user tolerance and continuance intention after hallucination events, and Senadheera et al. (2026), drawing on a cross-national survey, reported that persuasive framing of responsible AI practices strengthened trust and lowered perceived risk, with the effect depending on local levels of institutional trust. Read together, this work suggests that the effect of transparency depends less on whether information is disclosed than on how it is framed. It is on this premise that the Transparency Framing Style construct is built.

### 2.1.2. Through which routes is trust in AI formed?

A second line of research asks whether trust in AI is essentially a cognitive judgement of performance or whether it also has an affective, relational component. Pathak and Bansal (2024), working with a multidimensional trust framework in service interface contexts, found that output quality built cognitive trust while social and relational factors opened a separate affective channel. Kim et al. (2025) reported a comparable dual pattern among undergraduates, for whom perceived responsiveness shaped both learning motivation and emotional attachment, and these together drove continuance intention. Pitts and Motamedi (2025) went further: their experimental evidence suggests that chatbot interactions produce a qualitatively different kind of trust from that generated by traditional technologies. If trust in conversational AI rests partly on relational cues, then the way a tool communicates, and not only what it produces, becomes a legitimate basis on which users form trust.

Studies that locate trust within institutional and organisational settings support this relational view from another angle. Korzyński et al. (2026) found that structural framing among professionals did not drive usage directly but served as a bridge toward institutional trust. Through a longitudinal case study in banking, Hanelt et al. (2026) traced how domain experts managed the tension between opacity and performance across three progressive stages of vulnerability, presenting trust as something that evolves over time rather than a fixed evaluation. In higher education, Alotaibi (2025) identified personal technical knowledge as a strong mediator of technology trust and task performance, while Verboom et al. (2025) used a socio-technical systems approach to map institutional stakeholders' readiness and their divided anxieties. What these studies have in common is a picture of trust as relational and dynamic. None of them, however, examines researchers using AI tools for their own knowledge work, a setting in which the professional cost of misplaced trust is unusually high.

### 2.1.3. When does trust carry over into continued use?

A third group of studies considers what sustains use once trust has formed. Accuracy appears to be central. Xiong et al. (2026) identified response accuracy as the primary upstream driver of system trust and the bridge to continued use, and Ding et al. (2026) found that transparency and accuracy together reduced users' fears of algorithmic opportunism, thereby protecting continuance intention. User control and data handling form a second set of conditions. Rahman and Liang (2026) demonstrated that restricting user control over AI outputs triggered psychological reactance, producing immediate distrust and a broader aversion to the system. Huynh and Aichner (2025) showed that information sensitivity heightened privacy concerns while meaningful control over user data restored systemic trust, and Cloarec et al. (2024) noted that personalisation delivers immediate utility even as it amplifies privacy anxieties. The error-framing results of Lee et al. (2025), discussed above, add that the consequences of system failures for continued use depend partly on how those failures are communicated. This body of work establishes that the relationship between trust and continued use is conditional. Yet the conditions examined so far are, for the most part, properties of the system itself, namely its accuracy, the control it allows and the way it handles data. Whether the communicative presentation of transparency conditions the same relationship has not been tested.

### 2.1.4. Synthesis

The literature supports three conclusions that shape the present study. (i) Transparency does not uniformly strengthen trust, and the conflicting findings become intelligible once the framing of disclosures is considered. (ii) Trust in conversational AI rests on relational and communicative cues as well as on output quality, which gives the presentation of information a theoretical basis for influencing trust. (iii) The link between trust and continued use is conditional, but the conditions studied to date are features of the system rather than features of how the system communicates. Two gaps follow from this. The framing of transparency has been manipulated in individual experiments but has not been conceptualised as a construct that users perceive and that can be measured; and its role as a condition on. Table 1 captures the synthesis of extant studies.

**Table 1: Summary of prior studies on transparency, trust and continuance intention in AI**

| Study                      | Context / design                                     | Key finding                                                                                                        |
|----------------------------|------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| Wang and Ding (2024)       | AI explanations                                      | Explanations improved behavioural trust and decision accuracy; placebo effect among less technically capable users |
| Lee and Cha (2025)         | ChatGPT users                                        | Accuracy and clarity components of transparency; predictability reduced systemic mistrust                          |
| Glassberg et al. (2025)    | Workplace AI                                         | Prior perceptions resisted change; systematic transparency still mediated trust                                    |
| Silalahi (2025)            | Sustainability-oriented GenAI use                    | Interaction transparency and data clarity strengthened trust; structural complexity weakened it                    |
| Schilke and Reimann (2025) | Disclosure of AI involvement; thirteen experiments   | Transparency dilemma: disclosure lowered trust by undermining legitimacy                                           |
| Varnum et al. (2025)       | Transparency disclosures on interfaces; eye-tracking | Position and contrast of disclosures shaped attention and ethical perception                                       |
| Senadheera et al. (2026)   | Responsible AI practices; cross-national survey      | Persuasive framing strengthened trust and lowered perceived risk, contingent on institutional trust                |
| Pathak and Bansal (2024)   | Service interfaces                                   | Output quality built cognitive trust; social and relational factors built affective trust                          |
| Kim et al. (2025)          | Undergraduate students                               | Perceived responsiveness shaped learning motivation and emotional attachment, driving continuance intention        |
| Pitts and Motamedi (2025)  | Chatbots; experimental                               | Chatbot interaction produced a qualitatively different kind of trust from traditional technologies                 |
| Korzyński et al. (2026)    | Professionals                                        | Structural framing did not drive usage directly but bridged toward institutional trust                             |
| Hanelt et al. (2026)       | Banking; longitudinal case study                     | Experts managed opacity-performance tension across three stages of vulnerability                                   |

|                          |                                                                 |                                                                                                      |
|--------------------------|-----------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| Alotaibi (2025)          | Higher education                                                | Personal technical knowledge a strong mediator of technology trust and task performance              |
| Verboom et al. (2025)    | Higher education stakeholders; socio-technical systems approach | Mapped institutional readiness and divided anxieties                                                 |
| Xiong et al. (2026)      | AI systems                                                      | Response accuracy the primary driver of system trust and bridge to continued use                     |
| Ding et al. (2026)       | AI systems                                                      | Transparency and accuracy reduced fears of algorithmic opportunism, protecting continuance intention |
| Rahman and Liang (2026)  | User control over AI outputs                                    | Restricted control triggered reactance, distrust and aversion                                        |
| Huynh and Aichner (2025) | Generative AI                                                   | Information sensitivity heightened privacy concern; data control restored systemic trust             |
| Cloarec et al. (2024)    | Personalisation                                                 | Personalisation delivered utility while amplifying privacy anxieties                                 |
| Lee et al. (2025)        | Hallucination events                                            | Error framing shaped user tolerance and continuance intention                                        |

Table 1 shows that the relationship between trust and continuance intention has not been examined, least of all among academics who rely on AI tools for scholarly work. The present study addresses both gaps. Research on trust and continuance intention toward AI-powered systems has grown quickly, drawing on varied theoretical lenses that collectively treat transparency as a pivotal yet contested driver of user behavior.

## 2.2. ABI Model of Trustworthiness

The ABI model of trustworthiness, introduced by Mayer et al. (1995), has become one of the most influential frameworks in the study of trust. At its core, the model suggests that people judge whether someone; or something; is worthy of trust based on three key dimensions: ability, benevolence, and integrity. Together, these dimensions capture the skills, intentions, and moral character that shape how trust is formed. Over the years, the ABI model has been applied and validated across a wide range of fields, from organizational behavior (e.g., Sondern & Hertel, 2024) to information systems (e.g., McKnight et al., 2002) and even human-computer interaction (e.g., Toreini et al., 2020). Because of its versatility, it remains a cornerstone for understanding how trust develops in both personal relationships and technology-driven environments. In this study, the model serves as a guiding lens to explore how individuals assess the trustworthiness of AI-powered research tools.

Ability is essentially about whether someone; or something; can do the job well. It refers to the skills, expertise, and knowledge that allow a trustee to act reliably and effectively. In academic terms, Mayer et al. (1995) describe ability as the capacity of an agent to achieve intended outcomes within a given context. When people judge trustworthiness, they often look at evidence of past performance, demonstrated competence, and relevant experience (Hoff & Bashir, 2015). For technological systems like AI, ability shows up in measurable qualities such as accuracy, predictive strength, and overall functionality (Lockey et al., 2021). A system that consistently performs well is seen as more capable of handling its tasks, which makes users more willing to depend on it. In this way, ability directly shapes trust; when stakeholders believe the trustee has the necessary competence, they are more inclined to place confidence in its decisions and actions (McKnight et al., 2002).

Benevolence is about whether we believe someone/something truly cares about others and acts with kindness (Mayer et al., 1995). It is the sense that a person; or even a system; is motivated by goodwill, not just self-interest (McAllister, 1995). In organizations, benevolence often shows up as empathy, fairness, and a genuine effort to support and protect the people they work with (Eisenberger et al., 1986). With technology, especially AI, benevolence takes on a new meaning. It's tied to how well systems are designed to align with human values and put people first (Friedman & Kahn, 2003). Tools that prioritize user well-being, minimize harm, and respect human dignity are more likely to be seen as benevolent (Shahriari & Shahriari, 2017).

At its core, integrity is about reliability in character. It is the reputation an individual or a system builds by staying fair and honest across the board (Mayer et al., 1995). When one's actions (person or system) align with ethical standards every single time, people perceive them as having true integrity. Within the trust literature, integrity is closely associated with sustained ethical conduct and reliability over time (Simons, 2002). Reliable systems earn trust by staying consistent, following their own rules, and keeping their processes transparent (High-Level Expert Group on Artificial Intelligence, 2019). Technical integrity relies on systems that stay open, explain their own logic, and take ownership of the results (Floridi & Cowls, 2022).

Previous research on trust in AI relies predominantly on the ABI theory, with its focus on perceptions of ability, benevolence, and integrity as measures of trustworthiness. Converging evidence from empirical studies supports the importance of these dimensions in the construction of human–AI trust. For instance, Liu et al. (2025) show that AI systems which are perceived as benevolent; ones that are designed with the wellbeing of users in mind; greatly increase users' trust and willingness to work together. Similarly, Novozhilova et al. (2024) demonstrate that although AI is commonly reported to possess remarkable capabilities, concerns over benevolence and integrity persist, resulting in a complex and at times contradictory trust dynamic. In organizational settings, Li and Bitterly (2024) point out that the perceived absence of emotional capacity in AI negatively influences benevolence perceptions and therefore trust, especially in empathy-intensive situations. Broadening, Dang and Liu (2025) argue that trust not only pertains to AI systems, but the user as well, which is once again moderated by ability, benevolence, and integrity in certain situations. Moreover, Cheung and Ho (2025) demonstrate that explainability boosts all three, thereby bolstering trust and intentions.

Ability, understood here as accuracy in retrieving, summarising and analysing scholarly material, is a precondition for use, yet researchers are trained to verify outputs and may treat competence as a baseline expectation rather than as a basis for trust. Benevolence is conceptually more difficult to apply to a non-human agent, because goodwill presupposes intention; for AI research tools it is more plausibly inferred from design choices that serve the user's goals than from any attributed motive (McKnight et al., 2011). Integrity may be particularly salient, since the main risks researchers associate with generative tools, fabricated references and unsupported claims (Walters & Wilder, 2023), are failures of honesty rather than of capability. Where a single fabricated citation can carry professional consequences, truthful and accurate reporting may matter more than either competence or apparent goodwill. We therefore expect each dimension to relate positively to trust, while treating their relative strength in this context as an empirical question.

Empirical work on AI systems supports a positive association between perceived ability and trust. In experimental research with a fictional AI chatbot, Lalot and Bertram (2025) found that competence was one of the two attributes, together with integrity, that most strongly led people to regard the chatbot as trustworthy. Sarker and Rashid (2026) reported a comparable result for generative AI, where perceived competence significantly predicted users' trust. For AI-powered research tools, ability corresponds to the accuracy and relevance of what the tool retrieves, summarises and analyses, and researchers who judge a tool to perform these tasks well should be more willing to rely on it. Hence:

**H1:** *Perceived ability positively impacts trust in AI-powered research tools.*

Evidence on benevolence is more recent but consistent in direction. Across four pre-registered experiments and a supplemental study (N = 2,100), Li and Bitterly (2024) showed that benevolence is an important and previously neglected predictor of trust in AI management, with the perception that AI lacks benevolence lowering trust. Liu et al. (2025) varied the behaviour of an AI assistant in a video game and found that an assistant acting in the user's interest was trusted significantly more, and elicited greater willingness to cooperate, than one that did not. Lalot and Bertram (2025) add a qualification: benevolence mattered for trust in an AI chatbot, but less than competence and integrity. Although a research tool cannot hold intentions, researchers may still infer goodwill from whether the tool is oriented toward their research goals. Hence:

**H2:** *Perceived benevolence positively impacts trust in AI-powered research tools.*

Integrity has emerged as a particularly strong basis for trust in conversational AI. Lalot and Bertram (2025) found integrity to be a more important determinant of trust in an AI chatbot than benevolence and, alongside competence, one of the main criteria by which people judged the chatbot to be reliable. Sarker and Rashid (2026) likewise reported that perceived integrity significantly influenced trust in generative AI. In research use, integrity is reflected in whether a tool reports information honestly, including the limits of what it knows, rather than presenting fabricated references or unsupported claims as fact (Walters & Wilder, 2023). Hence:

**H3:** *Perceived integrity positively impacts trust in AI-powered research tools.*

### 2.3. Trust in AI

Trust in AI plays an important role in how users interact with it and whether they are willing to adopt new technologies. Researchers often describe trust in AI as a mix of beliefs, expectations, and judgments, not just about how the system works, but also about how it fits into human relationships (Glikson & Woolley, 2020). From a user's point of view, trust depends on whether AI feels transparent, fair, and easy to understand, and whether its creators and organizations act responsibly (Doshi-Velez & Kim, 2017; Scharowski et al., 2023). When AI systems are clear, fair, and aligned with human values, people are more likely to trust them. That means trust isn't only about whether the technology works reliably; it also involves ethics, accountability, and good governance (Schmidt et al., 2020).

Empirical research consistently reported that trust is one of the crucial factors in whether people decide to adopt AI, often studied through models like the Technology Acceptance Model (TAM) (e.g., Abulail et al., 2025; Bedué & Fritzsche, 2022; Choung et al., 2023). Most of the focus so far has been on how trust shapes people's willingness to try AI in the first place. But there's been much less attention on what happens after that, whether trust influences people's decision to keep using AI over time (Shen et al., 2025). Some studies hint that trust is not static; it can grow stronger or fade as people continue interacting with AI (Hoff & Bashir, 2015). However, in the context of academic and research activities, limited empirical studies are conducted to understand how directly trust affects long-term, sustained use of AI-powered research tools. Current studies mostly focused on investigating usage intention or actual usage of AI-assisted tools (e.g., Anani et al., 2025; Prasad et al., 2026; Shahzad et al., 2025; Singh & Strzelecki, 2026; Tang et al., 2024; Zou & Huang, 2023), while very few and limited studies empirically studied continuance usage of such tools (Jung & Jo, 2025).

Because AI research tools often come with uncertainty, perceived risks, and ethical questions, trust becomes even more important for keeping people engaged over time (i.e., continued usage). Users are more likely to stick with these tools when they see them as reliable, transparent, and aligned with their own values and expectations. Building on earlier studies that focus mainly on first-time adoption, this work argues that trust is also a key driver of whether people continue using AI in the long run. Hence, drawing from existing studies on trust in AI and continuance intention across various contexts (e.g., Kang et al., 2024; Salih et al., 2025; Zhou & Ma, 2025), we can posit that:

**H4:** *Trust in AI positively impacts the continuance intention to use AI-powered research tools.*

### 2.4. Theoretical Foundations of TFS

Three theoretical perspectives explain why TFS should be related to trust and why it should condition the effect of trust on continued use. Framing theory (Entman, 1993; Goffman, 1974) holds that the selection and emphasis given to particular aspects of a message shape how it is interpreted, even when its factual content is unchanged. Applied to AI research tools, this implies that information about data sources, reasoning steps and limitations will be read differently depending on whether it is presented in a structured, specific and candid manner or in a vague and defensive one.

Signalling theory (Connelly et al., 2011) explains why this interpretation matters for trust. Researchers cannot observe the internal operation of a large language model, so they rely on observable cues to infer whether it is trustworthy, and the way a tool communicates about its own processes is one of the few such cues available at the point of use. A clear and specific account of how an output was produced is therefore likely to be read as a signal of competence and honesty.

Cognitive fit theory (Vessey, 1991) explains the moderating role. Trust establishes a willingness to rely on a tool, but whether that willingness becomes a commitment to continued use depends on whether the tool's explanations can be absorbed into the researcher's task without additional verification effort. When the presentation of transparency information matches the way researchers evaluate evidence, the cost of acting on trust falls and trust should translate more readily into continuance intention. When it does not, residual uncertainty remains and the same level of trust is expected to carry less weight.

## 2.5. Transparency Framing Style (TFS)

Transparency has become one of the key enablers of user trust in AI-driven systems (High-Level Expert Group on Artificial Intelligence, 2019). As AI increasingly mediates research processes, users encounter opacity from the design to the evaluation and presentation of the outputs (Burrell, 2016; Pasquale, 2015). Previous research consistently finds that transparency; defined as the level of system openness regarding the information available related to procedures, data sources, and decision logic; impacts the user's perception of the system's trustworthiness and trust (e.g., Bhegade, 2026; Durán & Pozzi, 2025; Glassberg et al., 2025). With information-based research tools powered by AI, this concept of transparency is even more critical, since users depend on these tools for their knowledge generation, synthesis, and decision support. A lack of transparency often leads to uncertainty, skepticism, and diminished trust, which ultimately leads to decreased willingness by users to use such tools (Glassberg et al., 2025).

Within AI, transparency is most commonly seen as a static or binary system property (present or absent), and attention is typically lacking on how transparency-related information is communicated and given meaning (Burrell, 2016; Doshi-Velez & Kim, 2017). Previous research mostly considers information quantity or availability rather than presentation (Burrell, 2016; Shen et al., 2026). In the context of AI transparency, few studies have investigated how different presentations of transparency cues might influence user perceptions and interpretations. Users do not merely receive transparency information; they interpret it.

The insights from framing theory, which was introduced by Goffman (1974) and later elaborated on by Entman (1993), are enlightening for them. Framing theory suggests that the kind of information that is presented or "framed" can greatly influence the way people perceive, judge, and make decisions (Entman, 1993). It is not content itself that is transformed, rather different frames are created that emphasize certain features of information that are perceived in a certain manner when they are interpreted or processed (Entman, 1993). Framing guides how a message is understood and interpreted, and also affects attitude, risk perception, and trust as part of the presentation context (including technological and media communication contexts) (Lindgren et al., 2024; Wiedemann et al., 2022).

Similarly, if framing is employed in AI transparency, it proves that, depending on how this transparency information is framed (possibly against the system capabilities, limitations, uncertainty, and ethical safeguards it covers), one kind may give rise to different trust (Haresamudram et al., 2023; Hoffman et al., 2018; Lee & See, 2004). According to this view, this research presents a novel construct, which is the Transparency Framing Style (TFS). TFS is how information regarding transparency about an AI system, including system processes, limitations, and decision logic, is presented to users through structure, emphasis, and tone. Unlike standard transparency framing that guide users, the construct incorporates an interpretive layer of interpretation, which is utilized by users to gain clarity of transparency cues. TFS is a crucial issue to the establishment of user trust in the openness and degree of honesty of the system.

Based on early studies in information systems (IS) and human–AI interaction, this paper extends the hypothesis that TFS dictates user trust levels in AI systems, and whether continuance intention is influenced by TFS. Transparency is believed to be an important antecedent of trust in intelligent systems since it attenuates informational asymmetry and helps the consumer to understand the behaviour of the system (e.g., Durán & Pozzi, 2025; Fisher & Hopp, 2020; Glassberg et al., 2025; Sagona et al., 2025; Tang et al., 2024). In AI systems where users find out about data sources, decision-making logic, or limitations about their systems, the tendency towards users treating them as predictable and responsible

is higher and thereby perceiving the trustworthiness and trust (Lee & See, 2004). Hence, we argue that TFS, the manner in which transparency information is organised, emphasised, and conveyed, is positively associated with user trust in AI systems by affecting the way in which users view and interpret transparency stimuli.

Apart from the overt effect on trust, we also demonstrate that TFS moderates the relationship between trust and continuance intention. Most IS studies have shown that trust plays an important role in determining the subsequent use of IS, particularly when the IS environment is uncertain/automated (Gefen et al., 2003; Pavlou & Gefen, 2004). In contrast, recent studies show that the trust–behavior relationship is strongly correlated to contextual and perceptual features, particularly system-level transparency and explainability (Hoff & Bashir, 2015). Indeed, TFS is expected to help facilitate the translation of trust into continuance intention by enhancing the interpretability and credibility of transparency information. If transparency has a clear, meaningful framing, users are more likely to rely on their trust beliefs when making continuance intentions; conversely, weak framing of transparency will create residual uncertainty and a negative potential to strengthen the relationship. In this way, TFS not only directly fosters trust, but it also conditions trust’s relation to ongoing use of AI-related systems. Hence, the following hypotheses are posited:

**H5:** TFS positively impacts trust in AI-powered research tools.

**H6:** TFS positively moderates the relationship between trust and continuance intention to use AI-powered research tools.

## 2.6. Research Model

According to the theoretical underpinnings discussed and the hypotheses developed, the proposed research model is illustrated in Figure 1.

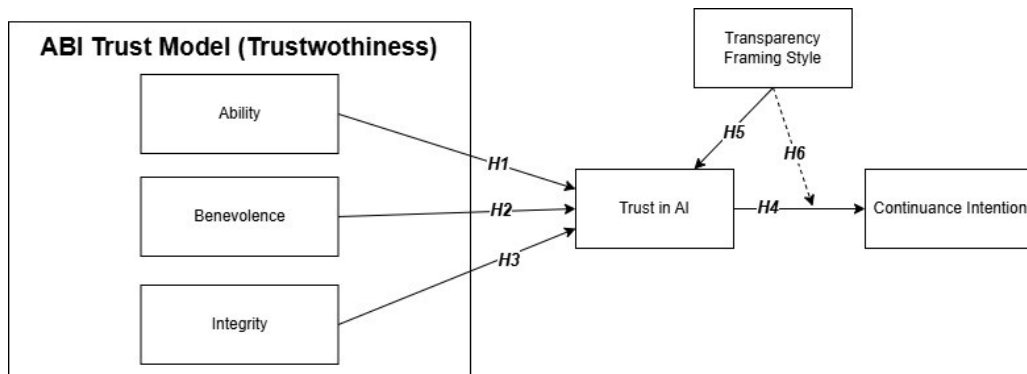


Figure 1. Research Model

## 3. Research Methodology

### 3.1. Measurement Items

Table 2 measurement items for all constructs were adapted from prior validated scales to ensure content validity. Accordingly, the items measuring ability, benevolence, and integrity were developed by Huynh and Aichner (2025). The trust construct was measured using the items developed by Chen and Cheng (2026), while the continuance intention construct was measured using the items developed by Bhattacharjee (2001). The TFS construct was developed by the researcher and applied in accordance with the context of the present study. All measurement items were adapted to the context of the current study, which focuses on the application of AI-powered research tools by academics and researchers. All constructs were measured using a 7-point Likert scale, where 1 indicated ‘strongly disagree’ and 7 indicated ‘strongly agree’. Table 2 presents the summary of measurement items.

The TFS items were developed in three stages. First, items were generated from the construct definition, drawing on framing theory (Entman, 1993) and on prior work on the framing of transparency messages (Fisher & Hopp, 2020). Second, the draft items were reviewed by a panel of five experts, three with expertise in human-AI interaction or information systems and two in research methodology. For face validity, the experts assessed the clarity, wording and readability of each item. For content validity, they rated the relevance of each item to the construct definition on a four-point scale, and the item-level content validity index (I-CVI) was computed. Third, the full questionnaire was pilot tested with 45 respondents drawn from the target population, exceeding the minimum of 30 recommended for pilot studies in scale development (Johanson & Brooks, 2010). The pilot data showed satisfactory reliability for TFS. Pilot respondents were excluded from the main sample.

**Table 2. Measurement items.**

| Construct                               | Items                                                                                                                         | Source                   |
|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Ability (ABLTY)</b>                  | AI-powered research tools interact with me in ways that are effective and helpful for literature review and research writing. | Huynh and Aichner (2025) |
|                                         | AI-powered research tools perform their research support tasks very well.                                                     |                          |
|                                         | AI-powered research tools handle research-related tasks skilfully and accurately.                                             |                          |
|                                         | AI-powered research tools provide useful and relevant information for my research work.                                       |                          |
| <b>Benevolence (BNVLN)</b>              | AI-powered research tools consider my needs when providing help with research tasks.                                          |                          |
|                                         | When I ask for assistance, AI-powered research tools try their best to support me.                                            |                          |
|                                         | AI-powered research tools seem interested in helping me make progress in my research.                                         |                          |
| <b>Integrity (INTGRT)</b>               | AI-powered research tools provide information that is accurate and reliable.                                                  |                          |
|                                         | AI-powered research tools present ideas and suggestions honestly.                                                             |                          |
|                                         | AI-powered research tools communicate in a straightforward and sincere manner.                                                | Chen and Cheng (2026)    |
| <b>Trust (TRUST)</b>                    | I feel confident in the suggestions provided by AI-powered research tools.                                                    |                          |
|                                         | I believe AI-powered research tools provide reliable guidance for literature review and research writing.                     |                          |
|                                         | I feel comfortable depending on AI-powered research tools to support my research work.                                        |                          |
|                                         | I trust the outputs of AI-powered research tools even when I cannot fully verify them myself.                                 |                          |
| <b>Continuance Intention (CI)</b>       | I intend to continue using AI-powered research tools in the future.                                                           | Bhattacharjee (2001)     |
|                                         | I plan to continue using AI-powered research tools for my research tasks.                                                     |                          |
|                                         | I will frequently use AI-powered research tools in my academic work.                                                          |                          |
| <b>Transparency Framing Style (TFS)</b> | The way information is presented by the AI tool is part of how I interpret its outputs.                                       | Researcher developed     |
|                                         | The presentation style used by the AI tool is part of how I understand its responses.                                         |                          |
|                                         | The manner in which the AI tool communicates information is related to how I perceive its recommendations.                    |                          |
|                                         | The same information can be understood differently depending on how it is presented by the AI tool.                           |                          |

### 3.2. Sample and Data Collection

The ten universities, three public and seven private, were selected to include both sectors of Malaysian higher education and to cover institutions with active postgraduate research programmes across a range of disciplines. Institutions were identified through the authors' professional academic networks, which provided access to institutional mailing lists and research groups. The survey link was circulated over a four-week period in January 2026 through these mailing lists and through professional academic networks, and a reminder was sent to the same channels two weeks after the initial invitation.

Because the survey link was distributed through institutional mailing lists and professional academic networks, the total number of individuals who received the invitation could not be determined, and a conventional response rate (completed surveys as a proportion of invitations) could not therefore be calculated. Of the 412 responses logged, 43 were excluded at the screening stage because the respondent had not used AI-powered research tools in the previous six months, leaving 369 eligible responses. A further 95 were removed because more than 10% of items were unanswered or the response showed a straight-lining pattern, yielding 274 usable cases. This represents a usable response ratio of 66.5% of all logged responses and 74.3% of eligible responses. Straight-lining was identified as the selection of the same response option across all items of the questionnaire, following Hair et al. (2022). For the retained cases, the remaining missing values were replaced using mean replacement in SmartPLS, as recommended by Hair et al. (2022) when the proportion of missing data is low.

To assess non-response bias, the responses of the first 25% and the last 25% of respondents were compared on all study constructs, on the assumption that late respondents resemble non-respondents (Armstrong & Overton, 1977). Independent-samples t-tests showed no significant differences between the two groups ( $p > 0.05$ ), suggesting that non-response bias is not a serious concern in this study. The minimum sample size was estimated using the inverse square root and gamma-exponential methods (Kock & Hadaya, 2018). Assuming a minimum absolute path coefficient of 0.197 at a 5% significance level and 80% statistical power, the inverse square root method requires 160 observations and the gamma-exponential method requires 146. A power analysis conducted in G\*Power (Faul et al., 2007) for the model's most complex regression, in which trust is predicted by four constructs, indicated a minimum of 85 cases ( $f^2 = 0.15$ ,  $\alpha = 0.05$ , power = 0.80). The final sample of 274 exceeded all of these thresholds.

### 3.3. Common Bias Method

To mitigate common method bias (CMB), procedural remedies were applied during questionnaire design and data collection following Podsakoff et al. (2003). First, items were adapted from validated studies and worded clearly to reduce ambiguity. Second, questionnaire items were randomised to limit order effects and reduce respondents' ability to infer the proposed relationships. Third, the cover letter assured respondents that their responses were anonymous, confidential and used only for research purposes, thereby reducing evaluation apprehension and socially desirable responding. These procedures were intended to reduce the likelihood that the observed relationships were driven by CMB rather than the underlying theoretical relationships.

### 3.4. Data Analysis

Partial least squares structural equation modelling (PLS-SEM) was conducted using SmartPLS to analyse the proposed research model. PLS-SEM was selected because it accommodates complex models containing direct, mediating and moderating relationships and places fewer distributional demands on the data. It also permits the measurement and structural models to be assessed within a single analytical framework.

We followed the procedure recommended by Hair et al. (2022) to test measurement and structural models. The measurement model was evaluated for internal reliability using outer loadings, with values exceeding 0.70; for internal consistency reliability using composite reliability, with values above 0.70; for convergent validity using the average variance extracted (AVE), with values greater than 0.50; and for discriminant validity using the Heterotrait-Monotrait Ratio (HTMT), with values below 0.85.

To assess the structural model, the path coefficients were tested for significance using the bootstrapping method with 5,000 resamples; the coefficient of determination ( $R^2$ ) was used to examine the variance explained in the dependent variables; effect size ( $f^2$ ) was calculated to determine the contribution of each independent variable to  $R^2$ ; the predictive relevance of the model was evaluated using the blindfolding procedure and  $Q^2$  values; and collinearity among the independent variables was assessed using the variance inflation factor (VIF).

## 4. Results

### 4.1. Demography

Table 3 presents the descriptive metadata for the final dataset ( $N = 274$ ). The sample shows a reasonably balanced gender distribution, with 54.0% female ( $n = 148$ ) and 46.0% male ( $n = 126$ ) respondents. Age-wise, the sample skews toward early-to-mid career profiles, with 63.9% of participants falling in the 24 to 33 age bracket ( $n = 175$ ), followed by the 34 to 43 cohort at 16.8% ( $n = 46$ ). In terms of institutional role, postgraduate research students at the Master's and PhD levels make up the majority at 67.9% ( $n = 186$ ), while formal faculty members across lecturer, assistant professor, associate professor, and full professor ranks account for the remaining 32.1%. Consistent with this distribution, 50.7% of the sample hold a Master's degree as their highest qualification ( $n = 139$ ), 22.3% hold a Doctorate ( $n = 61$ ), and 9.9% are at the postdoctoral level ( $n = 27$ ). Disciplinary representation spans multiple domains, with the largest concentrations in Natural Sciences (31.0%,  $n = 85$ ) and Engineering and Technology (25.5%,  $n = 70$ ), alongside meaningful representation from the Social Sciences (16.4%) and Humanities (12.8%).

**Table 3. Respondent Demographic Characteristics and AI Usage Profile ( $N = 274$ )**

| Variable              | Classification                 | n (%)       | Variable            | Classification                     | n (%)       |
|-----------------------|--------------------------------|-------------|---------------------|------------------------------------|-------------|
| Gender                | Male                           | 126 (46.0)  | Academic Experience | Less than 1 year                   | 110 (40.1)  |
|                       | Female                         | 148 (54.0)  |                     | 1–3 years                          | 85 (31.0)   |
|                       | Total                          | 274 (100.0) |                     | 4–7 years                          | 45 (16.4)   |
| Age Distribution      | 24–33 years                    | 175 (63.9)  |                     | 8–15 years                         | 25 (9.1)    |
|                       | 34–43 years                    | 46 (16.8)   |                     | More than 15 years                 | 9 (3.4)     |
|                       | 44–53 years                    | 27 (9.9)    |                     | Total                              | 274 (100.0) |
|                       | 54–63 years                    | 19 (6.9)    | Research Discipline | Natural Sciences                   | 85 (31.0)   |
|                       | 64–73 years                    | 5 (1.8)     |                     | Engineering & Technology           | 70 (25.5)   |
|                       | Over 73 years                  | 2 (0.7)     |                     | Social Sciences                    | 45 (16.4)   |
|                       | Total                          | 274 (100.0) |                     | Humanities                         | 35 (12.8)   |
| Highest Qualification | Bachelor's degree              | 47 (17.2)   |                     | Medical & Health Sciences          | 25 (9.1)    |
|                       | Master's degree                | 139 (50.7)  |                     | Other Domains                      | 14 (5.1)    |
|                       | Doctorate (PhD)                | 61 (22.3)   |                     | Total                              | 274 (100.0) |
|                       | Postdoctoral / Other           | 27 (9.9)    | AI Usage Frequency  | Rarely (< once/month)              | 5 (1.8)     |
|                       | Total                          | 274 (100.0) |                     | Occasionally (1–3 times/month)     | 16 (5.8)    |
| Academic Position     | Lecturer / Assistant Professor | 45 (16.4)   |                     | Frequently (1–3 times/week)        | 88 (32.1)   |
|                       | Associate Professor            | 18 (6.6)    |                     | Very frequently (daily/near-daily) | 165 (60.2)  |

| Variable                                                                                                                                                                                                                                                                                           | Classification                   | n (%)              | Variable                          | Classification                  | n (%)              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|--------------------|-----------------------------------|---------------------------------|--------------------|
|                                                                                                                                                                                                                                                                                                    | Professor                        | 3 (1.1)            |                                   | <b>Total</b>                    | <b>274 (100.0)</b> |
|                                                                                                                                                                                                                                                                                                    | Researcher / Postdoctoral Fellow | 22 (8.0)           | <b>Primary Purpose of AI Use*</b> | Literature review               | 185 (67.5)         |
|                                                                                                                                                                                                                                                                                                    | Student (Master's / PhD)         | 186 (67.9)         |                                   | Research writing / drafting     | 210 (76.6)         |
|                                                                                                                                                                                                                                                                                                    | <b>Total</b>                     | <b>274 (100.0)</b> |                                   | Data analysis / interpretation  | 120 (43.8)         |
|                                                                                                                                                                                                                                                                                                    |                                  |                    |                                   | Reference management / citation | 70 (25.5)          |
|                                                                                                                                                                                                                                                                                                    |                                  |                    |                                   | Other activities                | 15 (5.5)           |
|                                                                                                                                                                                                                                                                                                    |                                  |                    |                                   |                                 |                    |
| <i>Note.</i> Values are reported as frequency, <i>n</i> (percentage). *Primary purpose of AI use was a multiple-response item; therefore, respondents could select more than one option and percentages do not sum to 100%. Minor discrepancies from 100% in other categories are due to rounding. |                                  |                    |                                   |                                 |                    |

Table 3 usage patterns further reveal that generative AI tools are deeply embedded in participants' daily workflows, with a combined 92.3% reporting either frequent (32.1%, *n* = 88) or very frequent (60.2%, *n* = 165) use. These tools are applied primarily to complex academic tasks, most commonly research writing and drafting (76.6%, *n* = 210) and systematic literature reviews (67.5%, *n* = 185). Together, these figures describe a highly trained, digitally engaged academic sample that treats AI platforms as core cognitive instruments rather than peripheral administrative conveniences.

#### 4.2. Measurement Model Assessment

The measurement model demonstrates satisfactory reliability and validity across the constructs. As shown in Table 4, all item loadings exceed 0.70, indicating adequate indicator reliability. Composite reliability values range from 0.847 to 0.878, while AVE values range from 0.716 to 0.782 and exceed the 0.50 threshold, supporting convergent validity.

**Table 4. Reliability and Validity Indicators**

| Construct                         | Items   | Outer loading | Composite reliability | AVE   |
|-----------------------------------|---------|---------------|-----------------------|-------|
| <b>Ability</b>                    | ABLT1   | 0.872         | 0.876                 | 0.725 |
|                                   | ABLT2   | 0.851         |                       |       |
|                                   | ABLT3   | 0.851         |                       |       |
|                                   | ABLT4   | 0.831         |                       |       |
| <b>Benevolence</b>                | BNVL1   | 0.870         | 0.847                 | 0.763 |
|                                   | BNVL2   | 0.869         |                       |       |
|                                   | BNVL3   | 0.882         |                       |       |
| <b>Continuance Intention</b>      | CI1     | 0.886         | 0.848                 | 0.761 |
|                                   | CI2     | 0.848         |                       |       |
|                                   | CI3     | 0.881         |                       |       |
| <b>Integrity</b>                  | INTGRT1 | 0.883         | 0.862                 | 0.782 |
|                                   | INTGRT2 | 0.890         |                       |       |
|                                   | INTGRT3 | 0.880         |                       |       |
| <b>Transparency Framing Style</b> | TFS1    | 0.857         | 0.878                 | 0.732 |
|                                   | TFS2    | 0.872         |                       |       |
|                                   | TFS3    | 0.854         |                       |       |
|                                   | TFS4    | 0.838         |                       |       |
| <b>Trust in AI</b>                | TRUST1  | 0.846         | 0.869                 | 0.716 |
|                                   | TRUST2  | 0.836         |                       |       |
|                                   | TRUST3  | 0.852         |                       |       |
|                                   | TRUST4  | 0.851         |                       |       |

Table 5 also indicates acceptable discriminant validity under the HTMT criterion, with all ratios below the conservative threshold of 0.85. The measurement model was therefore considered suitable for structural model assessment.

**Table 5. Discriminant Validity Assessment using HTMT**

|                            | ABLTY | BNVLN | CI    | INTGRT | TFS |
|----------------------------|-------|-------|-------|--------|-----|
| Ability                    |       |       |       |        |     |
| Benevolence                | 0.609 |       |       |        |     |
| Continuance Intention      | 0.301 | 0.193 |       |        |     |
| Integrity                  | 0.574 | 0.672 | 0.296 |        |     |
| Transparency Framing Style | 0.475 | 0.474 | 0.476 | 0.522  |     |

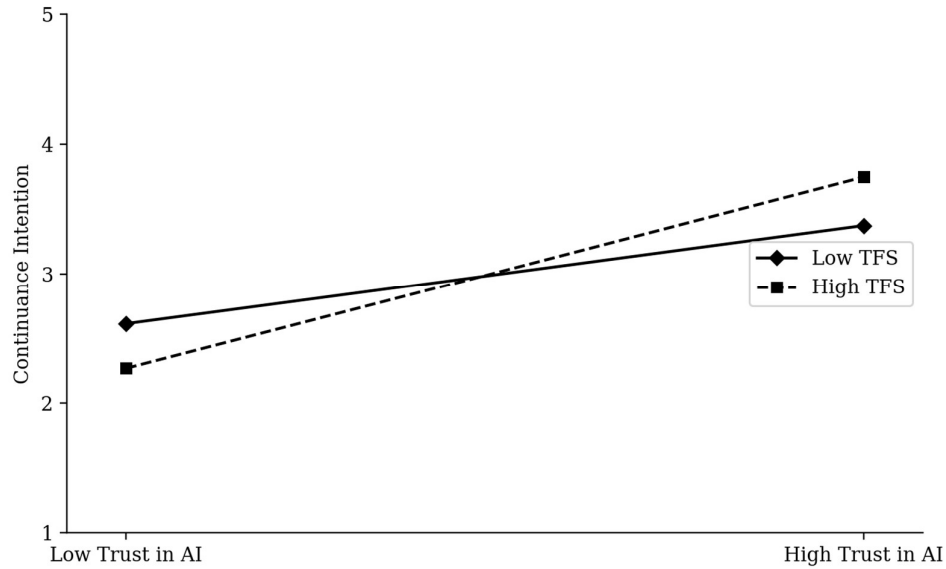


Figure 2. Simple Slope Analysis

#### 4.3. Structural Model Assessment

The structural model results in Table 6 show that ability has a positive and significant effect on trust in AI ( $\beta = 0.165$ ,  $t = 3.55$ ,  $p < 0.001$ ), while benevolence has a smaller but significant effect ( $\beta = 0.081$ ,  $t = 1.671$ ,  $p = 0.047$ ). Integrity also has a positive and significant effect on trust ( $\beta = 0.251$ ,  $t = 4.496$ ,  $p < 0.001$ ). TFS has a strong positive effect on trust in AI ( $\beta = 0.531$ ,  $t = 7.771$ ,  $p < 0.001$ ), and trust in AI positively predicts continuance intention ( $\beta = 0.561$ ,  $t = 7.585$ ,  $p < 0.001$ ). The interaction between TFS and trust is also significant ( $\beta = 0.180$ ,  $t = 4.587$ ,  $p < 0.001$ ), consistent with the simple-slope pattern in Figure 2. The model explains 68.6% of the variance in trust and 39.8% of the variance in continuance intention. Predictive relevance is high for trust ( $Q^2 = 0.669$ ) and weak to moderate for continuance intention ( $Q^2 = 0.151$ ). Overall, the results support all six hypotheses.

#### 5. Discussion

This study examined the socio-technical mechanisms shaping user trust and continuance intention toward AI-powered research tools. Viewing the findings through classical trust theory, it becomes clear that researchers evaluate algorithmic systems through a relational lens not unlike the one they apply to interpersonal relationships. Mayer et al. (1995) anchored trustworthiness in three dimensions: ability, benevolence, and integrity. The present findings confirm that these human-centered attributes translate systematically to digital and algorithmic trustees (Glikson & Woolley, 2020). Users do not simply ask whether an AI system is capable of synthesizing data; they also consider whether it operates in alignment with their objectives and adheres to a consistent, ethically grounded logic. In effect, researchers engage

in a continuous, largely implicit evaluation of whether these systems qualify as viable collaborative partners.

Integrity showed a larger estimated association with trust than benevolence. In the high-stakes domain of academic research, this pattern suggests that scholars may place greater weight on methodological consistency and operational reliability than on perceived relational warmth. This interpretation is consistent with McKnight et al.'s (2011) observation that behavioural reliability and structural integrity become especially important under conditions of heightened risk.

**Table 6. Structural Model Results**

| Relationship                                        | Path (O) | Mean (M) | STDEV | t-value | p-value | Effect ( $f^2$ / $Q^2$ )                                                                                     |
|-----------------------------------------------------|----------|----------|-------|---------|---------|--------------------------------------------------------------------------------------------------------------|
| Ability → Trust in AI                               | 0.165    | 0.165    | 0.046 | 3.55    | <0.001  | $f^2$ = 0.056 (small)                                                                                        |
| Benevolence → Trust in AI                           | 0.081    | 0.078    | 0.048 | 1.671   | 0.047   | $f^2$ = 0.012 (negligible)                                                                                   |
| Integrity → Trust in AI                             | 0.251    | 0.248    | 0.056 | 4.496   | <0.001  | $f^2$ = 0.117 (small–moderate)                                                                               |
| TFS → Trust in AI                                   | 0.531    | 0.536    | 0.068 | 7.771   | <0.001  | $f^2$ = 0.658 (large)                                                                                        |
| TFS × Trust in AI → Continuance Intention           | 0.18     | 0.18     | 0.039 | 4.587   | <0.001  | $f^2$ = 0.080 (small–moderate interaction)                                                                   |
| Trust in AI → Continuance Intention                 | 0.561    | 0.562    | 0.074 | 7.585   | <0.001  | $f^2$ = 0.220 (moderate)                                                                                     |
| <b>Trust in AI (Predictive relevance)</b>           |          |          |       |         |         | $R^2$ = 0.686 (strong)<br>$Q^2$ = 0.669 (high predictive relevance);<br>RMSE = 0.580; MAE = 0.423            |
| <b>Continuance Intention (Predictive relevance)</b> |          |          |       |         |         | $R^2$ = 0.398 (moderate)<br>$Q^2$ = 0.151 (weak–moderate predictive relevance);<br>RMSE = 0.930; MAE = 0.736 |

Signaling Theory (Connelly et al., 2011) offers a useful lens for interpreting this dynamic. Because users cannot directly observe the internal mechanics of generative AI models, they rely on visible operational signals to gauge epistemic risk (Doshi-Velez & Kim, 2017). When a tool consistently delivers valid, verifiable outputs and demonstrates disciplinary competence, those signals function as uncertainty-reducing mechanisms, consistent with foundational technology acceptance models that tie sustained adoption to demonstrated utility and performance consistency (Davis, 1989; Venkatesh et al., 2003). The largest estimated direct coefficient for trust was associated with Transparency Framing Style (TFS), rather than with the ABI dimensions alone. This pattern suggests that the way an AI system communicates its underlying logic is an important component of trust calibration alongside the quality of its outputs (Miller, 2019).

This reflects the Framing Effect (Tversky & Kahneman, 1981), whereby the contextual presentation of information fundamentally shapes how users process and evaluate it. Given that most advanced AI models operate as opaque black boxes, clear and user-centered explanations serve as essential cognitive heuristics (Doshi-Velez & Kim, 2017), signaling to researchers that outputs are credible and verifiable. This finding converges directly with contemporary Explainable AI (XAI) perspectives, which hold that algorithmic explanations are only useful when they align with human cognitive structures. Computational accuracy, in other words, must be paired with interpretable framing to genuinely build trust.

The robust path coefficient linking trust to continuance intention further validates the ECM in generative AI contexts (Bhattacharjee, 2001). Trust here is not a static judgment made at the point of initial adoption; it is a dynamic psychological state that is continuously recalibrated through iterative experience (Bhatnagar et al., 2024). Because academic inquiry is inherently iterative, researchers must repeatedly rely on algorithmic outputs to inform consequential intellectual decisions. When trust is not reinforced through consistent interaction cycles, the tool shifts from an academic asset to an operational liability, ultimately

driving discontinuance. For an AI system to maintain viability in scholarly workflows, it must re-verify its reliability across successive use cycles rather than relying on early positive impressions.

The empirical confirmation of TFS as a significant moderator of the trust-continuance relationship represents a central theoretical contribution of this study. A researcher may trust an AI tool in the abstract while still hesitating to commit to it as a permanent workflow component, and that gap matters. Cognitive Fit Theory (Vessey, 1991) helps explain this distinction: when a system's explanation structure matches the user's cognitive processing style, it produces a state of internal alignment that transforms passive trust into active behavioral reliance. Rather than approaching outputs with residual skepticism, the researcher understands the system's epistemic path and integrates it confidently into their core methodology. Simple slope analysis showed that the relationship between trust and continuance intention was stronger at high levels of TFS (+1 SD) than at low levels (-1 SD). This supports emerging perspectives on human-AI complementarity (Brynjolfsson et al., 2017; Jussupow et al., 2021), suggesting that an AI tool's value depends not only on its computational power but on how seamlessly it integrates with human cognitive and decision-making processes.

Taken together, these findings position trust in academic AI tools as a multifaceted construct shaped by the interplay between objective system integrity and communicative transparency framing. Transparency serves a dual function in this ecosystem: it anchors initial trust and sustains continuance over time. This reframes technology adoption away from static linear models and toward a more fluid, recursive co-adaptation between technical capability and human cognitive need.

## 6. Implications

The findings suggest a set of propositions for developers and institutions. These propositions follow from the relationships observed in this study but have not themselves been tested, and they should be read as a basis for future design research and evaluation rather than as validated guidance.

### 6.1. Theoretical Implications

This study contributes to the literature on trust in AI-enabled environments in several distinct ways. First, by applying Mayer et al.'s (1995) framework to generative AI contexts, the study confirms that ability, benevolence, and integrity remain valid trustworthiness dimensions even when the trustee is a non-human system, extending classical trust theory into algorithmic territory.

Second, it advances Signaling Theory by demonstrating that under conditions of algorithmic opacity, users rely heavily on observable trust cues embedded in system explanations. This positions transparency as a cognitive signaling mechanism that reduces uncertainty and calibrates trust, rather than treating it as a simple interface feature (Connelly et al., 2011).

Third, the study expands the ECM (Bhattacharjee, 2001) by reconceptualizing trust as a dynamic, recursive mechanism that sustains continuance intention through continuous reinforcement across iterative tasks, rather than a static post-adoption outcome measured at a single point in time.

Fourth, and most importantly, it operationalises TFS as a boundary condition by integrating Cognitive Fit Theory (Vessey, 1991) into the human-AI collaboration context. The findings demonstrate that how transparency is framed determines whether trust actually converts into behavioral reliance, underscoring the necessity of cognitive alignment between explanation design and users' mental models.

### 6.2. Practical Implications

To translate these empirical insights into system design and institutional policy, this study proposes a comprehensive, five-stage framework for managing trust-driven continuance in AI-powered research environments.

#### *Stage 1: The Trust Anchor (Development Layer)*

System developers must prioritize establishing algorithmic integrity as a fundamental prerequisite for user engagement. Because researchers face significant epistemic risks, software engineers must design

generative tools to provide explicit verification pathways. Instead of rendering unverified text outputs, the system interface must structurally present source data provenance, explicit citation pathways, and reasoning logs. Furthermore, integrating confidence metrics that signal when evidence is sparse allows users to calibrate their reliance accurately, removing the ambiguity that leads to algorithmic aversion.

*Stage 2: The Cognitive Alignment (Interface Layer)*

Once technical integrity is established, system architectures must optimize for Cognitive Fit to prevent user cognitive overload. AI developers should implement layered transparency framing mechanisms. This interface design provides an initial high-level summary of the logic, followed by expandable reasoning steps, and finally, direct access to the raw underlying evidence. Crucially, these explanation interfaces should be adaptive, tailoring the granularity of the contextual data to the user's domain expertise, thereby transforming complex algorithmic outputs into intuitive decision-making tools.

*Stage 3: Workflow Integration (Usage Layer)*

To move user interaction from experimental utilization to permanent adoption, the AI utility must transition into a seamless workflow component. Proper transparency framing provides researchers with justified confidence, reducing the need for exhaustive secondary verification of every output. By making the tool's logical constraints clear, researchers can safely delegate specific information-processing tasks, embedding the system as a reliable decision-support mechanism within their core research routines.

*Stage 4: Institutional Governance (Governance Layer)*

To ensure long-term stability within the broader academic ecosystem, higher education institutions, funding bodies, and publishers must establish formal frameworks for AI integration. Higher education administrators should implement critical algorithmic literacy programs that train scholars to interrogate AI explanation pathways rather than accepting outputs passively. Concurrently, academic publishers and institutional review boards must define clear reporting standards for AI-assisted inquiries, ensuring that tool adoption aligns with established research ethics and transparency guidelines.

*Stage 5: Feedback-Driven Trust Calibration (Ecosystem Learning Layer)*

Finally, a continuous feedback loop should be established across the academic ecosystem, where user interactions, trust perceptions, and usage patterns are systematically collected to refine both AI systems and institutional policies. Developers can utilize these empirical usage metrics to iteratively optimize transparency framing mechanisms. Simultaneously, institutional bodies can update governance and compliance standards based on evolving user needs and technological advancements, ensuring that human-AI collaboration remains robust, ethical, and sustainable.

### **6.3. Social Implications**

The findings also bear on the wider research community and on public confidence in science. First, they point to a risk as well as an opportunity. Trust in AI research tools was strongly associated with how transparency information was framed, which means that persuasive presentation could raise trust without any corresponding improvement in the reliability of outputs. If researchers rely on well-presented but unverified material, errors may pass into published work and, through it, into policy and public understanding. The appropriate goal is therefore calibrated rather than maximal trust, and transparency framing should be judged partly by whether it helps users recognise when not to rely on a tool.

Second, most respondents were early-career researchers and postgraduate students who use these tools daily. This group has the least accumulated experience against which to judge AI outputs, and unequal access to training in AI literacy may widen existing differences in research capability between better and less well-resourced institutions. Third, as AI-assisted research becomes common, public trust in scientific findings will depend increasingly on whether the use of these tools is disclosed and accountable. Shared conventions for reporting AI assistance, supported by publishers and funders, would help to protect the credibility of research in the eyes of the public.

## 7. Limitations and Future Studies

This study has several limitations that suggest directions for future research. First, the cross-sectional design does not support causal inference or reveal how trust and TFS change as users gain experience with AI-powered tools; longitudinal designs could address these dynamics. Second, the sample is limited to academics and researchers in Malaysian universities, which constrains generalisability to other countries and professional settings with different cultural and institutional approaches to AI and transparency. Third, TFS was examined as an aggregate construct. Future studies could separate dimensions such as linguistic tone, level of detail and visual presentation to determine which aspects of framing most effectively reduce epistemic risk and cognitive load. Finally, future research could examine how perceived AI personality or empathy interacts with the ABI dimensions as AI systems become more autonomous.

## 8. Conclusion

The findings show that trust in AI-powered research tools is shaped by perceived ability, benevolence and integrity. Integrity displayed a larger estimated association with trust than benevolence, while TFS was a significant direct predictor of trust and a significant moderator of the trust-continuance relationship. This indicates that how transparency information is framed can influence whether trust translates into continued use. For AI tools to be integrated effectively into scientific workflows, developers and institutions should therefore consider algorithmic reliability together with interpretive compatibility and communicative transparency that support sustained, appropriately calibrated human-AI collaboration.

The data were collected at a single point in time from academics and postgraduate researchers in Malaysian universities, and the sample was weighted towards early-career researchers who use AI tools frequently. The relationships reported here are therefore associations rather than causal effects, and their strength may differ in countries with other research cultures, regulatory approaches to AI or levels of institutional trust. Longitudinal and multi-country studies are needed to establish whether the role of TFS observed here holds over time and across settings.

## Declarations

### Funding

This research received no external funding.

### Conflict of Interest

The author declares no conflict of interest.

### Data Availability

The data supporting the findings of this study are available from the corresponding author upon request.

## References

- Abulail, R. N., Badran, O. N., Shkoukani, M. A., & Omeish, F. (2025). Exploring the factors influencing AI adoption intentions in higher education: An integrated model of DOI, TOE, and TAM. *Computers*, 14(6), Article 230. <https://doi.org/10.3390/computers14060230>
- Alharbi, B. S., Aljabri, M. M., & Ali, E. A. (2026). Determinants of trust in artificial intelligence (AI) for health-related decision-making among adults in Saudi Arabia: A cross-sectional study. *Healthcare*, 14(4), Article 506. <https://doi.org/10.3390/healthcare14040506>
- Alonso, M., Astobiza, A. M., & Ortega Lozano, R. (2025). AI-mediated healthcare and trust. A trust-construct and trust-factor framework for empirical research. *Artificial Intelligence Review*, 58(11), Article 337. <https://doi.org/10.1007/s10462-025-11306-7>

- Alotaibi, S. M. F. (2025). Determinants of generative artificial intelligence (GenAI) adoption among university students and its impact on academic performance: The mediating role of trust in technology. *Interactive Learning Environments*, 33(6), 4159–4188. <https://doi.org/10.1080/10494820.2025.2492785>
- Anani, G. E., Nyamekye, E., & Bafour-Koduah, D. (2025). Using artificial intelligence for academic writing in higher education: The perspectives of university students in Ghana. *Discover Education*, 4(1), Article 46. <https://doi.org/10.1007/s44217-025-00434-5>
- Andersen, J. P., Degn, L., Fishberg, R., Graversen, E. K., Horbach, S. P. J. M., Schmidt, E. K., Schneider, J. W., & Sørensen, M. P. (2025). Generative artificial intelligence (GenAI) in the research process: A survey of researchers' practices and perceptions. *Technology in Society*, 81, Article 102813. <https://doi.org/10.1016/j.techsoc.2025.102813>
- Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14(3), 396–402. <https://doi.org/10.1177/002224377701400320>
- Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2024). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction*, 40(5), 1251–1266. <https://doi.org/10.1080/10447318.2022.2138826>
- Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, 35(2), 530–549. <https://doi.org/10.1108/JEIM-06-2020-0233>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Bhatnagar, P., Rajesh, A., & Misra, R. (2024). Continuous intention usage of artificial intelligence enabled digital banks: A review of expectation confirmation model. *Journal of Enterprise Information Management*, 37(6), 1763–1787. <https://doi.org/10.1108/JEIM-11-2023-0617>
- Bhattacharjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 25(3), 351–370. <https://doi.org/10.2307/3250921>
- Bhegade, V. (2026). A viewpoint on trust, ethics, and transparency in AI-driven learning organizations. *Development and Learning in Organizations*, 40(2), 27–30. <https://doi.org/10.1108/DLO-08-2025-0311>
- Brynjolfsson, E., Rock, D., & Syverson, C. (2017). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics* (NBER Working Paper No. 24001). National Bureau of Economic Research. <https://doi.org/10.3386/w24001>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4* (arXiv:2303.12712). arXiv. <https://doi.org/10.48550/arXiv.2303.12712>
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951715622512>
- Chen, C.-J., & Cheng, J. M.-S. (2026). When help feels creepy: A MOA model of AI smart-retail adoption via perceived intrusiveness and the moderating role of trust in AI. *Journal of Retailing and Consumer Services*, 90, Article 104699. <https://doi.org/10.1016/j.jretconser.2025.104699>
- Chen, S. C., & Dhillon, G. S. (2003). Interpreting dimensions of consumer trust in e-commerce. *Information Technology and Management*, 4(2–3), 303–318. <https://doi.org/10.1023/A:1022962631249>
- Cheung, J. C., & Ho, S. S. (2025). The effectiveness of explainable AI on human factors in trust models. *Scientific Reports*, 15(1), Article 23337. <https://doi.org/10.1038/s41598-025-04189-9>
- Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, 39(9), 1727–1739. <https://doi.org/10.1080/10447318.2022.2050543>

- Cloarec, J., Meyer-Waarden, L., & Munzel, A. (2024). Transformative privacy calculus: Conceptualizing the personalization–privacy paradox on social media. *Psychology & Marketing*, 41(7), 1574–1596. <https://doi.org/10.1002/mar.21998>
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39–67. <https://doi.org/10.1177/0149206310388419>
- Dang, J., & Liu, L. (2025). Public trust in artificial intelligence users. *Current Opinion in Psychology*, 66, Article 102148. <https://doi.org/10.1016/j.copsyc.2025.102148>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Ding, D., Li, J., & Jia, X. (2026). Exploring the antecedents and mitigators of user uncertainty in generative AI use: A principal-agent perspective. *Technovation*, 153, Article 103537. <https://doi.org/10.1016/j.technovation.2026.103537>
- Ding, L., Lawson, C., & Shapira, P. (2025). Rise of generative artificial intelligence in science. *Scientometrics*, 130(9), 5093–5114. <https://doi.org/10.1007/s11192-025-05413-z>
- Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning* (arXiv:1702.08608). arXiv. <https://doi.org/10.48550/arXiv.1702.08608>
- Durán, J. M., & Pozzi, G. (2025). Trust and trustworthiness in AI. *Philosophy & Technology*, 38(1), Article 16. <https://doi.org/10.1007/s13347-025-00843-2>
- Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology*, 71(3), 500–507. <https://doi.org/10.1037/0021-9010.71.3.500>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fehr, J., Citro, B., Malpani, R., Lippert, C., & Madai, V. I. (2024). A trustworthy AI reality-check: The lack of transparency of artificial intelligence products in healthcare. *Frontiers in Digital Health*, 6, Article 1267290. <https://doi.org/10.3389/fdgth.2024.1267290>
- Fisher, J., & Hopp, T. (2020). Does the framing of transparency impact trust? Differences between self-benefit and other-benefit message frames. *International Journal of Strategic Communication*, 14(3), 203–222. <https://doi.org/10.1080/1553118X.2020.1770767>
- Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 535–545). Wiley. <https://doi.org/10.1002/9781119815075.ch45>
- Friedman, B., & Kahn, P. H., Jr. (2003). Human values, ethics, and design. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 1177–1201). Lawrence Erlbaum Associates.
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), 51–90. <https://doi.org/10.2307/30036519>
- Glassberg, I., Brender-Ilan, Y., & Zwilling, M. (2025). The key role of design and transparency in enhancing trust in AI-powered digital agents. *Journal of Innovation & Knowledge*, 10(5), Article 100770. <https://doi.org/10.1016/j.jik.2025.100770>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.

- Hanelt, A., Wiesche, M., Benlian, A., & Hinz, O. (2026). Opening the network of trust: How domain experts in triadic relationships build trust in AI-based counterparts. *MIS Quarterly*, 50(1), 299–326. <https://doi.org/10.25300/MISQ/2025/18041>
- Haresamudram, K., Larsson, S., & Heintz, F. (2023). Three levels of AI transparency. *Computer*, 56(2), 93–100. <https://doi.org/10.1109/MC.2022.3213181>
- Hatherall, L., & Sethi, N. (2025). Exploring expert and public perceptions of answerability and trustworthy autonomous systems. *Journal of Responsible Technology*, 21, Article 100106. <https://doi.org/10.1016/j.jrt.2025.100106>
- High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). *Metrics for explainable AI: Challenges and prospects* (arXiv:1812.04608). arXiv. <https://doi.org/10.48550/arXiv.1812.04608>
- Huynh, M.-T., & Aichner, T. (2025). In generative artificial intelligence we trust: Unpacking determinants and outcomes for cognitive trust. *AI & Society*, 40(8), 5849–5869. <https://doi.org/10.1007/s00146-025-02378-8>
- Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, 70(3), 394–400. <https://doi.org/10.1177/0013164409355692>
- Jung, Y. M., & Jo, H. (2025). Understanding continuance intention of generative AI in education: An ECM-based study for sustainable learning engagement. *Sustainability*, 17(13), Article 6082. <https://doi.org/10.3390/su17136082>
- Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*, 32(3), 713–735. <https://doi.org/10.1287/isre.2020.0980>
- Kang, S., Choi, Y., & Kim, B. (2024). Impact of motivation factors for using generative AI services on continuous use intention: Mediating trust and acceptance attitude. *Social Sciences*, 13(9), Article 475. <https://doi.org/10.3390/socsci13090475>
- Khan, A. R., Mahmood, S. M. F., Akbar, M., & Hossain, S. (2026). Understanding generative AI continuance intention: A dual-factor perspective on facilitators and barriers. *The Electronic Library*, 44(4), 856–875. <https://doi.org/10.1108/EL-07-2025-0328>
- Kim, M., Kim, J., Knotts, T. L., & Albers, N. D. (2025). AI for academic success: Investigating the role of usability, enjoyment, and responsiveness in ChatGPT adoption. *Education and Information Technologies*, 30(10), 14393–14414. <https://doi.org/10.1007/s10639-025-13398-8>
- Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227–261. <https://doi.org/10.1111/isj.12131>
- Korzyński, P., Silva, S. C. e, Górski, A. M., & Mazurek, G. (2026). Trust in AI and top management support in generative-AI adoption. *Journal of Computer Information Systems*, 66(5), 586–600. <https://doi.org/10.1080/08874417.2024.2401986>
- Lalot, F., & Bertram, A.-M. (2025). When the bot walks the talk: Investigating the foundations of trust in an artificial intelligence (AI) chatbot. *Journal of Experimental Psychology: General*, 154(2), 533–551. <https://doi.org/10.1037/xge0001696>
- Lee, C., & Cha, K. (2025). Toward the dynamic relationship between AI transparency and trust in AI: A case study on ChatGPT. *International Journal of Human–Computer Interaction*, 41(13), 8086–8103. <https://doi.org/10.1080/10447318.2024.2405266>
- Lee, C., Kim, J., Lim, J. S., & Shin, D. (2025). Generative AI risks and resilience: How users adapt to hallucination and privacy challenges. *Telematics and Informatics Reports*, 19, Article 100221. <https://doi.org/10.1016/j.teler.2025.100221>

- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. [https://doi.org/10.1518/hfes.46.1.50\\_30392](https://doi.org/10.1518/hfes.46.1.50_30392)
- Li, M., & Bitterly, T. B. (2024). How perceived lack of benevolence harms trust of artificial intelligence management. *Journal of Applied Psychology*, 109(11), 1794–1816. <https://doi.org/10.1037/apl0001200>
- Lindgren, E., Lindholm, T., Vliegenthart, R., Boomgaarden, H. G., Damstra, A., Strömbäck, J., & Tsati, Y. (2024). Trusting the facts: The role of framing, news media as a (trusted) source, and opinion resonance for perceived truth in statistical statements. *Journalism & Mass Communication Quarterly*, 101(4), 981–1004. <https://doi.org/10.1177/10776990221117117>
- Liu, H., Xu, S., Wu, S., Warren, A. D., & Redick, T. S. (2025). The effect of AI benevolence on humans' trust and use of an AI assistant in a video game. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 69(1), 379–381. <https://doi.org/10.1177/10711813251357916>
- Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. <https://aisel.aisnet.org/hicss-54/os/trust/2>
- Marabelli, M., Davison, R. M., Gatti, G., Srivastava, A., & Tarafdar, M. (2026). Use of generative AI in scholarly research: Challenges and opportunities. *Communications of the Association for Information Systems*, 58, Article 36. <https://aisel.aisnet.org/cais/vol58/iss1/36>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
- McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24–59. <https://doi.org/10.5465/256727>
- McGrath, M. J., Duenser, A., Lacey, J., & Paris, C. (2025). Collaborative human-AI trust (CHAI-T): A process framework for active management of trust in human-AI collaboration. *Computers in Human Behavior: Artificial Humans*, 6, Article 100200. <https://doi.org/10.1016/j.chbah.2025.100200>
- McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2), 1–25. <https://doi.org/10.1145/1985347.1985353>
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. <https://doi.org/10.1287/isre.13.3.334.81>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Novozhilova, E., Mays, K., Paik, S., & Katz, J. E. (2024). More capable, less benevolent: Trust perceptions of AI systems across societal contexts. *Machine Learning and Knowledge Extraction*, 6(1), 342–366. <https://doi.org/10.3390/make6010017>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Pathak, A., & Bansal, V. (2024). AI as decision aid or delegated agent: The effects of trust dimensions on the adoption of AI digital agents. *Computers in Human Behavior: Artificial Humans*, 2(2), Article 100094. <https://doi.org/10.1016/j.chbah.2024.100094>
- Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, 15(1), 37–59. <https://doi.org/10.1287/isre.1040.0015>
- Pitts, G., & Motamedi, S. (2025). Understanding human-AI trust in education. *Telematics and Informatics Reports*, 20, Article 100270. <https://doi.org/10.1016/j.teler.2025.100270>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>

- Prasad, K. D. V., Singh, S., Srinivas, V., Kothari, H., & Pathak, A. (2026). Factors influencing the adoption and use of generative artificial intelligence tools among business school students in higher education. *Education and Information Technologies*, 31(10), 3255–3281. <https://doi.org/10.1007/s10639-026-13943-z>
- Radanliev, P. (2025). AI ethics: Integrating transparency, fairness, and privacy in AI development. *Applied Artificial Intelligence*, 39(1), Article 2463722. <https://doi.org/10.1080/08839514.2025.2463722>
- Rahman, M. J., & Liang, H. (2026). When do you want more control over AI? The paradox of autonomy, task stakes, and distrust in AI aversion. *International Journal of Information Management*, 89, Article 103060. <https://doi.org/10.1016/j.ijinfomgt.2026.103060>
- Sagona, M., Dai, T., Macis, M., & Darden, M. (2025). Trust in AI-assisted health systems and AI's trust in humans. *npj Health Systems*, 2(1), Article 10. <https://doi.org/10.1038/s44401-025-00016-5>
- Salih, L., Tarhini, A., & Acikgoz, F. (2025). AI-enabled service continuance: Roles of trust and privacy risk. *Journal of Computer Information Systems*. Advance online publication. <https://doi.org/10.1080/08874417.2025.2452544>
- Sarker, P., & Rashid, M. H. A. (2026). Trust in generative AI: Investigating the role of perceived trustworthiness in shaping user satisfaction. In A. Elbanna, M. Janssen, Y. K. Dwivedi, M. Hossain, & A. S. Islind (Eds.), *Digital adoption, diffusion and innovation in the augmented and digital society* (IFIP Advances in Information and Communication Technology, Vol. 779, pp. 130–143). Springer. [https://doi.org/10.1007/978-3-032-16779-8\\_9](https://doi.org/10.1007/978-3-032-16779-8_9)
- Scharowski, N., Perrig, S. A. C., Svab, M., Opwis, K., & Brühlmann, F. (2023). Exploring the effects of human-centered AI explanations on trust and reliance. *Frontiers in Computer Science*, 5, Article 1151150. <https://doi.org/10.3389/fcomp.2023.1151150>
- Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, Article 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>
- Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems*, 29(4), 260–278. <https://doi.org/10.1080/12460125.2020.1819094>
- Senadheera, S., Yigitcanlar, T., Desouza, K., Mossberger, K., Cheong, P. H., Corchado, J., & Paz, A. (2026). Is responsible AI a public demand? Trust, risk, and persuasion in local government AI. *Information Processing & Management*, 63(8), Article 104932. <https://doi.org/10.1016/j.ipm.2026.104932>
- Shahriari, K., & Shahriari, M. (2017). IEEE standard review: Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)* (pp. 197–201). IEEE. <https://doi.org/10.1109/IHTC.2017.8058187>
- Shahzad, M. F., Xu, S., & Asif, M. (2025). Factors affecting generative artificial intelligence, such as ChatGPT, use in higher education: An application of technology acceptance model. *British Educational Research Journal*, 51(2), 489–513. <https://doi.org/10.1002/berj.4084>
- Shen, J. H., Liu, K., Wang, A., Cen, S. H., Zhang, A. K., Meinhardt, C., Zhang, D., Klyman, K., Bommasani, R., & Ho, D. E. (2026). *The limits of AI data transparency policy: Three disclosure fallacies* (arXiv:2601.18127). arXiv. <https://doi.org/10.48550/arXiv.2601.18127>
- Shen, L., Qiu, N., & Wang, Z. (2025). Psychological safety and trust as drivers of teachers' continued use of AI tools in classrooms. *Scientific Reports*, 15, Article 31426. <https://doi.org/10.1038/s41598-025-13789-4>
- Silalahi, A. D. K. (2025). Can generative artificial intelligence drive sustainable behavior? A consumer-adoption model for AI-driven sustainability recommendations. *Technology in Society*, 83, Article 102995. <https://doi.org/10.1016/j.techsoc.2025.102995>

- Simons, T. (2002). Behavioral integrity: The perceived alignment between managers' words and deeds as a research focus. *Organization Science*, 13(1), 18–35. <https://doi.org/10.1287/orsc.13.1.18.543>
- Singh, S., & Strzelecki, A. (2026). Academics as adopters of generative AI: An application of diffusion of innovations theory. *Education and Information Technologies*, 31(2), 621–645. <https://doi.org/10.1007/s10639-025-13835-8>
- Sondern, D., & Hertel, G. (2024). Revisiting the classic ABI model of trustworthiness: Interactive effects of trustworthiness components on trust in mixed-motive social exchange contexts. *Journal of Trust Research*, 14(2), 213–236. <https://doi.org/10.1080/21515581.2024.2388659>
- Szatmary, R. (2025). Trust and security: How trust in autonomous systems shapes decision-making. In T. A. Kovács, R. G. Stadler, & N. Daruka (Eds.), *The impact of the energy dependency on critical infrastructure protection* (pp. 685–703). Springer. [https://doi.org/10.1007/978-3-031-78544-3\\_55](https://doi.org/10.1007/978-3-031-78544-3_55)
- Tang, A., Li, K.-K., Kwok, K. O., Cao, L., Luong, S., & Tam, W. (2024). The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing. *Journal of Nursing Scholarship*, 56(2), 314–318. <https://doi.org/10.1111/jnu.12938>
- Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 272–283). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372834>
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458. <https://doi.org/10.1126/science.7455683>
- Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H., & Seaborn, K. (2022). Trust in human-AI interaction: Scoping out models, measures, and methods. In *Extended abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). Association for Computing Machinery. <https://doi.org/10.1145/3491101.3519772>
- Varnum, C. N., Wojdyski, B. W., Binford, M. T., & Duncan, J. (2025). Designing for trust: How online news consumers view and interpret informational transparency boxes. *Digital Journalism*, 13(9), 1625–1646. <https://doi.org/10.1080/21670811.2024.2346749>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Verboom, A. D. P. R., Pais, L., Zijlstra, F. R. H., Oswald, F. L., & Santos, N. R. d. (2025). Perceptions of artificial intelligence in academic teaching and research: A qualitative study from AI experts and professors' perspectives. *International Journal of Educational Technology in Higher Education*, 22(1), Article 46. <https://doi.org/10.1186/s41239-025-00546-w>
- Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22(2), 219–240. <https://doi.org/10.1111/j.1540-5915.1991.tb00344.x>
- Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13(1), Article 14045. <https://doi.org/10.1038/s41598-023-41032-5>
- Wang, P., & Ding, H. (2024). The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-AI trust and decision performance. *Information Processing & Management*, 61(4), Article 103732. <https://doi.org/10.1016/j.ipm.2024.103732>
- Wiedemann, P. M., Kirsch, F., Lohmann, M., Böhl, G.-F., & Freudenstein, F. (2022). Effects of as-if risk framing of hazards on risk perception and its rebuttal. *Regulatory Toxicology and Pharmacology*, 136, Article 105282. <https://doi.org/10.1016/j.yrtph.2022.105282>
- Xiong, L., Zhou, Y., & Hua, Y. (2026). Why do users continue to use GCAI? A mixed-methods study. *Technology in Society*, 86, Article 103309. <https://doi.org/10.1016/j.techsoc.2026.103309>

- Zhou, T., & Ma, X. (2025). Examining generative AI user continuance intention based on the SOR model. *Aslib Journal of Information Management*, 78(4), 969–987. <https://doi.org/10.1108/AJIM-08-2024-0620>
- Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology*, 14, Article 1259531. <https://doi.org/10.3389/fpsyg.2023.1259531>

Copyright © 2026 The Author(s). Published by Asia Pacific University of Technology & Innovation (APU). This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International License  (<https://creativecommons.org/licenses/by/4.0/>).