



Article

Epistemic Trust in Generative AI as an Information Source: Development and Validation of the Trust in AI-Epistemic Scale (TAI-E)

László Balázs

Department of Organizational Development and Communication Science, University of Dunaújváros,
2400 Dunaújváros, Hungary; balazsl@uniduna.hu

Abstract

As generative AI increasingly mediates everyday information, users must judge not only whether a claim is true but whether the system conveying it is a legitimate source of belief. This article develops and validates the Trust in AI-Epistemic Scale (TAI-E), measuring epistemic trust in generative AI across three dimensions: attribution of epistemic authority, perceived epistemic reliability, and suspension of critical judgment. In a survey of part-time higher-education students ($N = 412$; ages 18–63), the three-factor structure was replicated across a calibration–validation split ($CFI = 0.944$; $RMSEA = 0.085$; subscale $\alpha = 0.84$ – 0.87). Granting a system epistemic authority and suspending critical judgment toward it behaved as partly opposing stances, so the scale captures a configuration rather than a single trust score. Self-reported reflective processing accompanied higher authority and reliability attributions together with greater critical engagement. Most strikingly, self-reported reflectiveness barely tracked behavioural cognitive reflection ($r = 0.14$), and only the self-report predicted trust—suggesting it reflects epistemic self-image as much as reflective capacity. Relations among constructs are reported as construct-validity associations rather than causal effects. The TAI-E provides a psychometrically grounded tool for studying epistemic trust, AI literacy, and well-founded reliance on AI.

Keywords: epistemic trust in AI; scale development; reflective information processing; sense of coherence; human–AI interaction; AI literacy; epistemic authority

1. Introduction

The rapid integration of artificial intelligence (AI) systems into information ecosystems is reshaping the conditions under which users orient themselves epistemically. AI is increasingly not merely a data-processing tool but a cognitive-mediating infrastructure that participates in the selection, structuring, and presentation of information; consequently, users often construct knowledge about reality indirectly, through technological filters (Glikson and Woolley 2020; Hepp 2020; Shneiderman 2020). Under these conditions, trust becomes a particularly important component of epistemic decision-making: users must evaluate not only informational content but also the epistemic legitimacy of the system through which that content is mediated (Jones and Chandrasekharan 2024; Shin 2025; Sahut and Tricot 2017; Origgi 2008).

These conditions also shape the environment in which generative AI operates: algorithmic curation, the constraints of the attention economy, and the ecology of disinformation place users in a complex and partly opaque knowledge landscape (Bawden and Robinson



Academic Editor:

Kavitha Balakrishnan

Received: 29 April 2026

Revised: 24 June 2026

Accepted: 29 June 2026

Published: 1 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

2009; Lewandowsky et al. 2017; Vosoughi et al. 2018). In this context, epistemic trust in generative AI entails a dual risk. Insufficient trust may lead to the underuse of systems that can genuinely extend users' epistemic reach, whereas excessive or insufficiently differentiated trust may foster the uncritical acceptance of fabricated content, sycophantic responses, and model-embedded biases (Hoff and Bashir 2015; Parasuraman and Riley 1997). A substantial research programme addresses one response to this tension under the heading of trust calibration, asking whether users' reliance tracks a system's actual capabilities and limitations (Wischnewski et al. 2023; Kaplan et al. 2023; Lee and See 2004; Bansal et al. 2019). The present study does not measure calibration in this technical sense: it does not assess system performance, and therefore makes no claim about whether users' trust is well matched to it. Instead, it addresses a logically prior and distinct question—how to characterize and measure the epistemic stance users adopt toward generative AI as a source of knowledge—which the calibration literature presupposes but does not directly operationalize.

The article makes two main contributions. First, it operationalizes epistemic trust in generative AI as a distinct construct: trust is treated not as a general attitude toward technology but as the epistemic legitimacy a user grants generative AI as a source of knowledge. To capture this stance, we developed and validated a three-dimensional scale (Trust in AI-Epistemic Scale, TAI-E), comprising epistemic authority attribution, perceived epistemic reliability, and suspension of critical judgment. Second, we situate the instrument within its nomological network by examining how its dimensions relate to conceptually adjacent constructs (self-reported reflective information processing, sense of coherence), to a behavioural index of cognitive reflection, and to criterion variables not used in scale construction (frequency of AI use, age). Together, these analyses provide evidence concerning the structural, convergent, discriminant, and criterion-related validity of the TAI-E. The article is organized accordingly. Following the theoretical background, the development and psychometric evaluation of the TAI-E are reported—item generation, exploratory and confirmatory factor analysis on a calibration–validation split, reliability, and convergent and discriminant validity (Section 3). The subsequent section then situates the instrument within its nomological network as further construct-validity evidence (Section 4). These analyses are exploratory and are interpreted as associations rather than as tests of a causal model.

2. Theoretical Background

2.1. Epistemic Challenges of the Digital Information Environment

Platform-based media systems, algorithmic content selection, and deep mediatization (Couldry and Hepp 2017; Hepp 2020) have made the information space increasingly personalized, dynamic, and partly opaque. The growth in available content is burdensome in itself, yet information overload is best understood not only as a quantitative problem but also as a structural one: fragmentation, loss of context, and redundancy may jointly weaken the coherence of knowledge construction (Arnold et al. 2023; Moko et al. 2023; Bawden and Robinson 2009; Eppler and Mengis 2004). The limits of users' cognitive capacities, together with the pressure to make rapid judgments, create conditions of uncertainty in which assessments of informational reliability are often made indirectly, through technological filters (Andrejevic 2013; Lorenz-Spreen et al. 2019).

The emergence of AI systems adds a further layer to this dynamic. AI not only filters information but also generates and interprets it, meaning that users receive a growing share of their knowledge from an intermediary whose internal operations remain partly or wholly hidden—the familiar problem of the “black box” (Zednik 2021; Reinhardt and Buchholz 2025; Kizilcec 2016; Doshi-Velez and Kim 2017). Consequently, information

processing becomes not only a cognitive task but also an epistemic decision: users must evaluate the content of information, the credibility of its source, and the mechanism by which it is mediated (Neuberger et al. 2025).

Analytical precision is needed because “AI” denotes a heterogeneous family of technologies with distinct epistemic roles. Recommender systems and social-media ranking algorithms shape which information a user encounters; search engines retrieve and order existing sources; AI-based decision-support systems issue recommendations within constrained task domains; and generative AI systems—large language models in particular—compose novel, fluent responses that users may treat as direct answers to epistemic questions. These classes raise different trust problems. A recommender’s influence is largely curatorial, whereas a generative system may function as an apparent interlocutor and source, producing content whose provenance and reliability are difficult to verify. The present inquiry therefore concerns the latter case: epistemic trust in generative AI used as a source of knowledge. Where evidence from adjacent system classes is invoked, it is treated as a conceptual parallel rather than assumed to generalize automatically.

2.2. Reflective Information Processing

Dual-process models (Evans and Stanovich 2013; Kahneman 2011) conceptualize human information processing in terms of two basic modes: a fast, automatic, and heuristic process (System 1) and a slower, resource-intensive, analytical process (System 2). Reflective processing is linked to the latter: the individual is able to review their intuitive responses, weigh contradictions, and consciously manage informational uncertainty. The characteristics of the digital environment—the speed of the stimulus flow, the constraints of the attention economy—typically reinforce the dominance of intuitive, heuristic strategies (Stanovich et al. 2016), yet the availability of reflective processing varies significantly across individuals and contexts.

In the current analysis, the self-report measure of reflective information processing refers to a self-attributed dispositional tendency to engage in analytic, effortful evaluation of information rather than accepting it at face value: questioning the plausibility of claims, attending to source and justification, considering alternative interpretations, and monitoring one’s initial responses. So defined, the construct is close to what the rationality literature calls reflective or actively open-minded thinking (Stanovich et al. 2016), and it is conceptually adjacent to—but not identical with—metacognitive monitoring and the disposition to override intuitive responses indexed by tasks such as the Cognitive Reflection Test (Frederick 2005).

At the same time, anchoring the construct in dual-process theory requires caution, because the mapping between “reflective” processing and the System 1/System 2 distinction remains contested. The two-systems framing has been criticized for collapsing heterogeneous processes into a single dichotomy and for implying that deliberate, analytic cognition is necessarily slow, conscious, and voluntary (Evans and Stanovich 2013). A parallel debate in the philosophy of mind questions whether reasoning and judgment are aptly construed as actions at all: on some accounts, forming and revising belief involves a distinctive, non-voluntary evaluative control rather than the managerial, effortful control suggested by the System 2 image (Hieronymi 2009; Vierkant 2022). We therefore do not treat reflective information processing as equivalent to “System 2 activation.” Instead, we use the term in the more modest sense specified above—a measurable disposition toward analytic, source-sensitive evaluation—while remaining agnostic about its precise cognitive architecture.

Two interpretive caveats follow from how the construct is measured. First, the instrument is a self-report scale: it captures respondents’ perceptions of their own cognitive

tendencies, not their observed processing behaviour. A well-documented gap exists between self-reported reflectiveness and performance on reflection-sensitive tasks (Stanovich et al. 2016), so the measure is most appropriately read as indexing a self-attributed disposition. Second, any association between this measure and epistemic trust should therefore be interpreted as a relationship between self-perceived reflectiveness and trust, not as evidence that individuals who actually process information more reflectively place more—or less—epistemic trust in generative AI. This limitation is revisited in the interpretation of the results.

In AI-mediated information situations, reflective processing is particularly relevant. The construct of automation bias—the tendency to accept a system’s outputs without adequate scrutiny—was originally documented in automated decision-support and human–automation settings (Cummings 2004; Parasuraman and Riley 1997; Alon-Barkat and Busuioc 2023). A comparable dynamic may plausibly extend to generative AI used as an information source, where fluent and confident outputs can be accepted uncritically (Romeo and Conti 2026), although this transfer across system classes should be treated with caution. In such contexts, automatic acceptance of system outputs may reduce the quality of information processing—the depth and accuracy with which outputs are scrutinized—and increase the risk of erroneous judgments. The relationship between epistemic trust in generative AI and reflective processing may therefore have implications for the epistemic quality of the resulting beliefs (their accuracy and justification), while remaining an empirical association in the present design.

2.3. Epistemic Trust in AI as a Multidimensional Construct

Before operationalizing the construct, it is necessary to clarify which notion of trust is at issue, since trust and epistemic trust are often conflated in the AI literature. A substantial body of work in the philosophy of technology questions whether AI systems are appropriate objects of trust in the full sense, arguing that trust involves vulnerability, normative expectations, and accountability that artificial systems may not possess; on this view, what users extend to AI is more accurately described as reliance than as trust (Ryan 2020; Bryson 2018; Deroy 2023; Dorsch and Deroy 2024). The present study does not take a position on whether AI can be trusted in this thick interpersonal sense. Instead, it adopts the narrower notion of epistemic trust: the extent to which a user treats a source as a legitimate basis for belief and accepts its outputs as justified without independent verification (Origgi 2008; Alvarado 2023a, 2023b; Moll and Dorsch 2025). Epistemic trust is thus distinct both from generalized trust and from mere reliance; it concerns the conditions under which a user may be epistemically justified in believing what a generative AI system conveys. This is the sense in which the term is used throughout the paper.

In the early wave of the AI-trust literature, trust was primarily viewed as a precursor to technology acceptance (Davis 1989; Hoff and Bashir 2015). However, recent approaches clearly emphasize the multidimensional and context-dependent nature of trust (Kaplan et al. 2023; Choung et al. 2023; Glikson and Woolley 2020; Thiebes et al. 2021). Mayer et al.’s (1995) classic three-part framework—competence, benevolence, and integrity—appears in an adapted form in AI research, supplemented by the dimensions of transparency, accountability, and explainability (Doshi-Velez and Kim 2017; Miller 2019).

The present investigation focuses specifically on epistemic trust in generative AI. Within social epistemology, epistemic trust refers to the extent to which an individual regards a source as legitimate and accepts the content it conveys as valid and usable (Neuberger et al. 2025; Bartsch et al. 2024; Origgi 2008; Sperber et al. 2010). This issue is especially relevant for AI because direct verification of outputs is often difficult; decisions to accept them therefore rely partly on the epistemic authority attributed to the system. A

conceptual distinction is useful here: epistemic authority concerns the credibility ascribed to AI as a source, whereas a separate judgment concerns the perceived accuracy of the outputs it produces. The present construct addresses the former (authority attributed to the source) and the latter (perceived reliability of outputs) as related but distinct dimensions.

To make this stance measurable, the TAI-E operationalizes three interrelated but analytically distinct dimensions. Attribution of *epistemic authority* refers to the extent to which users regard a generative AI system as a legitimate source of knowledge whose outputs may warrant belief. *Perceived epistemic reliability* refers to the degree to which users assess the system's outputs as accurate, sound, and truth-conducive. The definition deliberately avoids using the word trust, so that this dimension can help explain epistemic trust rather than presuppose it. *Suspension of critical judgment* refers to the extent to which a user is willing to temporarily set aside independent critical evaluation when engaging with AI outputs. Because the present study relies on questionnaire self-report, this dimension captures users' reported disposition to suspend critical judgment rather than observed suspension in practice; distinguishing genuine suspension from ordinary, evidence-responsive belief revision would require a behavioural paradigm, such as a change-of-mind design.

2.4. Sense of Coherence (SOC) in the Context of Digital Information

In Antonovsky's (1987, 1993) salutogenic model, the sense of coherence (SOC) consists of three dimensions: comprehensibility, manageability, and meaningfulness. SOC does not capture stress resistance, but rather the ability to constructively confront complex and unpredictable situations—in this respect, it differs from the concept of resilience, with which it is sometimes mistakenly equated in the literature (Mittelmark et al. 2022). This study consistently upholds this terminological distinction.

In a digital information context, the theoretical relevance of SOC lies in the fact that individuals with a higher sense of coherence may perceive the uncertainty generated by information overload in a more structured and controllable manner (Stoyanova and Stoyanov 2021; Arnold et al. 2023; Bawden and Robinson 2009). At the same time, the effect of SOC in AI-mediated situations is not necessarily direct: in a mediated information environment, psychological resources primarily exert their influence through mediating mechanisms. A direct association between SOC and AI trust or reflective processing is therefore not necessary even at the theoretical level.

In the present analyses, SOC is included solely as a control variable. It is not part of the epistemic-trust construct, nor is a direct association with epistemic trust theoretically expected. In a mediated information environment, general psychological resources of this kind are likely to operate, if at all, through more proximal cognitive mechanisms. SOC is therefore retained to rule out the possibility that observed associations reflect a general dispositional resource rather than processing-specific variance.

2.5. Expected Relations in the Nomological Network

Because the present research is intended to develop and validate the TAI-E rather than to adjudicate causal hypotheses, we do not formulate directional predictions about a mediating mechanism. Instead, we outline a set of theoretically grounded expectations regarding how a valid measure of epistemic trust in generative AI would be likely to behave within its nomological network. These associations are interpreted as construct-validity evidence in the sense of Cronbach and Meehl (1955), and are specified in line with contemporary recommendations for transparent construct validation (Flake et al. 2017; Clark and Watson 1995).

First, the three dimensions of the TAI-E should be empirically separable from one another and from conceptually adjacent constructs. This distinction is particularly important

in relation to the self-reported reflective information processing measure, which may partially overlap with suspension of critical judgment. Establishing such differentiation follows the classic logic of convergent and discriminant validation (Campbell and Fiske 1959).

Second, insofar as epistemic trust is connected to how users evaluate and process information, the trust dimensions would be expected to show meaningful, but not necessarily uniform, associations with self-reported reflective information processing. A more analytic and source-sensitive disposition may plausibly be associated with more differentiated authority attribution and stronger critical engagement, rather than with uniformly higher or lower levels of trust. This interpretation is consistent with accounts suggesting that deeper user processing can support calibrated reliance instead of simply reducing trust (Lee and See 2004; Wischniewski et al. 2023; Stanovich et al. 2016).

Third, the TAI-E should display theoretically interpretable associations with criterion variables that were not used in its construction, such as frequency of AI use and age (Logg et al. 2019; Gerlich 2024; Hargittai 2010). At the same time, it would be expected to remain largely independent of broader, non-epistemic psychological resources, such as sense of coherence (Antonovsky 1987; Mittelmark et al. 2022). Such a pattern would indicate that the scale captures a domain-specific epistemic orientation toward AI rather than a generalized psychological disposition.

Finally, because reflective information processing is assessed by self-report, we also examine whether it converges with a behavioural index of cognitive reflection (the three-item Cognitive Reflection Test; Frederick 2005; see Section 4.1.2). Any divergence between these indicators is treated as theoretically informative, as it may clarify whether the self-report measure primarily captures perceived analytic disposition, actual reflection-sensitive performance, or partly distinct aspects of information processing. This precaution is warranted by the well-documented gap between self-reported analytic tendencies and cognitive reflection performance (Frederick 2005; Stanovich et al. 2016).

These expectations guide the analyses reported in Section 4. They are therefore framed as validity-relevant associations rather than as tests of a causal model.

3. Development and Validation of the TAI-E Scale

The aim of this paper is to develop and validate a psychometrically sound scale suitable for measuring epistemic trust in AI. Below, we describe the logic behind item development, the sample and methodology, followed by the results of the EFA, CFA, and reliability, convergent, and discriminant validity analyses.

3.1. Method

3.1.1. Construct-Theoretical Approach and Item Development

The three frameworks were chosen because each anchors a different facet of the construct. Origi's (2008) social-epistemology account supplies its core—the legitimacy a hearer grants a source as a basis for belief—and motivates the authority dimension; Siau and Wang's (2018) taxonomy specifies the AI-specific bases on which such legitimacy is granted or withheld; and Thiebes et al.'s (2021) reliability framework grounds the perceived-reliability dimension. Together they map onto the three TAI-E dimensions and onto the suspension-of-critical-judgment pole that complements them. The item development was based on the three-dimensional concept of epistemic trust mentioned in Section 2.3. Initial item pool: 30 items. To ensure the content validity of the measurement instrument, we employed an expert review process. The questionnaire initially contained 30 items, which were evaluated by three experts representing complementary fields (psychology, computer science, linguistics). Although the panel did not include a philosopher, the epistemological grounding of the construct was drawn directly from the social-epistemology literature

(Origgi 2008; Sperber et al. 2010; Alvarado 2023a), which informed both the definition of the three dimensions and the wording of the items. The multidisciplinary approach was intended to ensure that cognitive-psychological soundness, technological relevance, and linguistic and interpretive clarity were all considered. Experts assessed relevance, clarity, and alignment with the construct. As a result of this process, 8 items were discarded, leaving 22 items for further analysis. The retention rate of the items (73.3%) suggests adequate construct coverage while allowing the item pool to be narrowed. Based on the feedback, several items were refined, especially those addressing epistemic dimensions such as perceived reliability and trust in justification, where conceptual and linguistic precision is particularly important. Thus, content validity was provisionally supported before the factor-analytic phase. Items with low content validity were removed prior to the EFA; the initial 22-item scale was subsequently reduced to 12 items through the EFA procedure.

3.1.2. Sample and Procedure

Data collection took place via an online questionnaire using Google Forms during the first three weeks of April 2026. The questionnaire was distributed to students enrolled in correspondence (part-time) programmes at several higher-education institutions via institutional communication channels. The inclusion criteria required that the respondent (1) be at least 18 years of age, (2) be a native Hungarian speaker or possess a high level of Hungarian-language proficiency, and (3) use generative AI tools at least occasionally (e.g., ChatGPT, Claude, Gemini, Copilot, Midjourney). Both the eligibility screening and the usage-frequency item referred explicitly to generative AI, so respondents answered the trust items with generative AI systems in mind even where the item wording referred generically to “AI”. This criterion was included because epistemic trust in generative AI can be meaningfully assessed only among respondents with at least direct or indirect experience of such systems. Exclusion criteria comprised incomplete responses, being under 18 years of age, and reporting no generative AI experience; 20 cases (4.6% of the initial responses) were removed through application of these criteria rather than because of missing data, yielding a final analytic sample of $N = 412$ (listwise deletion).

Because part-time programmes typically serve working adults, the sample covered a wide age range and heterogeneous occupational backgrounds rather than a traditional-age undergraduate population. Respondents ranged in age from 18 to 63 years ($M = 36.56$), and 62% were women. Educational attainment was distributed as follows: primary and secondary education 75.7% (secondary education 59.7%) and higher education 24.2% (BA/BSc 15.5%; MA/MSc 8.7%). The occupational distribution showed relative overrepresentation in engineering (24.8%) and IT (18.9%), indicating a technology-oriented, AI-exposed population that should be considered when assessing generalizability.

Participation was voluntary and anonymous; no incentives, compensation, or course credit were offered. Participants were informed in advance about the purpose of the study and the terms of data processing, and by completing the questionnaire they consented to the use of their data. The procedure complied with relevant ethical and data-protection regulations. For the factor analyses, the sample was randomly split into a calibration subsample for the EFA and an independent validation subsample for the CFA (see Section 3.2.1).

3.1.3. Measures

On the 22-item baseline scale, respondents rated each item on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree). The items on the Suspension of Critical Judgment subscale were reverse-coded. The final 12 items retained after the exploratory factor analysis are listed, in both English and Hungarian, in Appendix A. For reasons of brevity,

the instrument retains the label “Trust in AI.” This label, however, should not be read as referring to AI in an unrestricted sense. In the present questionnaire, the generic item wording was embedded in a generative-AI-oriented context: both the screening item and the AI-use-frequency item explicitly referred to tools such as ChatGPT, Claude, Gemini, Copilot, and Midjourney. It is therefore reasonable to interpret respondents’ answers as referring primarily to generative AI systems. On this basis, the validated construct is specified as epistemic trust in generative AI as a source of knowledge.

3.2. Results

3.2.1. Exploratory Factor Analysis (EFA)

Before conducting the factor analysis, the Kaiser–Meyer–Olkin value ($KMO = 0.898$) and Bartlett’s test of sphericity ($\chi^2 = 2751.823$; $p < 0.001$) indicated that the data were suitable for factor analysis. Given the deviation from normality in the sample, we used principal axis factoring (PAF) with oblimin rotation. This choice does not assume orthogonal factors and is theoretically appropriate because the dimensions of epistemic trust are expected to be conceptually interrelated.

Item reduction from the original 22-item pool to the final 12 items was performed on the full sample using parallel analysis, the eigenvalue–scree criterion, and loading-based retention rules. To reduce the risk that exploratory and confirmatory analyses on the same sample would overstate stability, the retained 12-item structure was then cross-validated using a random split: an EFA on a calibration subsample ($N = 206$) and a CFA on an independent validation subsample ($N = 206$). The two halves did not differ on age ($t = -0.34$, $p = 0.731$) or on the TAI-E dimension scores. In the calibration subsample, the three-factor solution replicated cleanly, explaining 66.7% of the total variance, with all items loading on their intended factor ($\lambda = 0.59$ – 0.92) and no cross-loadings ≥ 0.32 . The EFA results are summarized in Table 1.

Table 1. TAI-E EFA results: factor loadings and communalities (calibration subsample, $n = 206$).

Item (Code)	F1 Authority	F2 Reliability	F3 Suspension	h^2
TRUST_1	–	0.92	–	0.89
TRUST_2	–	0.87	–	0.79
TRUST_3	–	0.73	–	0.55
TRUST_6	–	0.73	–	0.58
AUTH_1	0.82	–	–	0.70
AUTH_2	0.72	–	–	0.60
AUTH_3	0.82	–	–	0.69
AUTH_5	0.77	–	–	0.61
CRIT_1_R	–	–	0.83	0.70
CRIT_2_R	–	–	0.59	0.47
CRIT_3_R	–	–	0.72	0.62
CRIT_5_R	–	–	0.86	0.78
Eigenvalue/%	2.77/23%	2.75/23%	2.47/21%	66.7%

Note. $N = 206$ (calibration subsample); PAF, oblimin rotation; loadings < 0.32 omitted; R = reverse-coded.

The interpretation of the three factors is consistent with the theoretical construct. F1: Attribution of epistemic authority encompasses items that recognize AI as a legitimate source of knowledge. F2: Perceived Epistemic Reliability captures the perception of the reliability of information provided by AI. F3: Suspension of critical judgment measures the willingness to temporarily suspend one’s own critical judgment. Factor loadings appeared on the expected factors in all cases ($\lambda = 0.59$ – 0.92). Due to the oblimin rotation, the secondary loadings of some items are negative, which is a natural consequence of the correlations between factors and does not affect the interpretation. The correlations between factors are

of moderate to high strength ($r = 0.422\text{--}0.627$), indicating both interrelatedness among the constructs and adequate distinctiveness.

3.2.2. Confirmatory Factor Analysis (CFA)

To test the three-factor structure, we conducted confirmatory factor analysis (CFA) using the SEM module of the Jamovi software (version 2.7.30. based on Lavaan), employing the maximum likelihood (ML) estimation method. The fit indices are presented in Table 2.

The CFA was conducted on the validation subsample ($N = 206$) using robust maximum likelihood (MLR), with the full sample ($N = 412$) included as a supplementary robustness model. In the validation subsample, the three-factor model showed acceptable-to-marginal fit ($CFI = 0.916$; $TLI = 0.891$; $RMSEA = 0.099$). In the full sample, fit improved to the range reported for the development model ($CFI = 0.944$; $TLI = 0.927$; $RMSEA = 0.085$; $SRMR = 0.048$). Standardized loadings were significant and substantively adequate ($\lambda = 0.58\text{--}0.88$). The three-factor structure replicated across the calibration-validation split; that the more marginal fit emerged in the smaller validation half ($N = 206$) is consistent with the reduced precision expected at this subsample size rather than with a problem in the structure. Independent replication on a separately collected sample would nonetheless further strengthen the instrument. The correlations between latent factors fell within the moderate-to-strong range ($r = 0.537\text{--}0.752$). Convergent validity was assessed using average variance extracted, and discriminant validity using the Fornell–Larcker criterion. The CFA results are illustrated in Figure 1.

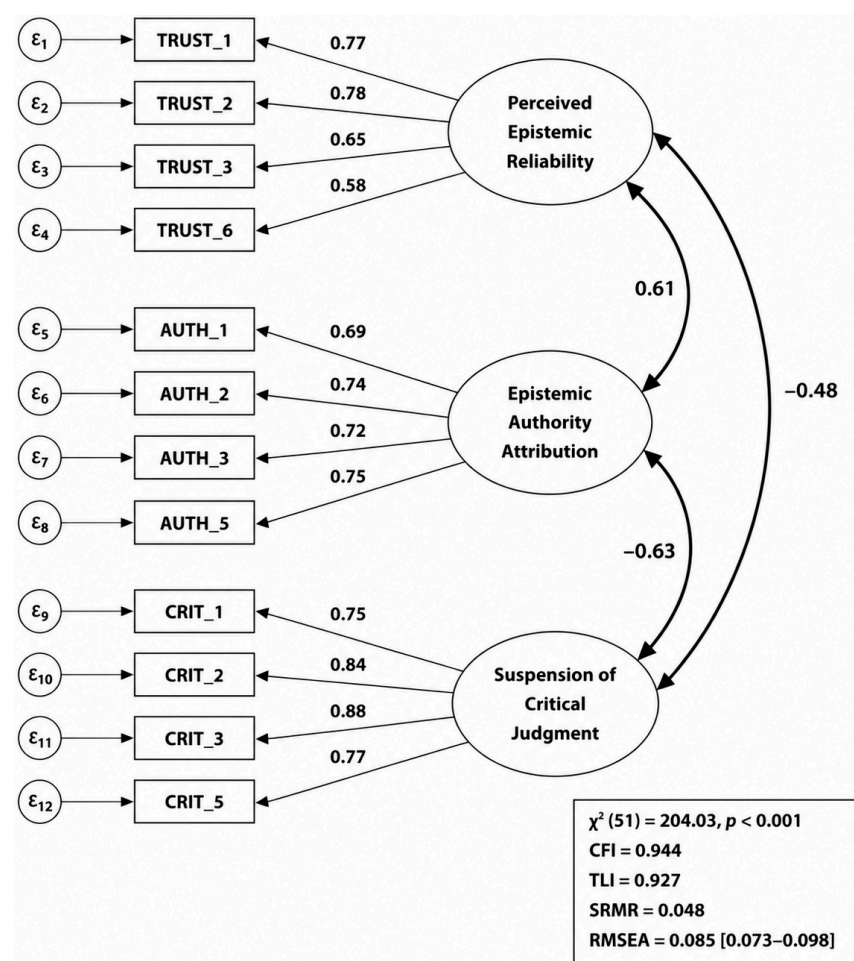


Figure 1. Three-factor confirmatory factor model of the TAI-E scale (standardized estimates).

Table 2. TAI-E CFA fit indices.

Index	Validation (N = 206)	Full (N = 412)	Threshold
χ^2/df	153.79 (51)	204.03 (51)	-
CFI	0.916	0.944	≥ 0.95 good
TLI	0.891	0.927	≥ 0.95 good
SRMR	-	0.048	≤ 0.08 good
RMSEA	0.099	0.085	≤ 0.08 good

Note. CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean Square Error of Approximation.

3.2.3. Reliability

The three TAI-E subscales show good-to-excellent internal consistency: Epistemic Authority Attribution ($\alpha = 0.85$, $\omega = 0.90$), Perceived Epistemic Reliability ($\alpha = 0.84$, $\omega = 0.89$), and Suspension of Critical Judgment ($\alpha = 0.87$, $\omega = 0.91$). A single overall TAI-E composite, however, is not supported: the inter-factor correlation matrix is not positive-manifest—Suspension of Critical Judgment correlates negatively with both Epistemic Authority ($r = -0.63$) and Perceived Reliability ($r = -0.48$)—so the three dimensions do not form a unidimensional higher-order factor. Consistent with this, a 12-item total score yields poor internal consistency ($\alpha = 0.47$, $\omega = 0.56$). This pattern is theoretically coherent rather than anomalous: attributing epistemic authority and reliability to AI is, by construction, the opposite pole of suspending one’s critical judgment. We therefore score and analyse the TAI-E at the subscale level throughout, and do not use a single composite score. The reliability coefficients for the three TAI-E subscales and the related validation measures are reported in Table 3.

Table 3. Internal consistency of the TAI-E subscales.

Scale/Subscale	Number of Items	Cronbach- α	ω	Evaluation
TAI-E (treated as three subscales; no overall composite)				
F1: Epistemic authority attribution	4	0.85	0.90	Good
F2: Perceived epistemic reliability	4	0.84	0.89	Good
F3: Suspension of critical judgment	4	0.87	0.91	Good
Self-reported reflective information processing	5	0.75	0.83	Acceptable
Sense of coherence (SOC-13)	13	0.78	0.83	Acceptable
Cognitive Reflection Test (CRT, 3 items)	3	0.60 ^a	-	Acceptable (binary)

Note. α = Cronbach’s alpha; ω = McDonald’s omega. Evaluation thresholds follow [George and Mallery \(2005\)](#): $\alpha \geq 0.90$ = excellent; ≥ 0.80 = good; ≥ 0.70 = acceptable. A single TAI-E composite is not reported because the three dimensions are not positive-manifest (see Section 3.2.3); the scale is analysed at the subscale level. ^a For the binary CRT items, the coefficient is KR-20 (the Kuder–Richardson equivalent of α); ω is not reported for dichotomous items.

3.2.4. Convergent and Discriminant Validity

Convergent validity was supported by significant standardized loadings ($\lambda = 0.58$ – 0.88 , all $p < 0.001$) and composite reliabilities ($\text{CR} = 0.79$ – 0.88). Average variance extracted was adequate for Authority ($\text{AVE} = 0.52$) and Suspension ($\text{AVE} = 0.66$) and marginally below the 0.50 benchmark for Reliability ($\text{AVE} = 0.49$). Given the strong CR values and substantively adequate loadings, convergent validity is considered acceptable. Discriminant validity was examined using the Fornell–Larcker criterion and the HTMT ratio ([Henseler et al. 2015](#)). According to the Fornell–Larcker criterion, the square root of each dimension’s AVE (Authority $\sqrt{\text{AVE}} = 0.72$; Reliability $\sqrt{\text{AVE}} = 0.70$; Suspension of Critical Judgment $\sqrt{\text{AVE}} = 0.81$) exceeded that dimension’s correlations (in absolute value) with every other dimension; the largest association was $|r| = 0.63$ between Authority and Suspension. This

pattern supports discriminant validity among the three TAI-E factors. The same criterion separated RIP from Suspension of Critical Judgment (RIP $\sqrt{\text{AVE}} = 0.86$ vs. $|r| = 0.33$), converging with the HTMT and the two-factor CFA in indicating that the two constructs are distinct. Given the conceptual proximity between Suspension of Critical Judgment and self-reported reflective information processing (RIP), we also estimated a four-construct model including RIP. HTMT (Suspension, RIP) = 0.40, well below the 0.85 threshold, and a model forcing Suspension and RIP onto a single factor fit poorly (CFI = 0.70) compared with a two-factor model (CFI = 0.95; $\Delta\text{CFI} = 0.26$). These results provide additional support for empirical distinctiveness. All HTMT values among the three TAI-E dimensions were ≤ 0.73 . The HTMT ratios supporting discriminant validity are presented in Table 4.

Table 4. Discriminant validity of the TAI-E dimensions and self-reported reflective information processing: heterotrait–monotrait (HTMT) ratios.

	AUTH	TRUST	CRIT	RIP
AUTH	–	0.72	0.73	0.44
TRUST		–	0.56	0.40
CRIT			–	0.40
RIP				–

Note. HTMT = heterotrait–monotrait ratio of correlations (Henseler et al. 2015); values < 0.85 indicate adequate discriminant validity. AUTH = Epistemic Authority Attribution; TRUST = Perceived Epistemic Reliability; CRIT = Suspension of Critical Judgment; RIP = Self-reported Reflective Information Processing. The HTMT between Suspension of Critical Judgment and RIP (0.40) supports the interpretation that the two constructs are empirically distinct.

3.3. Summary

Overall, the findings provides support for the psychometric validity of the TAI-E's three-factor structure. The subscales show good-to-excellent internal consistency, strong and significant factor loadings, and a structure that is broadly consistent across exploratory and confirmatory analyses. Model fit indices are generally acceptable, although not uniformly optimal. Convergent validity was adequate, and discriminant validity was supported by the Fornell–Larcker criterion and HTMT analyses, despite moderate overlap between some dimensions.

Nevertheless, further replication would strengthen confidence in the instrument. The 12-item three-factor structure replicated across a calibration–validation split, and all three subscales were internally consistent. Because the dimensions are not positive-manifest, the TAI-E is treated as a genuinely multidimensional instrument and is analysed at the subscale level in the validity analyses that follow; no single composite is used.

4. Construct Validity: Relations of the TAI-E with Related Constructs and Criterion Variables

Since the three-factor structure had already been established in the initial scale-development phase, this section uses the same sample to situate the TAI-E within its nomological network and to provide construct-validity evidence rather than test causal hypotheses. We examine (a) convergent and divergent relations with conceptually related constructs (self-reported reflective information processing, sense of coherence), (b) relations with criterion variables not used in scale construction (frequency of AI use, age), and (c) a behavioural check on reflective processing using a cognitive-reflection task. These analyses are exploratory and were not preregistered; all coefficients are interpreted as associations, not as evidence of causal or temporal ordering.

4.1. Method

4.1.1. Sample

The analyses reported in this section draw on the same sample described in Section 3.1.2 ($N = 412$ after listwise handling); no additional data were collected.

4.1.2. Measuring Instruments

To represent epistemic trust in generative AI, we used the three TAI-E subscales—Epistemic Authority Attribution, Perceived Epistemic Reliability, and Suspension of Critical Judgment—scored separately. To measure sense of coherence, we used the SOC-13 scale (Antonovsky 1987) in its validated Hungarian version (Balajti et al. 2007). Respondents rated their answers on a 7-point Likert scale. The scale's $\alpha = 0.779$. Principal component analysis identified three components (total explained variance: 52.94%), consistent with the theoretical structure; the use of the aggregate indicator is therefore supported.

To measure self-reported reflective engagement with generative AI, we used a 5-item, 6-point Likert scale ($\alpha = 0.747$; see Appendix B). The scale is a self-developed measurement tool based on the theoretical frameworks of reflective, analytical information processing and was developed as a result of targeted operationalization, with particular regard to dual-process models (e.g., Kahneman 2011; Evans and Stanovich 2013). EFA identified a unidimensional structure (explained variance: 39.61%; factor loadings: 0.323–0.738). The low factor loading of one item ($\lambda = 0.323$, communality = 0.104) warrants attention. However, due to the item's theoretical relevance and the brevity of the scale, we retained it in the analysis, as its removal would not have resulted in a significant improvement in reliability indices, while it would have reduced the construct's content coverage; therefore, we retained the item to preserve content validity; however, further refinement or retesting may be warranted in subsequent studies.

The frequency of AI use was measured as a single-item self-report variable on a 5-point ordinal scale (1 = never, 5 = daily). Age was considered as a continuous variable during the analysis.

As a behavioural benchmark for the self-report reflective-processing measure, the questionnaire also included the three classic items of the Cognitive Reflection Test (CRT; Frederick 2005)—the bat-and-ball, machines, and lily-pad problems—administered between January and March 2026 together with all other measures and scored as the proportion of correct responses. These items were part of the original data collection but were not analysed in the initial submission; the comparison reported in Sections 4.2.4 and 5.3 was added during revision in response to reviewer feedback, and is therefore exploratory and post hoc. It is interpreted as a convergent-validity check rather than as a test of a causal model.

Consistent with the conceptualization in Section 2.2, the scale measures a self-attributed disposition toward reflective processing rather than observed cognitive behaviour; this is taken into account when interpreting the findings.

4.2. Results

4.2.1. Descriptive Statistics

On the 6-point metric, Epistemic Authority Attribution fell below the scale midpoint ($M = 3.12$), Perceived Epistemic Reliability near it ($M = 3.58$), and Suspension of Critical Judgment—reverse-coded—above it ($M = 4.34$), indicating that respondents leaned toward critical engagement rather than uncritical acceptance. The dispersion was substantial on all three dimensions ($SD = 1.18$ – 1.30), consistent with international findings of ambivalent user attitudes toward AI (Gerlich 2024, 2025; Henrique and Santos 2025). Descriptive statistics for all main variables are summarized in Table 5.

Table 5. Descriptive statistics for the main variables.

Variable	N	Min	Max	M	SD	Scale Midpoint
F1: Epistemic authority attribution	412	1.00	6.00	3.12	1.27	3.50
F2: Perceived epistemic reliability	412	1.00	6.00	3.58	1.18	3.50
F3: Suspension of critical judgment	412	1.00	6.00	4.34	1.32	3.50
Self-reported reflective information processing	412	1.00	6.00	3.63	1.32	3.50
Sense of coherence (SOC-13)	412	1.00	6.38	4.52	0.85	4.00
Cognitive Reflection Test (CRT-3, prop. correct)	412	0	1	0.52	0.37	-
Age (years)	412	18	63	36.56	12.32	-

Note. No overall TAI-E composite is reported (see Section 3.2.3). F3 is reverse-coded, so a higher score reflects greater critical engagement (less suspension of critical judgment). CRT-3 = proportion of correct responses on the three classic items (Frederick 2005).

Self-reported reflective information processing also yielded a moderate mean score ($M = 3.63$, scale midpoint: 3.50), accompanied by considerable dispersion ($SD = 1.32$). Sense of coherence was above the scale midpoint ($M = 4.52$, scale midpoint: 4.00), and its smaller standard deviation ($SD = 0.85$) indicates a comparatively more homogeneous distribution.

4.2.2. Correlation Analysis

Self-reported reflective information processing showed the theoretically expected pattern across the three dimensions: it correlated positively with Epistemic Authority Attribution ($r = 0.35$, $p < 0.001$) and Perceived Epistemic Reliability ($r = 0.32$, $p < 0.001$), and negatively with Suspension of Critical Judgment ($r = -0.33$, $p < 0.001$). In other words, respondents who reported more reflective processing attributed somewhat more epistemic authority and reliability to AI while remaining more critically engaged—a configuration inconsistent with a simple “more reflection → less trust” account. Sense of coherence was essentially unrelated to the trust dimensions ($|r| \leq 0.15$), supporting its divergent status. The full correlation matrix is presented in Table 6.

Table 6. Pearson correlations among the TAI-E dimensions and related variables ($N = 412$).

Variable	1	2	3	4	5	6	7
1. Authority	–						
2. Reliability	0.61 **	–					
3. Suspension (R)	−0.63 **	−0.48 **	–				
4. RIP	0.35 **	0.32 **	−0.33 **	–			
5. SOC	0.08	0.15 **	0.10 *	0.13 **	–		
6. CRT-3	0.11 *	0.05	0.00	0.14 **	0.05	–	
7. Age	−0.01	0.04	0.15 **	0.18 **	0.42 **	0.12 *	–
8. AI-use freq.	−0.23 **	−0.15 **	0.07	−0.01	0.07	−0.11 *	0.02

Note. ** $p < 0.01$, * $p < 0.05$ (two-tailed). Suspension of Critical Judgment is reverse-coded (R). RIP = self-reported Reflective Information Processing; CRT-3 = three-item Cognitive Reflection Test (proportion correct).

Frequency of AI use related negatively to Authority ($r = -0.23$, $p < 0.001$) and Reliability ($r = -0.15$, $p = 0.002$) but not to Suspension, indicating that more frequent users attributed less epistemic authority to AI. Age was unrelated to Authority and Reliability but correlated modestly with greater critical engagement (Suspension $r = 0.15$, $p = 0.002$).

4.2.3. Criterion-Related Validity

To assess criterion-related validity, we examined how the three TAI-E dimensions relate to an external variable not used in scale construction—self-reported frequency of AI use—using linear regression with HC3 heteroscedasticity-robust standard errors. Greater frequency of AI use was predicted by lower Epistemic Authority Attribution ($\beta = -0.23$,

$p < 0.001$) and lower Perceived Epistemic Reliability ($\beta = -0.15$, $p = 0.002$), whereas Suspension of Critical Judgment was unrelated to use ($\beta = 0.07$, $p = 0.141$). Age showed the complementary pattern: it was unrelated to Authority and Reliability but predicted greater critical engagement (Suspension $\beta = 0.15$, $p = 0.002$). These associations are modest in magnitude but theoretically coherent: more frequent users attribute less epistemic authority to AI, consistent with an exposure-based moderation of trust (Hoff and Bashir 2015; Logg et al. 2019). The regression results for the two external criteria are reported in Table 7.

Table 7. Criterion-related validity: TAI-E dimensions predicting frequency of AI use and age.

Predictor	Outcome: AI-Use β	p	Outcome: Age β	p
Epistemic Authority Attribution	−0.23	<0.001	−0.01	0.880
Perceived Epistemic Reliability	−0.15	0.002	0.04	0.452
Suspension of Critical Judgment	0.07	0.141	0.15	0.002

Note. Standardized coefficients (β); HC3 robust SE; N = 412. AI-use and age are external criteria not used in scale construction.

4.2.4. Self-Reported Versus Behavioural Reflection

Because the RIP reflective-processing measure is a self-report, we tested whether it tracks an objective behavioural index of reflection—the three-item Cognitive Reflection Test (CRT; Frederick 2005), administered as part of the same questionnaire (see Section 4.1.2). Self-reported reflective processing correlated only weakly with behavioural CRT performance ($r = 0.14$, $p = 0.005$), sharing roughly 2% of variance. More importantly, the two indicators differed in their relations to epistemic trust. Self-reported reflective processing was moderately associated with the trust dimensions (Authority $r = 0.35$; Reliability $r = 0.32$; Suspension $r = -0.33$, all $p < 0.001$), whereas behavioural CRT performance was at most weakly related to Authority ($r = 0.11$, $p = 0.023$) and unrelated to Reliability ($r = 0.05$) and Suspension ($r = 0.00$). Thus, the association between “reflectiveness” and epistemic trust in AI appears to be carried primarily by self-perceived reflectiveness rather than by demonstrated cognitive reflection. This dissociation is itself informative: it suggests that the self-report measure captures, at least in part, an epistemic self-image rather than reflective capacity as measured behaviourally, and it cautions against reading the reflection–trust association as evidence that more reflective processors trust AI more (see Section 5). Table 8 summarizes the relations of self-reported and behavioural reflection to the TAI-E dimensions.

Table 8. Self-reported (RIP) versus behavioural (CRT-3) reflection in relation to the TAI-E dimensions.

	Authority	Reliability	Suspension (R)	RIP
Self-reported RIP	0.35 **	0.32 **	−0.33 **	—
Behavioural CRT-3	0.11 *	0.05	0.00	0.14 **

Note. ** $p < 0.01$, * $p < 0.05$ (two-tailed); N = 412. RIP = self-reported Reflective Information Processing; CRT-3 = three-item Cognitive Reflection Test (proportion correct). Suspension is reverse-coded (R).

Taken together, the four sets of analyses provide convergent construct-validity evidence for the TAI-E: the three dimensions are empirically distinct from one another and from self-reported reflective information processing, relate coherently to external criteria, and behave as a configuration of source legitimation and critical engagement rather than as a single trust quantity. The divergence between self-reported and behavioural reflection further specifies what the reflection measure captures. These patterns are interpreted in the Discussion.

5. Discussion

The present study set out to operationalize epistemic trust in generative AI as a source of knowledge and to situate the resulting instrument within its nomological network. Three observations organize the discussion: epistemic trust in generative AI appears to be multidimensional and structurally bipolar rather than a single more-or-less continuum; self-reported reflective information processing relates to this trust configurally rather than monotonically; and the relationship between “reflectiveness” and trust is carried primarily by self-perceived rather than behaviourally demonstrated reflection. Because the design is cross-sectional and relies on a single self-report sample, all relationships are interpreted as associations and as construct-validity evidence, not as causal or temporal claims.

5.1. A Multidimensional, Bipolar Structure of Epistemic Trust

The three TAI-E dimensions—attribution of epistemic authority, perceived epistemic reliability, and suspension of critical judgment—replicated across a calibration–validation split, yet they did not cohere into a single composite: the inter-factor matrix was not positive-manifest, with suspension of critical judgment correlating negatively with both authority attribution ($r = -0.63$) and perceived reliability ($r = -0.48$), and a 12-item total yielding poor internal consistency ($\alpha = 0.47$). Rather than indicating a psychometric defect, this pattern may be theoretically informative. It suggests that ascribing epistemic authority and reliability to a generative system and suspending one’s critical judgment toward it are not simply two ends of a single dimension but partly opposed epistemic stances that can vary independently. This interpretation is consistent with accounts in social epistemology in which acceptance of a source and continued vigilance toward it can coexist rather than exclude one another (Origgi 2008; Sperber et al. 2010; Neuberger et al. 2025): an epistemically competent user may grant a source substantial authority while retaining the disposition to verify. Treating epistemic trust in AI as a single quantity—“more” versus “less” trust—therefore risks obscuring the configuration that may matter most for epistemic quality, namely how authority attribution and critical engagement are jointly held. The instrument’s value may lie precisely in preserving these dimensions as analytically distinct constructs.

5.2. Reflective Processing Relates to Epistemic Trust Configurally, Not Monotonically

Self-reported reflective information processing did not map onto epistemic trust in a simple direction. It was positively associated with authority attribution ($r = 0.35$) and perceived reliability ($r = 0.32$), and negatively associated with suspension of critical judgment ($r = -0.33$). In other words, respondents reporting more reflective processing attributed somewhat greater epistemic authority and reliability to generative AI while remaining more critically engaged. This configuration does not fit either of two simplified narratives that frequently appear in the surrounding literature: the assumption that analytic engagement necessarily reduces trust (Alon-Barkat and Busuioc 2023; Romeo and Conti 2026; Stanovich et al. 2016), or the assumption that trust primarily reflects uncritical acceptance (Davis 1989; Hoff and Bashir 2015). It is more consistent with the broader logic of the trust-calibration literature, in which the depth of user processing is treated as a condition for appropriately differentiated reliance (Lee and See 2004; Wischnewski et al. 2023; Kaplan et al. 2023). However, the present study does not measure calibration in that literature’s strict sense, because system performance was not assessed. The supported claim is therefore narrower: a more reflective self-presentation is associated with a more differentiated epistemic stance, combining higher attributed authority with greater, not lesser, critical engagement.

5.3. The Self-Report Problem: Reflectiveness as Epistemic Self-Image

Perhaps the most consequential finding is interpretive rather than confirmatory. Self-reported reflective processing correlated only weakly with behavioural cognitive reflection as measured by the Cognitive Reflection Test ($r = 0.14$), sharing roughly two percent of variance, and the two indices diverged in their relation to epistemic trust. Whereas self-reported reflectiveness was moderately associated with the trust dimensions ($|r| = 0.32\text{--}0.35$), behavioural CRT performance was at most weakly related to authority attribution ($r = 0.11$) and unrelated to perceived reliability and suspension of critical judgment. The association between “reflectiveness” and epistemic trust therefore appears to be carried largely by self-perceived reflectiveness, not by demonstrated cognitive reflection. This dissociation speaks to a longstanding caution in the dual-process literature, where the gap between self-reported analytic dispositions and reflection-sensitive performance is well documented (Frederick 2005; Stanovich et al. 2016), and to philosophical debates questioning whether the deliberate, “System 2” image of reflective judgment adequately captures how beliefs are actually formed and revised (Evans and Stanovich 2013; Hieronymi 2009; Vierkant 2022). A plausible reading is that the self-report captures, in part, an epistemic self-image—a person’s view of themselves as a careful processor of information—rather than reflective capacity as behaviourally indexed. This warrants caution against interpreting the reflection–trust association as evidence that individuals who actually process information more reflectively extend more epistemic trust to AI. The data are also compatible with the possibility that confidence in one’s own discernment accompanies a readiness to credit AI as a knowledge source. Disentangling these accounts would require behavioural or longitudinal designs.

5.4. Criterion Relations: Exposure, Age, and the Divergence of General Resources

The TAI-E dimensions related to external criteria in theoretically coherent ways. More frequent AI use was associated with lower attribution of epistemic authority ($\beta = -0.23$) and lower perceived reliability ($\beta = -0.15$) but was unrelated to suspension of critical judgment. This pattern is consistent with exposure-based accounts in which familiarity with a system’s failures may temper the authority granted to it (Logg et al. 2019; Gerlich 2024, 2025). Age, by contrast, was unrelated to authority and reliability and was associated only with greater critical engagement (suspension $r = 0.15$). This complicates the assumption of a uniform age-related decline in technological trust (Hargittai 2010): in these data, older respondents did not attribute less authority or reliability to generative AI; if anything, they appeared somewhat more critically engaged. Finally, sense of coherence was essentially unrelated to the trust dimensions ($|r| \leq 0.15$), reinforcing its status as a divergent, general-purpose psychological resource. This is consistent with evidence that the salutogenic value of sense of coherence is most evident in general stress and life-management contexts (Antonovsky 1987; Stoyanova and Stoyanov 2021) and adds specificity: a global sense of comprehensibility and manageability does not necessarily translate into a particular epistemic stance toward AI as a knowledge source.

5.5. Implications for the Conceptualization and Measurement of Trust in AI

Taken together, the results support treating trust in generative AI as a specifically epistemic construct, distinct from generalized attitudes and from mere behavioural reliance. The dominant tradition in the field has often framed trust as a precursor to technology acceptance (Davis 1989; Mayer et al. 1995; Choung et al. 2023), where the function of trust is adoption rather than the characterization of knowledge quality. Philosophical work questions whether artificial systems are appropriate objects of trust in the full interpersonal sense at all, suggesting that what users extend is often better described as reliance (Ryan

2020; Bryson 2018; Dorsch and Deroy 2024). The TAI-E does not adjudicate that debate; instead, it operationalizes the narrower and more tractable notion of epistemic trust—the legitimacy a user grants a source and the acceptance of its outputs as warranting belief (Origgi 2008; Alvarado 2023b; Neuberger et al. 2025). The instrument's bipolar structure, together with the divergence between self-reported and behavioural reflection, suggests that epistemic trust in AI may be best studied not as a quantity to be maximized but as a configuration of source legitimation and critical engagement whose quality depends on how the two are held together.

6. Limitations

Cross-sectional design and the question of causality: The cross-sectional design identifies patterns of association between antecedent variables and trust in AI, not verified causal mechanisms. Any indirect-effect or sequence-based interpretation remains statistical and exploratory and does not in itself establish temporal ordering (MacKinnon et al. 2007). A longitudinal measurement design would be necessary for a causal interpretation of an age → self-reported reflective processing → AI trust chain. This limitation does not undermine the descriptive and hypothesis-generating value of the results, but it requires caution in formulating implications.

Sample composition and limitations of external validity: Data collection was limited to distance-learning students in higher education. The relative overrepresentation of engineering and IT fields (43.7%) and high AI exposure (95%) suggest a population that is more technologically oriented and experienced with AI than many average adult samples. The sample size ($N = 412$) is adequate for the analyses conducted, but the convenience-sampling strategy limits generalizability. Probability-based and multi-context sampling would be desirable in future research.

Self-report measurement and common method bias: All variables were measured using a self-report format within the same questionnaire, which may generate common method bias (Podsakoff et al. 2003). The self-reported reflective-processing measure is particularly susceptible to this problem, as the discrepancy between self-report and actual cognitive behaviour has been extensively documented (Stanovich et al. 2016). This concern does not invalidate the results. The present study took a first step toward addressing it by including a behavioural cognitive-reflection measure (CRT-3) alongside the self-report; the divergence between the two (Section 5.3) reinforces the concern and indicates that future research should privilege behavioural or performance-based indicators of reflection where possible.

The need for cross-validation of the TAI-E instrument: Although the factor structure was cross-validated on a calibration-validation split within a single dataset, the two subsamples are not equivalent to independently collected samples. Replication on a separate sample therefore remains necessary. The CFA RMSEA (0.085; validation-subsample upper 90% CI approaching 0.10) further justifies refinement and independent replication. The expert panel also did not include a philosopher specializing in epistemology, philosophy of technology, or AI ethics. Given that epistemic trust is a fundamentally epistemological construct, future content-validation work would benefit from such expertise to refine the conceptual boundaries of the dimensions further.

7. Conclusions

This study introduced and validated the Trust in AI-Epistemic Scale (TAI-E), an instrument that operationalizes epistemic trust in generative AI as the legitimacy users grant it as a source of knowledge. The scale recovered a replicable three-factor structure—epistemic authority attribution, perceived epistemic reliability, and suspension of critical

judgment—whose dimensions are not reducible to a single composite but appear to form opposing poles of source legitimation and critical engagement. Beyond the instrument itself, the study offers a cautionary methodological lesson: the association between reflective processing and epistemic trust was driven by self-reported, rather than behaviourally demonstrated, reflection. Self-perceived reflectiveness should therefore not be mistaken for reflective capacity. The TAI-E may provide a psychometrically grounded basis for future research on AI literacy, epistemic competence, and the conditions of well-founded reliance on AI-generated information, particularly when used with representative samples, behavioural indicators, and longitudinal designs.

Funding: This research received no external funding. The APC was funded by the University of Dunaújváros.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Scientific Council of the University of Dunaújváros (protocol code DUE-KB_5_2026_ENG; 30 March 2026).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical and data protection restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. The Trust in AI-Epistemic Scale (TAI-E) Items.

Item	Dimension	Item Text (EN)	Tétel Szövege (HU)	Coding
1	Epistemic Authority Attribution	I consider AI a more competent source of knowledge than many human experts.	Az AI-t kompetensebb tudásforrásnak tekintem sok emberi szakértőnél.	—
2	Epistemic Authority Attribution	I tend to accept AI's opinion as a decisive argument in professional discussions.	Hajlamos vagyok az AI véleményét döntő érvként elfogadni szakmai vitákban.	—
3	Epistemic Authority Attribution	AI-generated knowledge appears to me as an “objective truth.”	Az AI tudása számomra „objektív igazságként” jelenik meg.	—
4	Epistemic Authority Attribution	I view AI as a kind of “universal encyclopedia” whose authority rarely needs to be questioned.	Úgy tekintek az AI-ra, mint egy „univerzális enciklopédiára”, amelynek ritkán kell megkérdőjelezni a tekintélyét.	—
5	Perceived Epistemic Reliability	AI responses are consistently accurate and free from factual errors.	Az AI válaszai következetesen pontosak és mentesek a ténybeli tévedésektől.	—
6	Perceived Epistemic Reliability	I trust that AI provides the most relevant and up-to-date information.	Megbízom benne, hogy az AI a legrelevánsabb és legfrissebb tudást közvetíti felém.	—
7	Perceived Epistemic Reliability	I consider AI a stable reference point in my information-seeking process.	Az AI-t stabil viszonyítási pontnak tekintem a tájékozódásom során.	—
8	Perceived Epistemic Reliability	I trust that the technical precision of AI guarantees the accuracy of the information provided.	Bízom abban, hogy az AI technikai precizitása garantálja a közölt adatok pontosságát.	—

Table A1. Cont.

Item	Dimension	Item Text (EN)	Tétel Szövege (HU)	Coding
9	Suspension of critical judgment	I often incorporate AI responses into my work or thinking without verification.	Gyakran ellenőrzés nélkül illeszttem be az AI válaszait a saját munkámba vagy gondolataimba.	*R
10	Suspension of critical judgment	When AI responds confidently, I do not feel the need to critically evaluate the source.	Ha az AI magabiztos stílusban válaszol, nem érzem szükségét a forráskritikának.	*R
11	Suspension of critical judgment	I rarely question conclusions drawn by AI.	Ritkán kérdőjelezem meg az AI által levont következtetéseket.	*R
12	Suspension of critical judgment	I do not consider it necessary to verify AI-provided information using other sources.	Nem tartom szükségesnek, hogy az AI által közölt információkat más forrásból is ellenőrizsem.	*R

Note: All items were rated on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree; 1 = egyáltalán nem értek egyet, 6 = teljes mértékben egyetértek). *R = reverse-coded items (fordítottan kódolt tételek). The label “Trust in AI” is retained for brevity; in the present study the items concern generative AI, consistent with the questionnaire’s generative-AI screening and usage items (see Section 3.1.3).

Appendix B

Table A2. Self-Reported Reflective Information Processing (RIP) Items.

Item	Item Text (EN)	Tétel Szövege (HU)	Coding
1	I deliberately check AI’s responses before accepting them.	Az AI válaszait tudatosan ellenőrzöm, mielőtt elfogadnám őket.	—
2	I often verify the information AI provides using other sources.	Gyakran utánanézek az AI által adott információknak más forrásokból.	—
3	I do not automatically accept AI’s first answer.	Nem fogadom el automatikusan az AI első válaszát.	—
4	When AI’s response seems overly confident, I become suspicious.	Ha az AI válasza túl magabiztosnak tűnik, gyanakodni kezdek.	—
5	I consider multiple sources when using information provided by AI.	Több forrást is figyelembe veszek, amikor AI által szolgáltatott információkat használok.	—

Note. All RIP items were rated on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree; 1 = egyáltalán nem értek egyet, 6 = teljes mértékben egyetértek). The scale is a self-report measure of a dispositional tendency toward analytic, source-sensitive evaluation of generative AI outputs (see Section 2.2); it indexes self-attributed reflectiveness rather than observed processing behaviour.

References

- Alon-Barkat, Saar, and Madalina Busuioc. 2023. Human–AI Interactions in Public Sector Decision Making: ‘Automation Bias’ and ‘Selective Adherence’ to Algorithmic Advice. *Journal of Public Administration Research and Theory* 33: 153–69. [\[CrossRef\]](#)
- Alvarado, Ramon. 2023a. AI as an Epistemic Technology. *Science and Engineering Ethics* 29: 32. [\[CrossRef\]](#) [\[PubMed\]](#)
- Alvarado, Ramon. 2023b. What Kind of Trust Does AI Deserve, If Any? *AI and Ethics* 3: 1169–83. [\[CrossRef\]](#)
- Andrejevic, Mark. 2013. *Infoglut: How Too Much Information Is Changing the Way We Think and Know*. Abingdon: Routledge.
- Antonovsky, Aaron. 1987. *Unraveling the Mystery of Health: How People Manage Stress and Stay Well*. San Francisco: Jossey-Bass.
- Antonovsky, Aaron. 1993. The Structure and Properties of the Sense of Coherence Scale. *Social Science and Medicine* 36: 725–33. [\[CrossRef\]](#) [\[PubMed\]](#)
- Arnold, Miriam, Mascha Goldschmitt, and Thomas Rigotti. 2023. Dealing with Information Overload: A Comprehensive Review. *Frontiers in Psychology* 14: 1122200. [\[CrossRef\]](#) [\[PubMed\]](#)
- Balajti, Ilona, Zoltán Vokó, Róza Ádány, and Karolina Kósa. 2007. A koherenciaérzés mérőeszközének hazai adaptálása. *Mentálhigiéné és Pszichoszomatika* 8: 147–61. [\[CrossRef\]](#)
- Bansal, Gagan, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7: 2–11. [\[CrossRef\]](#)

- Bartsch, Anne, Marie-Louise Mares, Johanna Schindler, Jessica Kühn, and Ina Krack. 2024. Trust but Verify? A Social Epistemology Framework of Knowledge Acquisition and Verification Practices for Fictional Entertainment. *Human Communication Research* 50: 194–207. [CrossRef]
- Bawden, David, and Lyn Robinson. 2009. The Dark Side of Information: Overload, Anxiety, and Other Paradoxes and Pathologies. *Journal of Information Science* 35: 180–91. [CrossRef]
- Bryson, Joanna J. 2018. *AI & Global Governance: No One Should Trust AI*. New York: United Nations University, Centre for Policy Research. Available online: <https://unu.edu/cpr/blog-post/ai-global-governance-no-one-should-trust-ai> (accessed on 16 February 2026).
- Campbell, Donald Thomas, and Donald Winslow Fiske. 1959. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin* 56: 81–105. [CrossRef]
- Choung, Hyesun, Prabu David, and Arun Ross. 2023. Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction* 39: 1727–39. [CrossRef]
- Clark, Lee Anna, and David Watson. 1995. Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment* 7: 309–19. [CrossRef]
- Couldry, Nick, and Andreas Hepp. 2017. *The Mediated Construction of Reality*. Cambridge: Polity Press.
- Cronbach, Lee Joseph, and Paul Everett Meehl. 1955. Construct Validity in Psychological Tests. *Psychological Bulletin* 52: 281–302. [CrossRef] [PubMed]
- Cummings, Missy L. 2004. Automation Bias in Intelligent Time Critical Decision Support Systems. Paper presented at AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA, September 20–22; pp. 557–62. [CrossRef]
- Davis, Fred D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13: 319–40. [CrossRef] [PubMed]
- Deroy, Ophelia. 2023. The Ethics of Terminology: Can We Use Human Terms to Describe AI? *Topoi* 42: 881–89. [CrossRef]
- Dorsch, John, and Ophelia Deroy. 2024. Quasi-Metacognitive Machines: Why We Don't Need Morally Trustworthy AI and Communicating Reliability Is Enough. *Philosophy & Technology* 37: 62. [CrossRef]
- Doshi-Velez, Finale, and Been Kim. 2017. Towards a Rigorous Science of Interpretable Machine Learning. *arXiv* arXiv:1702.08608. [CrossRef]
- Eppler, Martin J., and Jeanne Mengis. 2004. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society* 20: 325–44. [CrossRef]
- Evans, Jonathan St. B. T., and Keith E. Stanovich. 2013. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science* 8: 223–41. [CrossRef] [PubMed]
- Flake, Jessica Kay, Jolynn Pek, and Eric Hehman. 2017. Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science* 8: 370–78. [CrossRef]
- Frederick, Shane. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19: 25–42. [CrossRef]
- George, Darren, and Paul Mallery. 2005. *SPSS for Windows Step by Step: A Simple Guide and Reference*, 4th ed. Boston: Allyn & Bacon.
- Gerlich, Michael. 2024. The Societal Perceptions of Artificial Intelligence and Its Impact on Marketing in 2024. Paper presented at 5th World Conference on Business, Management, Finance, Economics and Marketing, Vienna, Austria, May 16–17. [CrossRef]
- Gerlich, Michael. 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* 15: 6. [CrossRef]
- Glikson, Ella, and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals* 14: 627–60. [CrossRef]
- Hargittai, Eszter. 2010. Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the 'Net Generation'. *Sociological Inquiry* 80: 92–113. [CrossRef]
- Henrique, Bruno M., and Eugene Santos, Jr. 2025. Dynamic Trust Calibration Using Contextual Bandits. *arXiv* arXiv:2509.23497. [CrossRef]
- Henseler, Jörg, Christian M. Ringle, and Marko Sarstedt. 2015. A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling. *Journal of the Academy of Marketing Science* 43: 115–35. [CrossRef]
- Hepp, Andreas. 2020. *Deep Mediatization*. Abingdon: Routledge.
- Hieronymi, Pamela. 2009. Two Kinds of Agency. In *Mental Actions*. Edited by Lucy O'Brien and Matthew Soteriou. Oxford: Oxford University Press, pp. 138–62. [CrossRef]
- Hoff, Kevin A., and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57: 407–34. [CrossRef] [PubMed]
- Jones, Dominic Zaun Eu, and Eshwar Chandrasekharan. 2024. Measuring Epistemic Trust: Towards a New Lens for Democratic Legitimacy, Misinformation, and Echo Chambers. *Proceedings of the ACM on Human-Computer Interaction* 8: 462. [CrossRef]
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kaplan, Alexandra D., Theresa T. Kessler, Jonathan C. Brill, and Peter A. Hancock. 2023. Trust in Artificial Intelligence: Meta-Analytic Findings. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 65: 337–59. [CrossRef] [PubMed]

- Kizilcec, René F. 2016. How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 2390–95. [\[CrossRef\]](#)
- Lee, John D., and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46: 50–80. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lewandowsky, Stephan, Ullrich K. H. Ecker, and John Cook. 2017. Beyond Misinformation: Understanding and Coping with the ‘Post-Truth’ Era. *Journal of Applied Research in Memory and Cognition* 6: 353–69. [\[CrossRef\]](#)
- Logg, Jennifer M., Julia A. Minson, and Don A. Moore. 2019. Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes* 151: 90–103. [\[CrossRef\]](#)
- Lorenz-Spreen, Philipp, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann. 2019. Accelerating Dynamics of Collective Attention. *Nature Communications* 10: 1759. [\[CrossRef\]](#) [\[PubMed\]](#)
- MacKinnon, David P., Amanda J. Fairchild, and Matthew S. Fritz. 2007. Mediation Analysis. *Annual Review of Psychology* 58: 593–614. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mayer, Roger C., James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *Academy of Management Review* 20: 709–34. [\[CrossRef\]](#)
- Miller, Tim. 2019. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence* 267: 1–38. [\[CrossRef\]](#)
- Mittelmark, Maurice B., Shifra Sagy, Monica Eriksson, Georg F. Bauer, Jürgen M. Pelikan, Bengt Lindström, and Geir Arild Espnes, eds. 2022. *The Handbook of Salutogenesis*, 2nd ed. Cham: Springer. [\[CrossRef\]](#) [\[PubMed\]](#)
- Moko, Anasuodei, Maudlyn Victor-Ikogh, and Biobele Okardi. 2023. Information Overload: A Conceptual Model. *European Journal of Computer Science and Information Technology* 11: 19–29. [\[CrossRef\]](#)
- Moll, Maximilian, and John Dorsch. 2025. A Systematic Review of Human-Centered Explainability in Reinforcement Learning: Transferring the RCC Framework to Support Epistemic Trustworthiness. *Human-Intelligent Systems Integration* 7: 239–48. [\[CrossRef\]](#)
- Neuberger, Christoph, Sabrina Langenohl, and Christoph Nuernbergk. 2025. Epistemic Authority in the Digital Public Sphere: An Integrative Conceptual Framework and Research Agenda. *Communication Theory* 35: 37–56. [\[CrossRef\]](#)
- Origi, Gloria. 2008. Trust, Authority and Epistemic Responsibility. *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia* 23: 35–44. [\[CrossRef\]](#)
- Parasuraman, Raja, and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39: 230–53. [\[CrossRef\]](#)
- Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology* 88: 879–903. [\[CrossRef\]](#) [\[PubMed\]](#)
- Reinhardt, Karoline, and Oliver Buchholz. 2025. XAI: On Explainability and the Obligation to Explain. *Digital Society* 4: 36. [\[CrossRef\]](#)
- Romeo, Giuseppe, and Daniela Conti. 2026. Exploring Automation Bias in Human–AI Collaboration: A Review and Implications for Explainable AI. *AI & Society* 41: 259–78. [\[CrossRef\]](#)
- Ryan, Mark. 2020. In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics* 26: 2749–67. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sahut, Gilles, and André Tricot. 2017. Wikipedia: An Opportunity to Rethink the Links between Sources’ Credibility, Trust, and Authority. *First Monday* 22: 11. [\[CrossRef\]](#)
- Shin, Donghee. 2025. Automating Epistemology: How AI Reconfigures Truth, Authority, and Verification. *AI & Society* 41: 1553–59. [\[CrossRef\]](#)
- Shneiderman, Ben. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36: 495–504. [\[CrossRef\]](#)
- Siau, Keng, and Weiyu Wang. 2018. Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal* 31: 47–53.
- Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. 2010. Epistemic Vigilance. *Mind & Language* 25: 359–93. [\[CrossRef\]](#)
- Stanovich, Keith E., Richard F. West, and Maggie E. Toplak. 2016. *The Rationality Quotient: Toward a Test of Rational Thinking*. Cambridge: Boston Review. [\[CrossRef\]](#)
- Stoyanova, Kristina, and Drozdostoy Stoyanov. 2021. Sense of Coherence and Burnout in Healthcare Professionals in the COVID-19 Era. *Frontiers in Psychiatry* 12: 709587. [\[CrossRef\]](#) [\[PubMed\]](#)
- Thiebes, Scott, Sebastian Lins, and Ali Sunyaev. 2021. Trustworthy Artificial Intelligence. *Electronic Markets* 31: 447–64. [\[CrossRef\]](#)
- Vierkant, Tillmann. 2022. *The Tinkering Mind: Agency, Cognition, and the Extended Mind*. Oxford: Oxford University Press.
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. The Spread of True and False News Online. *Science* 359: 1146–51. [\[CrossRef\]](#) [\[PubMed\]](#)

- Wischnewski, Magdalena, Nicole Krämer, and Emmanuel Müller. 2023. Measuring and Understanding Trust Calibrations for Automated Systems: A Survey of the State-of-the-Art and Future Directions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. New York: Association for Computing Machinery, pp. 1–16. [[CrossRef](#)]
- Zednik, Carlos. 2021. Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy & Technology* 34: 265–88. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.