Do Children Trust AI, and Should They? Designing and Validating a Child-Centred K-AI Trust Scale for Intelligent Systems

Grazia Ragone
University of Bari Aldo Moro
Bari, Italy
grazia.ragone@uniba.it

Paolo Buono
University of Bari Aldo Moro
Bari, Italy
paolo.buono@uniba.it

Judith Good
University of Amsterdam
Amsterdam, Netherlands
j.a.good@uva.nl

Rosa Lanzilotti
University of Bari Aldo Moro
Bari, Italy
rosa.lanzilotti@uniba.it

Abstract

Most trust metrics for intelligent systems are developed for adults, relying on complex reasoning and language that do not align with children's developmental stages. As intelligent systems increasingly engage with young users, evaluating trust in child-AI interaction has become an urgent concern in HCI. In this paper, we present the iterative refinement and validation of the K-AI Trust Questionnaire, a child-centred instrument that integrates dispositional and situational trust components grounded in child-rights principles. Dispositional trust is captured through a child-adapted Propensity to Trust Technology (PTT), while situational trust is assessed through post-interaction items reflecting children's experience with AI. Starting with a sample of 289 children, we conducted psychometric analyses and exploratory testing, culminating in a confirmatory factor analysis on a subsample of 85 children. Results supported a unidimensional structure consistent with the PTT, and highlighted the limitations of adult-oriented scales, underscoring the need for developmentally appropriate tools for trustworthy child-AI design.

CCS Concepts

• **Human-centered computing** → *Empirical studies in HCI*; • **Social and professional topics** → **Children**.

Keywords

Trust in AI, Psychometric Validation, Child–Computer Interaction, Questionnaire Design, Human–AI Interaction, Trust Calibration, Educational Technology

ACM Reference Format:
Grazia Ragone, Paolo Buono, Judith Good, and Rosa Lanzilotti. 2026. Do Children Trust AI, and Should They? Designing and Validating a Child-Centred K-AI Trust Scale for Intelligent Systems. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3772318.3790765

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3790765

1 Introduction

As Artificial Intelligence (AI) technologies become increasingly embedded in everyday human activities, the capacity to systematically evaluate their trustworthiness has emerged as a key challenge for the Human-Computer Interaction (HCI) community. In particular, the deployment of AI systems in child-centred contexts, such as educational platforms and conversational agents, raises complex questions about how trust is established, calibrated, and maintained. Despite growing interest in ethical AI design, current evaluation frameworks [36, 44, 53] often rely on generalised usability metrics or adult-centric instruments that fail to account for the developmental, cognitive, and interpretive differences that shape children's experiences with intelligent systems.

Trust can be defined as "the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability" [33]. In this study, we conceptualise trust as the *perceived trustworthiness* of an AI system from the child's perspective. While trustworthiness denotes the system's actual qualities, such as competence, reliability, fairness, and transparency, trust reflects how these qualities are perceived and acted upon by the child. When not aligned, trust may lead to overtrust, where users depend too much on automation, or undertrust, where users stop engaging with it even though the system is actually helping them [48]. This also happens in child–AI interaction. For example, children may overtrust conversational agents, taking incorrect answers as true, or projecting human traits onto the system [73]. They may also stop using AI tools after a single error occurs, even when they are contributing to their learning [34]. These dynamics emphasise a core challenge in Child–Computer Interaction (CCI): intelligent systems can empower learning and exploration, but only when trust is appropriately calibrated.

These dynamics are not limited to child-AI interactions but are rooted in broader discussions of automation in HCI. Automation is the "execution by a machine agent (usually a computer) of a function that was previously carried out by a human" [48]. Within this definition, conversational agents, interactive automated systems powered by Large Language Models (LLMs) or simpler decision trees [13], process input, decide on output, and act without human mediation.

Children encounter such systems every day at home, in classrooms, and through voice assistants and educational platforms. The understanding of how trust in such technologies is built ensures that they support critical thinking rather than blind reliance. Children are often unable to distinguish between a system that appears intelligent and one that is truly reliable and responsive to their needs.

Although prior research has identified individual-level factors that influence trust in automation, such as self-confidence, personality, and the propensity to trust machines [43, 61], these studies have primarily targeted adult users, overlooking developmental differences in how children interact with and interpret intelligent systems. Much of this literature frames trust as a dispositional trait rather than a situated, relational construct, limiting our understanding of how it is co-constructed through social interaction. Trust is a dynamic process shaped by expectations, experiences, and interpretations deeply influenced by developmental stage and sociocultural context [60, 62], which is particularly relevant in the design of child-AI interaction. We therefore adopt a two-level framing of trust in child–AI interaction: a dispositional baseline, captured through a child-adapted Propensity to Trust Technology (PTT), and a situational, post-interaction judgment captured with our new K-AI Trust scale.

Emerging policy frameworks such as the EU AI Act [37] call for principles such as transparency, fairness, safety, and human agency as the cornerstones of trustworthy AI. However, these principles are rarely operationalised in empirical HCI studies, especially in systems designed for, or used by, children. A critical gap remains between high-level regulatory intentions and the concrete tools available to evaluate how children perceive and engage with AI systems in real-world settings [52]. Recent empirical work identifies five core principles: respect for autonomy, transparency, safety, fairness, and privacy, that should guide the design, development, and evaluation of AI systems [8]. We argue that these principles must be explicitly extended to child-AI interaction. Thus, child-AI design should be grounded in participatory, rights-aligned frameworks.

We suggest moving toward a mutual symbiosis model [75] in child-AI interaction, pointing to the redefinition of trust not simply as user compliance or technical reliability, but as a dynamic relationship, actively shaped by both sides. In this perspective, human and AI systems mutually support and adapt to one another, with shared understanding and goals that neither could achieve alone [38]. This vision involves children's gradual exploration of the AI system's affordances and boundaries, while the system complies with principles such as transparency, fairness, participation and responsiveness to the child's developmental needs. In this context, trust becomes part of an engaging, reflective and critical process, rather than passive dependence.

In this article, we introduce a novel questionnaire, *K-AI Trust*, specifically designed to assess children's trust in conversational AI systems after direct interaction. It was developed through a post-interaction study with children, drawing on integrated models of PTT, child-centred ethics, and empirical HCI research [1, 45]. Our contribution is a theoretically robust and developmentally informed tool that captures how children conceptualise and express trust in intelligent systems. Grounded in psychometric theory and aligned with international child rights and regulatory principles [1, 18, 46], the proposed questionnaire bridges the gap between abstract ethical principles and real-world child-AI interactions by operationalising constructs such as fairness, transparency, control, and responsiveness. In doing so, this work provides the HCI community with a robust, developmentally aligned tool for evaluating trust in child-AI interaction, thereby supporting the ethical and inclusive design of intelligent systems for children. In line with this framing, Study 1 establishes children's dispositional baseline using PTT, while Studies 2 and 3 design and validate *K-AI Trust* to measure situational trust after direct interaction with an AI system. This sequence clarifies how our contribution moves from trait to state, addressing a gap in child-centred HCI measures.

We articulate our investigation around the following research questions:

RQ1: Which dimensions best capture the trust construct when creating a child-centred scale?

RQ2: How do children interpret and respond to items related to ethical AI principles (e.g., consent, data use) within a trust questionnaire context?

To address these questions, we employed a mixed-methods approach, combining psychometric validation, exploratory factor analysis, and qualitative analysis of children's post-interaction data.

2 Related work

Trust is a multifaceted construct, central to human engagement with AI systems [10, 20, 51] and a core motivation for explainable AI (XAI). However, it remains conceptually fragmented and empirically underdefined [26, 69]. Different theoretical approaches use overlapping terms, including:

- dispositional trust (or propensity to trust): a trait-like baseline tendency to trust technology, independent of specific situations [42];
- reliance: behavioural responses to perceived system performance [50];
- perceived trustworthiness: subjective impressions shaped by cues or interface design [72];
- situational trust: context-dependent responses shaped by environmental and task-related factors [24];
- calibrated trust: alignment between user trust and actual system capabilities [32];
- warranted trust: trust justified by evidence of the system's competence and intentions [28].

This heterogeneity has led to theoretical ambiguity and divergent measurement strategies, especially in HCI and AI evaluation research [58]. Existing questionnaires primarily assess dispositional or trait-like measures, i.e. users' general tendency to trust machines or automation, e.g., [29, 43]. Currently, there are limited tools to assess post-interaction trust, or situational trust, i.e., how users' perceptions of trustworthiness evolve during interactions with AI systems.

Propensity to trust, in the broadest sense, is dispositional, reflecting a person's general tendency or natural inclination to trust others. Developmental research roots such dispositions in early interpersonal experiences, based on a person's generalised faith in human nature [55]. In infancy, consistent and sensitive caregiving fosters a general sense of safety and predictability [17], which evolves into cognitive expectations that others will act reliably [55, 56]. Across childhood, parental modelling of trustworthiness, cooperative peer relationships, and opportunities to detect and respond to deception shape both the tendency to trust and the discernment of when not to trust [5, 71]. With age, children's understanding of trust shifts from concrete, reward-based expectations toward more abstract notions of reciprocity, loyalty, and mutual vulnerability, particularly within close friendships. Webb and Worchel [71] emphasise that children actively construct trust beliefs through observation, reinforcement, and social comparison, using past experiences to form generalised expectations about others' reliability.

These interpersonal trust dispositions often transfer to novel domains, including human–automation interaction, as judgments about competence, predictability, and benevolence rely on similar cognitive and affective mechanisms. In technology contexts, PTT influences how children interpret early interactions with AI systems. High PTT may lead to overreliance, whereas low PTT may lead to premature rejection. Designing child-centred AI therefore requires supporting calibrated trust, helping children align their trust with actual system capabilities. This is all the more necessary for children, as compared to adults, given that their judgments of fairness, transparency, and comfort may shift rapidly in response to interactional cues. Because their source evaluation, critical stance, and privacy/moral reasoning are still developing, situational cues as feedback, perceived fairness, explainability, and data use strongly shape post-interaction trust [22, 35, 59].

2.1 Approaches to Measuring Trust in HCI

In the specific case of trust in automation, machines and/or technology, Lee and See [33] defined trust as a dynamic, situated attitude shaped by users' perception of system performance, process, and purpose. Hancock et al.'s [21] meta-analysis identified key factors influencing trust, including system reliability and user characteristics. These models have influenced the design of trust-aware systems in safety-critical high-stakes domains such as healthcare and defence[30].

Several questionnaires have been developed to capture trust in technology and automation. Jian et al.'s Trust in Automation Scale conceptualised trust and distrust as distinct but related dimensions [29], while McKnight et al. introduced a set of scales for online systems, distinguishing competence, benevolence, and integrity [40]. To contextualise these measures, McKnight et al.'s tripartite framework conceptualises trust in terms of Competence, Benevolence, and Integrity, capturing core cognitive beliefs about an AI system's ability, goodwill, and adherence to fair principles [40, 64]. In human–robot and autonomous systems research, Schaefer's Trust Perception Scale-HRI measured perceived reliability, dependability, predictability, and faith, and has since been adapted in studies of autonomous and AI-driven systems [57]. Ullman and Malle's measures have been widely applied, extending trust beyond reliability, predictability, and perceived intentions to include four subscales, Capable, Ethical, Sincere, and Reliable [65]. More recently, Körber advanced a multidimensional questionnaire for trust in automation, designed to disentangle cognitive, affective, and behavioural dimensions, including reliability, predictability, and intention, to assess adult users' trust in automated systems [31]. These tools are primarily adult-oriented, relying on domain-specific contexts such as aviation, e-commerce, or robotics, and employing technical terminology that may not align with children's developmental stages.

Some conceptual frameworks have begun to differentiate cognitive trust (e.g., trust based on evidence and system performance) from emotional trust (e.g., trust based on affect or intuition) [25], or to distinguish trust in competence (e.g., reliability and functionality) from trust in intentions [40, 41]. Yet, these distinctions are often not reflected in measurement tools. Furthermore, few empirical trust models meaningfully incorporate user-facing concepts such as transparency, explainability, or usability, principles that are increasingly relevant in contemporary child and user-rights frameworks [1, 11, 66, 67].

In recent years, HCI scholars have turned their attention to explainable AI (XAI), fairness, and algorithmic transparency, which directly relate to trust in competence (e.g., reliability, functionality) and trust in intentions (e.g., fairness, alignment with user goals) [14]. This body of work emphasises the importance of designing AI systems that communicate their reasoning, limitations, and data use to end-users in understandable ways, e.g., [6, 15, 70]. However, this research is often exploratory or design-focused, and validated questionnaires for evaluating how users experience trust in these systems remain sparse.

2.2 The need for measures of children's trust in AI

Understanding how children develop trust in intelligent systems is a growing concern in HCI and Child-Computer Interaction, especially as AI technologies increasingly shape children's learning, play, and communication environments. In particular, AI is increasingly being integrated into educational contexts and learning environments, including intelligent tutoring systems and conversational agents. However, understanding how students interpret AI behaviour, assign trust, and calibrate their reliance on these systems is still lacking [27]. Indeed, there is a notable lack of validated, multidimensional trust scales tailored to AI systems designed for children and adolescents, particularly in educational contexts[51, 54], and the HCI literature offers limited tools for measuring this kind of situated, post-interaction trust in ways that are developmentally appropriate and aligned with ethical AI principles.

Furthermore, while foundational models of PTT have been widely used to assess individuals' general tendency to trust automation (e.g., [9, 39]), they were originally designed for adult interpersonal and organisational contexts. In addition, these models of dispositional trust alone do not adequately account for how trust evolves during interaction, nor how system-specific factors (e.g., design features, feedback, agency) influence a child's perception of fairness, control, or safety.

To effectively support child-AI interaction, these models must be adapted to reflect children's developmental stages, lived experiences, and rights. Adapting these models provides a solid foundation for understanding how children's general attitudes shape their experience and expectations of AI.

2.3 A new model and measure for evaluating children's trust in AI

In order to address the issues identified above, we carried out a series of studies designed to create a new evaluation tool (*K-AI Trust*) that captures both dispositional and situational aspects of trust. In other words, it considers how naturally inclined children are to trust technology (dispositional trust) and how their trust is shaped during a specific interaction with an AI system (situational trust). When developing the questionnaire, we aimed to capture children's trust in AI as a holistic, interactional construct, grounded in their lived experiences.

The tool builds on theoretical insights from automation trust research, drawing on integrated models of PTT [43]. It combines these theoretical strands with principles from children's rights frameworks, particularly the UNCRC General Comment No. 25 [46], which calls for child-centred, age-appropriate design and evaluation practices. Concerns around fairness, transparency, responsiveness, and usability dimensions, increasingly foregrounded in policy frameworks such as the EU AI Act [47], the GDPR-K for kids [18], and others [1, 19, 45, 46], are also reflected in our questionnaire.

Accordingly, we identified three interrelated facets that shape how children experience and express trust. In our final questionnaire, these facets are not treated as separate dimensions, but rather as theoretical anchors that guide item development within a unified, developmentally grounded trust construct. They are the following:

- **Perceived Trustworthiness:** assessing judgments of competence, reliability, and fairness.
- **Transparency and Control:** gauging how well children understand the system's functioning and whether they feel able to manage their data and interactions.
- **Emotional Comfort and Engagement:** capturing affective responses such as enjoyment, safety, or unease when interacting with the system.

For the *perceived trustworthiness* facet, we focus on how children interpret the system's reliability, competence, and perceived intent. This facet aligns with multidimensional trust models that differentiate beliefs about a system's ability and intentions (e.g., competence, benevolence, integrity; [40]). In the context of child–AI interaction, these dimensions map onto children's judgments of whether an intelligent system is helpful, fair, and behaves ethically.

These qualities are particularly important in child-AI interaction because children often rely on cues such as consistency, accuracy, and fairness to judge whether a system can be depended upon [20, 72]. Additionally, frameworks such as the 5Rights principles [11] and the UNCRC [45] emphasise children's right to reliable and non-deceptive technologies. Items within this category were therefore designed to capture whether children perceive the intelligent system as dependable, competent in its responses, and aligned with benevolent intentions.

The *transparency and control* facet focuses on how children understand and manage their interactions with AI. We drew on models of explainability and user control in HCI and AI, particularly research highlighting that trust is reinforced when users can comprehend system actions and feel a sense of agency over them [24, 73]. Additionally, frameworks, such as the UNCRC General Comment No.25 [46] emphasise age-appropriate communication and children's agency in digital environments. Items within this category gauge whether children feel informed and in control during interaction.

For *emotional comfort and engagement*, we refer to emotional dimensions of trust that are especially salient in child-AI interaction, where feelings of safety, comfort, enjoyment, anxiety, or fear of betrayal strongly shape engagement. This facet is informed by affective-cognitive models of trust [64, 74] and by child-centred design frameworks that foreground emotional attunement, responsiveness, and user comfort in digital interactions, such as the 5Rights principles (e.g. Right to Safety and the Right to Support) and the UNCRS General Comment No. 25, which emphasises children's right to an emotionally safe and developmentally appropriate digital experience [11, 46]. We also draw on emerging CCI research highlighting the role of emotional support, relational cues, and user comfort in children's interactions with intelligent systems [7, 12]. Items within this category, therefore, capture whether children experience the system as approachable, enjoyable, and emotionally safe.

These theoretical iterations informed the development of a unified measurement tool. The questionnaire was ultimately validated as unidimensional, while our item selection and conceptual grounding reflect the integration of cognitive, emotional and interactional components of trust.

This structure helps researchers to assess both children's general tendency to trust technology and their trust in specific interactions. , Practitioners such as designers, educators, or policymakers can use it to monitor how children experience AI in real situations, spot early signs of distrust or anxiety, and support more transparent, inclusive interfaces. By grounding our scale development in both empirical data and real-world interaction contexts, we contribute to the growing call within the HCI community for metrics that support the responsible design and assessment of AI systems, particularly for vulnerable or underrepresented user groups, such as children.

The next sections outline the three studies that shaped the final *K-AI Trust* questionnaire.

3 Research Methodology

Study 1 examined children's dispositional trust using a modified version of the PTT [42]. The psychometric evaluation of this first questionnaire informed a refinement of the dispositional trust items. In Study 2, these refined items were combined with a new set of questions designed to measure children's situational trust after interacting with an AI system, yielding a combined dispositional-situational questionnaire. The psychometric analyses conducted in Studies 2 and 3 on this combined questionnaire led to the construction of the final *K-AI Trust* questionnaire for children, which jointly captures dispositional and situational trust.

Each study was carried out in the same primary school and involved multiple classrooms of year 4 and 5 pupils (aged 9 1o 11). The research protocol was reviewed and approved by the appropriate Institutional Review Board (protocol no. *2025-UNBACLE-0226225*). Before participation, written consent was obtained from parents or legal guardians of each child, in line with GDPR-K guidelines and local ethics requirements. Assent was also gathered from the participating children, in accordance with ethical guidelines for research with young people [11, 67, 68]. All analysis scripts are available at: osf.io/tew95.

3.1 Study 1

3.1.1 Participants. A total of 289 children participated in Study 1. The age range was 9–11 years (M = 9.75, SD = 0.68). The participants self-identified as Girl (52%), Boy (45.5%), and Not Disclosed (2.4%) (see Tab. 1).

Table 1: Participant Demographics for Study 1

Gender	Count	Mean Age	SD Age
Boy	132	9.82	0.70
Girl	151	9.68	0.65
ND	7	10.00	0.82

3.1.2 Materials. A **Prior Knowledge of AI Questionnaire** was developed to gauge children's knowledge of, and experience with, intelligent systems. The questionnaire included both factual and opinion-based items, designed to be accessible to children aged 9–11 and easy to complete independently (see Fig. 1).

Figure 1: Children completing the questionnaire individually in the school's hall.

These questions aimed to assess children's familiarity with AI and everyday technologies that incorporate intelligent behaviour. As such, the questions provided a comprehensive snapshot of the children's prior AI exposure, usage patterns, and early conceptualisations of intelligent systems, which were key contextual factors for interpreting their responses to the trust-related measures that followed. The Prior Knowledge of AI questionnaire can be found in Appendix A.

The **PTT Questionnaire** [42] was originally designed for adults. Therefore, it was discussed with teachers in the school, including the head teacher, in order to make age-appropriate modifications. It was then translated into the participants' native language, italian.

For example, the original PTT questionnaire referred broadly to "machines", a term that children found difficult to interpret. We therefore decided to focus on technological systems, such as computers or voice assistants (e.g., Alexa, Siri), to improve understandability. An English translation of the modified questionnaire can be found in Appendix B. We incorporated concrete, age-familiar examples (e.g., Alexa, TikTok) recommended by teachers to support comprehension and ecological validity, reflecting common forms of children's AI exposure: voice-based assistants and algorithmically curated content platforms. To ensure longevity, the questionnaire is designed so that platform names can be swapped with placeholders (e.g., "a voice assistant such as ___," "a social/media platform such as ___") to reflect future tools and local contexts.

Both questionnaires were extensively pilot tested and iteratively refined with a group of children in the target age range (9 to 11) before being used. Our item development was grounded in the theoretical and ethical frameworks reviewed earlier. PTT informed the baseline predisposition construct. [40] et al.'s competence–benevolence–integrity model and affective–cognitive views of trust guided the selection of child-friendly cues linked to ability, care, and confidence. Explainability and user-control frameworks supported items examining understandability and agency during interaction. Finally, UNCRC General Comment No. 25 [46], 5Rights [1, 11], and GDPR-K [18] principles shaped the inclusion of fairness, safety, and data-awareness elements to reflect children's digital rights. Together, these guided the three facets we target: competence, fairness, and comfort in child–AI interaction.

3.1.3 Procedure. After completing three demographic questions (age, class year, gender), children completed the "Prior Knowledge of AI" questionnaire, followed by the revised "Propensity to Trust" questionnaire.

The questionnaires were implemented as Microsoft Forms surveys and were accessible on individual tablet devices or school laptops. The questionnaires were self-administered to ensure standardisation and to respect each child's pace. The responses were collected digitally and stored securely.

3.1.4 Results. (Study 1)

In the **Prior Knowledge of AI Questionnaire** (see Appendix A), children were asked:

(Q1) *Whether they had ever heard of AI.* Their responses were: Yes: n = 258 (89%), No: n = 18 (6.2%) and Not sure: n = 14 (4.8%). Although AI familiarity is high across all genders, boys reported the highest awareness (91.7%) while the ND group showed slightly more uncertainty or lack of knowledge.

(Q2) *"Have you ever used any of these tools?"*, we found that Alexa (Amazon) was the most frequently reported tool (n = 152), followed by Siri (Apple; n = 135) and ChatGPT or similar systems (n = 124). Educational robots (e.g., Cozmo, BeeBot) appeared less often in children's responses (n = 25), and 40 participants reported not using any of the listed tools.

(Q3) *"When you use these tools, do they seem intelligent to you?"*, most children perceived the tools as intelligent to some degree, with 84.83% responding "Yes, a lot" or "A bit" (41.38%). Notably, 120 children selected "A bit," suggesting that many recognised some level of intelligence in AI, but in a more moderate and cautious way. A smaller proportion reported uncertainty ("I don't know", 10.69%), and only a few ("No," 4.48%). rejected the idea that the tools were intelligent. These results indicate that children tend to perceive AI as intelligence, although for a substantial subset, what "intelligence" means in the context of AI remains unclear or hard to interpret.

(Q4) *"What do you like to use technology for?"*, children most frequently reported listening to music (n = 157), learning (n = 148), and playing games (n = 129). *Talking to friends or family*was mentioned by 102 participants, while 14 selected other uses, including several unspecified "Other" responses (n = 14). These results suggest that while entertainment-oriented activities (e.g., music, games) dominate children's technology use, educational purposes and social connection also represent significant motivations. These motivations varied by gender. Girls reported using it more frequently to listen to music, while boys reported using it more often for learning and playing. Chi-square tests of independence were conducted to assess gender differences in these preferences. The difference in technology use for learning was not statistically significant. However, gender differences were significant for both listening to music, $\chi^2(2, N = 284) = 7.95, p = .019$, and playing, $\chi^2(2, N = 284) = 6.94$, $p = .031$.

(Q5) *"Do you think that what a voice assistant says is always true?"*, children provided a range of responses: "I never thought about it" was selected by 41 children (14 boys, 26 girls, 1 ND), "No" by 52 children (28 boys, 20 girls, 4 ND), "Sometimes" by 147 children (62 boys, 84 girls, 1 ND), and "Yes" by 50 children (28 boys, 21 girls, 1 ND).

Propensity to Trust results. To better understand children's overall attitudes toward intelligent systems, we computed descriptive statistics for the revised PTT scale. These statistics offer a snapshot of general trust levels across the full sample and help contextualise subsequent psychometric analyses (see Table 4).

Responses to the six items (with Q2 reverse-coded) were averaged per participant to obtain a composite trust score. On the 5-point scale, scores ranged from 1.17 to 4.67 ($M = 2.83$, $SD = 0.64$). As illustrated in Table 2, the distribution was slightly skewed, with a subgroup of children scoring below 2.33, reflecting relatively low trust. Subsequent analyses examined potential gender-based differences in trust.

Table 2: Descriptive Statistics for the revised Propensity to Trust (PTT) Scale

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
PTT Score	1.17	2.33	2.67	2.83	3.17	4.67

PTT Scale Reliability and Factor Structure. Internal consistency of the PTT was assessed using Cronbach's alpha, yielding a raw alpha of .60, with a standardised alpha of .59 and an average inter-item correlation of $r = .20$, which is within but at the lower end of the recommended range (.15–.50). The 95% confidence interval for alpha, calculated using the Feldt method, ranged from .52 to .67, indicating questionable internal consistency for the full scale. Reliability diagnostics flagged Item 2 as a weak contributor. It showed a low item–total correlation ($r = .12$) and a low item–rest correlation ($r_{\text{drop}} = .093$). When Item 2 was removed, Cronbach's alpha improved to .64, supporting its exclusion to enhance scale coherence.

An **Exploratory Factor Analysis (EFA)** using Maximum Likelihood extraction and no rotation confirmed a single-factor solution, consistent with the scale's original validation. Five items (Q1, Q3, Q4, Q5, Q6) loaded moderately to strongly on the extracted factor, with loadings ranging from .455 to .591. Item 2 did not load substantially and showed a very low communality of .013, further justifying its removal (see Fig. 3). The data were adequate for factor analysis (KMO = .74; Bartlett's test of sphericity, $\chi^2(15) = 158.65$, $p < .001$). The retained factor explained 22.6% of the total variance with an eigenvalue of 1.36. Although the retained factor explained only 22.6% of the variance, this is consistent with short scales that intentionally capture diverse aspects of a construct. Rather than indicating poor reflection of children's trust, the modest variance highlights the heterogeneity of children's experiences with AI and the breadth of the construct. Based on this analysis, the reduced version of the PTT scale excluded Item 2 and retained the five strongest items for subsequent analyses.

A **Confirmatory Factor Analysis (CFA)** was conducted to evaluate the factor structure of the revised PTT scale in a child sample. Based on prior reliability analysis, with Item Q2 excluded, the remaining five items (Q1, Q3, Q4, Q5, Q6) were specified to load on a single latent construct representing trust (see Fig. 2).

Model fit was assessed using maximum likelihood estimation. The model demonstrated excellent fit: $\chi^2(5) = 2.32$, $p = .803$, CFI = 1.00, TLI = 1.04, RMSEA = .000, 90% CI [0.000, 0.052], $p(\text{RMSEA} \leq .05) = .945$, and SRMR = .016. All fit indices exceeded conventional thresholds, indicating that the hypothesised one-factor model closely approximated the observed data in this child sample. The scale demonstrated modest but acceptable internal consistency (α = .64, ω = .64, CR = .64; see Table 3).

Table 3: Reliability Estimates for the revised PTT Scale

	Cronbach's α	McDonald's ω	Comp. Rel. (CR)
PTT (5 items)	.64	.64	.64

Standardised factor loadings ranged from .45 to .59, suggesting moderate to strong associations between the items and the latent construct. Squared multiple correlations (R^2) indicated that each item accounted for 23–36% of its variance through the latent factor. These final results of the revised PTT scale support the unidimensional structure of the revised PTT and its psychometric adequacy as a tool for measuring children's general trust in intelligent systems.

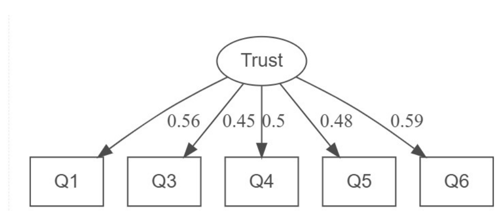

Figure 2: Confirmatory factor analysis model showing the single latent Trust factor and its standardised loadings onto items Q1, Q3, Q4, Q5, and Q6. Loadings range from .45 to .59, supporting a unidimensional structure of the PTT scale.

While the revised five-item PTT scale demonstrates statistical adequacy in model fit and internal consistency, the overall findings indicate the limited conceptual scope and developmental misalignment of the original scale, originally developed for adults, when applied to children. These issues, particularly the low reliability and poor variance explained, stress the need to develop a new, developmentally appropriate questionnaire that captures the multidimensional nature of children's trust in intelligent systems, including aspects such as transparency, perceived fairness, and emotional safety.

Table 4: Descriptive statistics for each item of the Propensity to Trust questionnaire administered to children.

Item	Mean	SD	Min	Max	Median
Q1	2.61	1.00	1	5	2
Q2	3.54	0.99	1	5	4
Q3	2.54	1.24	1	5	2
Q4	2.25	0.97	1	5	2
Q5	3.02	1.09	1	5	3
Q6	3.05	1.31	1	5	3

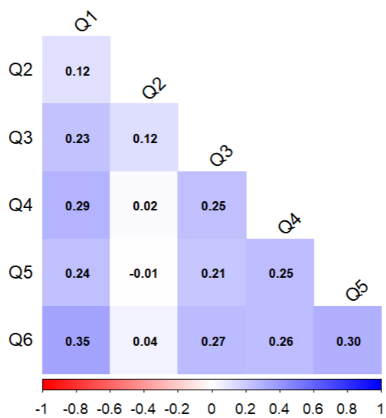

Figure 3: Inter-item correlation matrix for the six items of the PTT questionnaire. Coloured tiles and numerical values indicate the strength and direction of Pearson correlations between items.

3.2 Study 2

Building on Study 1, which established a general dispositional trust baseline, Study 2 focused on domain-specific dispositional trust in intelligent systems, probing children's beliefs about safety, fairness, and privacy in relation to "intelligent systems". These responses informed the refinement of the K-AI scale and laid the groundwork for the situational, post-interaction trust items introduced in the later study.

The aim of study 2 was to better understand how the interaction with an AI system might influence children's trust. The study involved having children interact with an AI ChatGPT and then answer a newly developed questionnaire: the preliminary **K-AI Trust Questionnaire**, which can be found in Appendix C.

3.2.1 Participants. Study 2 involved a subset of 85 children from Study 1 (see Tab 5). Although we would have hoped to include all the children in Study 1, the constraints of the end of the school year made this impossible.

Table 5: Sample 2 Participant Demographics

Gender	Count	Mean Age	SD Age
Boy	43	10.58	0.50
Girl	42	10.52	0.51

3.2.2 Materials.

K-AI Trust Questionnaire *(Preliminary version).* The preliminary version of the questionnaire was developed to assess children's post-interaction perceptions of AI systems, focusing on dimensions such as trust, perceived usefulness, transparency, fairness, emotional safety, control over personal data, and the quality of their experience with the AI system, including aspects related to ease of use, enjoyment, and perceived safety (see Appendix C for the full questionnaire). The questionnaire was based on earlier psychometric findings, informed by key principles from child rights and AI ethics frameworks, including UNCRC General Comment No.25 [46] and the 5Rights principles [11], which stress the need for age-appropriate, explainable, and rights-aligned technology design.

To ensure developmental appropriateness, we adapted item language and response formats based on established recommendations for survey research with children aged 9-11 [2]. The questionnaire was then pilot tested with a small group of ten children within the target age group to check for clarity, engagement, and comprehension in line with children's cognitive and emotional understanding. This informal testing led to minor revisions in wording and structure. Additionally, we consulted with the school's teachers and headteacher during the design process to ensure the questionnaire's developmental appropriateness and relevance. For instance, the original PTT item "I usually trust computers until they make mistakes" was adapted into "Do you think intelligent systems can make mistakes?" (item Q4 in the preliminary K-AI Trust), shifting the focus from self-reported disposition to a direct judgment about AI behaviour. Similarly, "For the most part, I distrust computers or systems like Alexa, Siri" was reframed as "Do you think that smart systems like Alexa or platforms like TikTok are safe to use?" (item Q1 in the preliminary K-AI Trust), a wording that is both simpler and aligned with children's daily digital environments.

Other adaptations involved broadening the functional framing of PTT items into evaluative judgments. For example, "In general, I would rely on a computer to assist me" became "Do you think that smart systems like ChatGPT are useful?" (item Q3 in the preliminary K-AI Trust). The PTT item "I am likely to trust a computer even when I have little knowledge about it" was inverted into two positively framed questions emphasising transparency: "Would you trust intelligent systems more if you knew exactly how they worked?" (item Q8 in the preliminary K-AI) and "Would you trust intelligent systems more if you knew exactly how they were using your data?" (item Q9 in the preliminary K-AI). Two PTT items, "My tendency to trust computers is high" and "It is easy for me to trust computers to do their job", were not retained. Both rely on abstract, self-reflective language that is developmentally demanding for younger respondents, and their conceptual content was partly redundant with other items addressing usefulness, reliability, and error recognition. To ensure developmental appropriateness and reduce redundancy, these items were excluded in favour of more concrete, contextually grounded questions.

Initially, the full questionnaire comprised 14 questions in total. Of these, 11 closed questions were statistically analysed and included in the psychometric analysis, while the remaining three questions were analysed qualitatively, due to their open-ended nature (see Appendix C for the full questionnaire).

3.2.3 Procedure. Data collection was carried out within the school setting during regular class hours, in close collaboration with teachers and school administrators. Children were invited to interact individually with ChatGPT-4o, accessed via a laptop in the school's entrance hall. The interaction was supervised by a researcher and designed to last only a few minutes. Children were instructed to use voice commands to ask the system at least two math-related questions of their own choosing. The questions varied in complexity: some children posed simple arithmetic problems (e.g., 2 + 7), while others asked more complex operations (e.g., 2345 × 569) and still others provided short math problems to solve (e.g., "If you have 3 apples and you get 2 more, how many apples do you have in total?").

This setting allowed us to observe how children approached the system and how they responded to its behaviour. Depending on ambient noise and input phrasing, ChatGPT did not always produce correct answers, which in turn elicited diverse reactions from the children including surprise, amusement, and confusion. These brief interactions served as a context for eliciting children's immediate impressions of the system's capabilities and trustworthiness.

For GDPR compliance and school data protection policies, a GoPro 4 webcam was used to capture only audio, ensuring no facial data were recorded. Following the interaction, children were directed to a second set of devices (tablets, laptops, mobile phones) at a different table, where they completed the preliminary version of the K-AI Trust Questionnaire. Responses were collected digitally and stored securely.

3.2.4 Results.

K-AI Trust Questionnaire (preliminary version). The overall trust score was calculated by averaging responses across the 11 closed items items, which were coded on a 5-point ordered response scale (1–5). Descriptive analysis revealed a moderate level of trust with limited dispersion in responses (see Tab. 6).

Table 6: Descriptive Statistics for the Preliminary K-AI Trust Questionnaire

	Mean	Median	SD	Variance
K-AI Trust Score	2.72	2.73	0.49	0.24

Exploratory Factor Analysis and Scale Refinement. To investigate the dimensionality of the preliminary K-AI Trust, an Exploratory Factor Analysis (EFA) was conducted using oblique (promax) rotation. Initial reliability analysis of the 11-item scale revealed low internal consistency (Cronbach's α = .46). After reverse-coding negatively worded items (Q8 and Q9), reliability improved substantially (α = .67; standardised α = .67). Guttman's λ_6 was .72. The average inter-item correlation was r = .16, which falls within the low to moderate range, suggesting that while items relate to a common construct, they may reflect slightly diverse aspects of trust. After removing items Q8, Q10, and Q11, internal consistency improved to an acceptable level (Cronbach's α = .70). We then conducted a parallel analysis to determine the underlying factor structure. The result suggested a two-factor solution for factor retention (see Fig. 4), although a single-factor model also appeared plausible. The EFA identified two interpretable components: the first reflected *Reliability and Usefulness*, and the second captured *Transparency and Critical Awareness*. However, several items displayed poor psychometric properties:

- **Q4**: Low communality (h^2 = .33).
- **Q6**: Very low communality (h^2 = .03).
- **Q7**: Very low communality (h^2 = .02).

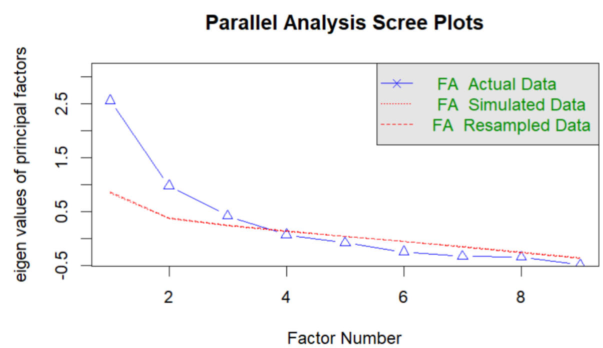

Figure 4: Parallel analysis scree plot for the preliminary K-AI Trust Questionnaire. The eigenvalues of the actual data (blue line with triangles) exceed those of the simulated and resampled data (red lines) only for the first component, suggesting the retention of a single factor.

A refined eight-item version of the scale was created by excluding problematic items. In this model, Q1–Q5 and Q9 loaded strongly on the *Reliability and Usefulness* factor, with loadings going from .49 to .63, while Q6, Q7 captured *Transparency and Critical Awareness* with loadings ranging from .54 to .74. This revised structure explained 38% of the total variance, an improvement over the original 11-item version (20%), and yielded clearer factor separation.

Each subscale was tested for internal consistency. The *Reliability and Usefulness* subscale (Q1–Q5 and Q9) demonstrated acceptable reliability (α = .73; standardized α = .73; average inter-item r = .13; Guttman's λ_6 = .57), while the *Transparency and Critical Awareness* subscale (Q6 and Q7) showed comparable performance (α = .62; standardised α = .62; average inter-item r = .45; Guttman's λ_6 = .45). Factor loadings were generally moderate, ranging from .16 to .64. Communalities varied widely, from 0.13 to 0.94

reflecting heterogeneity across items. Although these coefficients are modest, they are consistent with early-stage or exploratory scale development. These results support a multidimensional interpretation of children's attitudes toward intelligent systems, suggesting that trust and awareness function as distinct but related constructs. This refined two-factor structure was adopted for subsequent confirmatory analyses and for use in future iterations of the K-AI Trust scale.

Confirmatory Factor Analysis of the preliminary version of the K-AI Trust Questionnaire. A confirmatory factor analysis (CFA) was conducted using maximum likelihood estimation via the `lavaan` package (v0.6-19) to evaluate the structure of eight questionnaire items intended to measure children's trust in intelligent systems. Based on previous exploratory findings, a two-factor model was specified, with *Reliability_Usefulness* (Q1–Q5 and Q9) and *Data_Control* (Q6-Q7) as correlated latent constructs. The model was estimated using data from 85 participants, with one missing data pattern handled via full information maximum likelihood (FIML).

Parameter Estimates. All items loaded onto their respective factors, though some with weaker or problematic estimates:

- *Reliability_Usefulness:* Q1 = .53, Q2 = .62, Q3 = .41, Q4 = .64, Q5 = .61, Q9 = −.54
- *Transparency:* Q6 = 0.56, Q7 = .77

Q6 and Q7 loaded on the *Transparency* factor, with Q6 showing an anomalously high standardised loading ($\lambda = 1.23$) indicative of a Heywood case, and Q7 displaying a more moderate loading ($\lambda = .49$). The residual variance for Q7 was relatively high ($Var = .76$), suggesting that it captures substantial unique variance, while the variance for Q6 could not be reliably estimated. The two latent factors were moderately correlated ($r = .58$, $p < .001$), indicating related but distinct dimensions.

Model Fit. The one-factor model showed very poor fit, $\chi^2(20) = 80.85$, $p < .001$, CFI = .69, TLI = .56, RMSEA = .19, 90% CI [.15, .24], SRMR = .14. The two-factor model provided a better fit, $\chi^2(19) = 35.85$, $p = .011$, CFI = .91, TLI = .87, RMSEA = .10, 90% CI [.05, .15], SRMR = .10, though it still failed to reach conventional thresholds for adequate fit, reflecting the limited number of items loading on the *Transparency* factor.

Table 7: Model fit indices for the two-factor CFA model

Fit Index	Value
$\chi^2_{\text{scaled}}(19)$	35.85, $p = .011$
Comparative Fit Index (CFI)	.913
Tucker–Lewis Index (TLI)	.871
RMSEA	.103, 90% CI [0.048, 0.154]
SRMR	.097

Despite that, the factor loadings aligned with the intended distinction between *reliability/usefulness* and *transparency/awareness* (see Tab. 8). While fit indices[1] fell short of conventional thresholds, the model was built on theoretical grounds, providing preliminary support for the proposed tool and underscoring the need for further item refinement.

Table 8: Standardised factor loadings and residual variances for the two-factor CFA model.

Item	Factor	Stand. Loading	Resid. Var.
Q1	Reliability_Usefulness	0.596	0.645
Q2	Reliability_Usefulness	0.655	0.558
Q3	Reliability_Usefulness	0.530	0.719
Q4	Reliability_Usefulness	0.695	0.517
Q5	Reliability_Usefulness	0.665	0.558
Q9	Reliability_Usefulness	0.545	0.704
Q6	Transparency	1.229	NA
Q7	Transparency	0.493	0.757

Validity analysis showed that *Reliability_Usefulness* had marginal reliability (CR = .65) and insufficient convergent validity (AVE = .38 < .50). The *Transparency* factor (Q6 and Q7), yielded a Heywood case, preventing meaningful AVE and CR estimates. In sum, *Reliability_Usefulness* is only partly coherent and *Transparency* is psychometrically unstable and needs substantial item revision, including item refinement or the addition of further indicators, to adequately represent the intended construct.

Latent Variable Correlation. The estimated correlation between the two latent constructs was weak and positive ($r = .09$), indicating that children's evaluations of *Reliability_Usefulness* were largely independent from their perceptions of *Transparency*. This result suggests that the two factors are empirically distinct, though the instability of the *Transparency* factor, defined by only two items and affected by a Heywood case, limits the reliability of this estimate. Nonetheless, the emergence of a separate factor alongside *Reliability_Usefulness* provides preliminary support for the theoretical distinction between functional reliability/usefulness and ethical dimensions such as fairness and data control. These findings underscore the need for further item refinement and expansion to obtain stable and interpretable estimates of the relationship between these dimensions.

Quantitative Analysis. We conducted a thematic analysis of responses to two open-ended questions to gain insight into children's real-world interactions with AI systems: *Q13 "What have you used ChatGPT for?"* and *Q14 "Do you think these systems are intelligent"*? These questions were designed to uncover both behavioural and perceptual dimensions of children's experiences with intelligent systems.

Children's Use of ChatGPT. Children's responses to the open-ended question "What have you already used ChatGPT for?" (Q13) were analysed thematically and grouped into seven categories. The most frequently reported use was homework support, underscoring ChatGPT's perceived value as a learning aid. Many children

[1]Fit indices are reported using robust estimators (WLSMV). Although the two-factor model did not reach conventional thresholds for good fit (Robust CFI = .913, TLI = .871, RMSEA = .103, SRMR = .097), it is retained on theoretical grounds, with the recognition that further item refinement is needed, particularly for the *Transparency* factor, which is defined by only two indicators.

also stated they had never used the tool, revealing differing levels of familiarity within the sample. Other responses reflected a range of creative and exploratory uses, such as generating drawing ideas, storytelling, seeking information, and satisfying curiosity. Some children described multiple purposes, e.g., using ChatGPT for both educational and playful activities, while a smaller number mentioned emotional or entertainment-oriented interactions, for example, seeking comfort: 'When my best friend didn't want to be with me anymore, she, let's say, [the AI], consoled me.'"

These themes show that children engage with conversational AI in different ways. The identified categories range functional, imaginative, and affective dimensions, shedding light on children's emerging mental models of AI and the diverse contexts in which trust may begin to take shape.

- **Homework**: School-related uses such as solving math problems, writing assignments, or receiving help with grammar.
- **Creativity**: Imaginative or original tasks, like inventing rhymes, generating or transforming multimedia content, e.g., turning photos into cartoons, composing poems.
- **Play**: Recreational uses including playful dialogue, games, or exploring the system for fun.
- **Curiosity**: Exploratory behaviour aimed at testing the system's capabilities or understanding how it works.
- **Research**: Seeking factual information or asking knowledge-based questions (e.g., "Where is Vienna?").
- **Never Used**: Children reporting no prior interaction with ChatGPT due to lack of access, interest, or opportunity.
- **Entertainment**: Children using ChatGPT for playful or leisure activities such as games or jokes.
- **Emotional support**: Children describing ChatGPT as a source of comfort, companionship, or reassurance in moments of social or emotional need.

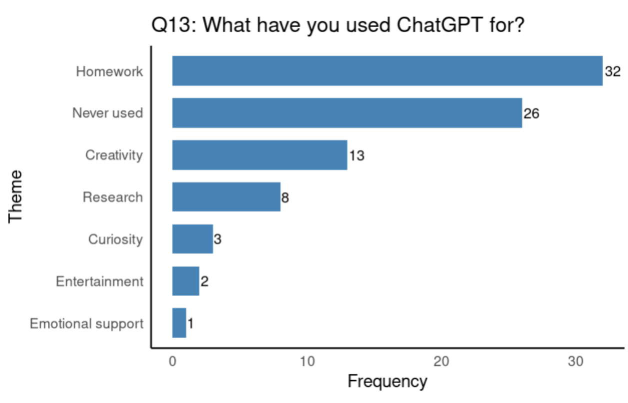

Figure 5: Thematic Categorisation of Children's Responses to Q13 ("What have you used ChatGPT for?")

Perceptions of Intelligence. Children's responses to the question "Do you think these systems are intelligent?" (Q14) were analysed thematically and grouped into five main categories:

- **High Perceived Intelligence** – Children explicitly describe the system as smart, capable, or comparable to human cognition.
  Example: "Yes, it answers everything!"
- **Moderate Perceived Intelligence** – Mixed or hesitant responses about the system's intelligence.
  Examples: "A bit", "More or less"
- **Low Perceived Intelligence / Distrust** – Clear skepticism about the system's intelligence or trustworthiness.
  Example: "I don't trust it"
- **Attribution to Human Creation** – Explanations attributing system performance or limitations to its human origin.
  Example: "No, because a human made it"
- **Uncertainty / Ambivalence** – Vague, unclear, or ambiguous answers showing indecision or unfamiliarity. *Example: "I don't know", "I'm not sure"*

The majority of responses (n = 81, 30.9%) fell under High Perceived Intelligence, in which children explicitly described AI systems as smart, capable, or comparable to human cognition (e.g., "Yes, it answers everything!"). A substantial proportion (n = 70, 26.7%) expressed Moderate or Uncertain Perceived Intelligence, using qualified terms such as "a bit," "more or less," or "I don't know," reflecting ambivalence or developmental hesitancy in judging technological capability.

Some children expressed Low Perceived Intelligence or Distrust (n = 49, 18.7%), voicing clear skepticism or rejection of AI's intelligence, often citing limitations or past negative experiences (e.g., "I don't trust it"). A similar category, Attribution to Human Creation (n = 35, 13.4%), included children who reasoned that because AI was made by humans, it could not truly be intelligent (e.g., "No, because a human made it"). Finally, Uncertainty or Ambivalence responses (n = 27, 10.3%) reflected confusion or unfamiliarity, often with brief or unclear statements.

These results suggest that children think about AI's intelligence in very different ways, and that many already take a fairly critical or reflective stance toward these systems. The fact that nuanced and even sceptical views appear among younger children as well underlines the need to build in clear explanations and trust-building features when designing AI systems for children.

3.3 Study 3

Study 3 finalised the validation of the K-AI Trust scale when administered immediately after child–AI interaction, confirming its factor structure and reliability. In this context, the K-AI scale captures post-interaction, domain-specific trust in intelligent systems, including situational aspects that children can readily articulate, whereas the revised PTT scale remains our purely dispositional baseline, used only in Study 1.

Based on limitations in internal consistency and factor structure of the previous questionnaires, a new 9-item trust scale was developed: the final K-AI Trust Questionnaire (see Appendix D).

3.3.1 Participants. Study 3 involved a subset of 85 children drawn from the same sample as Study 1, and largely overlapping with the sample for Study 2 (see Tab 9). There were minor variations in gender distribution due to differences in attendance on the day of data collection; however, the age range and school setting remained consistent across samples, ensuring comparability in the analysis. A subset of participants from Studies 1 and 2 also participated in Study 3 where the materials and interaction scenarios were different from earlier to avoid potential for familiarity with the setting and task structure and mitigate practice or recognition effects. Additionally, responses capture post-interaction trust, which is expected to be context-sensitive rather than memorisation-driven, further reducing the likelihood of repeated-exposure bias. At the same time, partial sample overlap supports comparability across iterations by holding school context and cohort characteristics constant, which reduces between-group variability during item refinement.

Table 9: Sample 3 Participant Demographics

Gender	Count	Mean Age	SD Age
Boy	39	10.1	0.72
Girl	46	10.02	0.8

3.3.2 Materials. Following an exploratory factor analysis of the preliminary version of the K-AI Trust questionnaire, administered in Study 2, a new version of the K-AI Trust questionnaire was defined based on this new modification. To strengthen the internal consistency of this new scale, we retained the most robust items, those with strong item-total correlations and a positive contribution to Cronbach's alpha. Additionally, three high-loading items from the modified PTT questionnaire, which had demonstrated acceptable reliability, were rephrased in a child-friendly manner and integrated into the revised questionnaire, the final K-AI Trust Questionnaire that can be found in Appendix D. Specifically, from the preliminary to the final version of the K-AI Trust Questionnaire, we retained items addressing safety, privacy, usefulness, mistakes, permission, and data use, while removing those on harm and fairness, which were less developmentally suitable. To strengthen the construct coverage, we introduced three items adapted from the PTT questionnaire. Specifically, Final Q8 (*"Would you rely on a computer to help you?"*) reflects PTT Q3 on reliance, Final Q9 (*"Are you likely to trust a computer even if you don't know how to use it?"*) parallels PTT Q6 on trusting stance, and Final Q5 (*"Would you trust intelligent systems more if you knew exactly how they worked?"*) echoes the PTT's focus on transparency as a precondition for trust. These adaptations resulted in a balanced 9-item instrument that combines child-centred concerns (safety, privacy, usefulness) with dispositional dimensions of trust established in prior scales.

This final data collection enabled psychometric testing and exploratory factor analysis to assess the internal structure and reliability of the updated questionnaire (see Table 9). As with the other studies, the questionnaire was delivered online via a secure form, and the responses were collected further to assess the reliability and structure of the updated scale.

3.3.3 Results.

Validation of K-AI Trust Questionnaire. This section explains the validation process and psychometric properties of the final K-AI Trust questionnaire. The validation included reliability analysis, exploratory and confirmatory factor analysis, construct validity, and measurement invariance testing across gender and age groups.

K-AI Trust Questionnaire Results. The scale included 9 items, each rated on a 5-point ordered response scale. The overall trust score, computed as the average of participants' responses across items, was $M = 3.03$, $SD = 0.67$, with a median of $Mdn = 3$ and a variance of $Var = 0.46$, indicating a moderate level of trust with moderate variability in responses, and was normally distributed (W = 0.98, p = .388). The mean item scores ranged from 2.33 to 4.05, with standard deviations between 1.05 and 1.44. The highest mean was observed for Q4 ("M" = 4.05, SD = 1.38), while Q2 had the lowest ("M" = 2.33, SD = 1.15).

Exploratory Factor Analysis EFA. To assess the internal consistency of the K-AI Trust Questionnaire, we conducted a series of psychometric evaluations. The scale demonstrated good reliability and internal consistency, with Cronbach's alpha = .72 (standardised $\alpha = .73$), Guttman's $\lambda_6 = .76$, 95% CI [.62, .80], and an average inter-item correlation of $r = .23$, indicating a moderate level of homogeneity among items [63].

Table 10: Reliability if an item is dropped

Item	Raw α	Std. α	G6 (smc)	Avg. r	S/N	SE	Var. r
Q1	0.68	0.71	0.71	0.21	2.09	0.05	0.031
Q2	0.71	0.72	0.74	0.24	2.58	0.047	0.025
Q3	0.70	0.71	0.75	0.23	2.51	0.042	0.0131
Q4	0.76	0.76	0.78	0.28	3.25	0.038	0.016
Q5	0.67	0.68	0.72	0.21	2.12	0.05	0.029
Q6	0.65	0.67	0.70	0.20	2.01	0.057	0.028
Q7	0.67	0.68	0.72	0.21	2.12	0.05	0.034
Q8	0.68	0.69	0.73	0.22	2.27	0.05	0.033
Q9	0.69	0.70	0.73	0.22	2.33	0.05	0.031

An item-level analysis (see Table 10) showed that most items contributed positively to internal consistency. Removing individual items resulted in only marginal changes to Cronbach's alpha (range: $\alpha = .68$ to .76), reinforcing the general coherence of the scale. However, Item Q4 showed poor psychometric performance, with a negligible factor loading ($\lambda = .07$), very low item-total correlation ($r_{\text{drop}} = .02$), and a high proportion of extreme responses. These indicators suggest that Q4 may not adequately reflect the latent trust construct and could benefit from rewording or conceptual clarification.

To further examine the underlying structure of the K-AI Trust Questionnaire, we conducted an exploratory factor analysis (EFA) using principal axis factoring with varimax rotation. Parallel analysis supported a unidimensional solution, with the first factor exceeding the eigenvalue threshold set by simulated data.

The one-factor solution accounted for 31.8% of the total variance and produced factor loadings ranging from .42 to .71. Most items loaded moderately to strongly on the latent trust factor, confirming their contribution to the core construct. However, consistent with the reliability findings, Q4 showed the weakest loading and lowest communality, indicating potential conceptual or phrasing issues.

A heatmap of factor loadings (see Fig. 7 and Tab. 11) visually reinforced these results, showing that all items contributed positively to the main factor except Q2 and Q4, which diverged from the general pattern. Despite these concerns, the one-factor solution remains theoretically coherent and psychometrically justifiable, offering a parsimonious model of children's trust in intelligent systems. We explored the impact of removing Q2 and Q4, which resulted in improved internal consistency (Cronbach's α = .76; standardised α = .76; Guttman's λ_6 = .76; average inter-item correlation = .29).

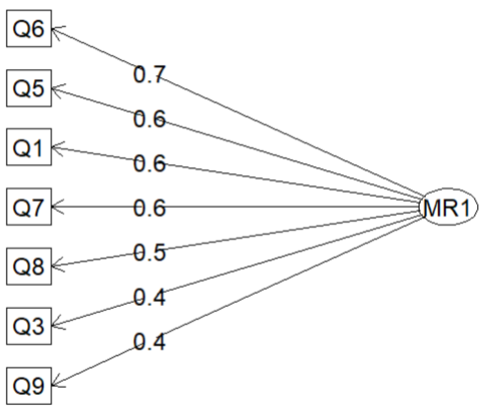

Figure 6: Final factor structure of the K-AI Trust Questionnaire. Exploratory factor analysis (EFA) with one factor solution indicated that seven items (Q1, Q3, Q5, Q6, Q7, Q8, Q9) loaded on a single latent construct of trust. Factor loadings ranged from .40 to .70, suggesting that all items contributed meaningfully to the latent factor.

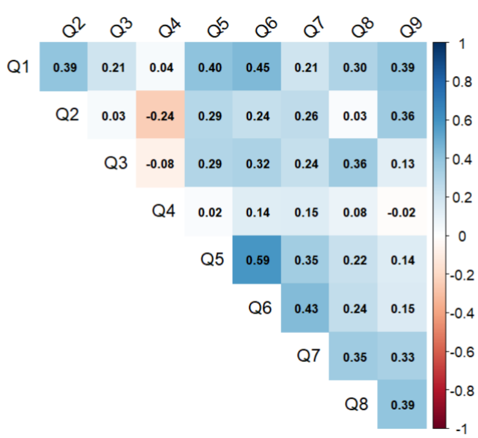

Figure 7: Heatmap of Pearson correlations between items in the new trust questionnaire (Q1–Q9). Darker blue shades indicate stronger positive correlations, while reddish hues indicate negative correlations. Diagonal values represent perfect self-correlation (r = 1).

Confirmatory Factor Analysis (CFA). A one-factor CFA model was tested using robust maximum likelihood estimation (MLR). Model fit indices showed poor results: $\chi^2(14) = 32.90$, $p < .003$, Comparative Fit Index (CFI) = .84, Tucker-Lewis Index (TLI) = .76, Root Mean Square Error of Approximation (RMSEA) = .13, and Standardised Root Mean Square Residual (SRMR) = .09. All standardised loadings were significant ($p < .001$) and ranged from .35 to .71 (see Table 11).

To further evaluate the factor structure, a two-factor CFA model was also tested, grouping items into *usefulness* and *perceived transparency* dimensions. This model did not yield better fit indices compared to the one-factor solution ($\chi^2(26) = 55.20$, $p = .001$, CFI = .80, TLI =.73, RMSEA = .12, SRMR = .09). Information criteria (AIC = 2362.52, BIC = 2408.93) were higher for the two-factor solution compared to the one-factor model (AIC = 1814.79, BIC = 1848.99), further indicating that the two-factor structure was not preferable. In the final CFA model, standardised loadings ranged from .41 to .73, with all items exceeding the recommended threshold of .40, except Q4, which loaded very weakly (.07) and was therefore excluded (see Table 11).

Table 11: Standardised Factor Loadings from Confirmatory Factor Analysis

Item	Standardized Loading
Q1	0.614
Q2	0.410
Q3	0.413
Q4	0.071
Q5	0.683
Q6	0.732
Q7	0.557
Q8	0.444
Q9	0.407

Given the more parsimonious structure, theoretical coherence, and comparable fit, the one-factor model was retained.

Construct Validity. To examine convergent validity, we calculated Composite Reliability (CR) and Average Variance Extracted (AVE) for the final unidimensional K-AI model. CR was acceptable (.76), indicating good internal consistency. AVE was lower (.30), which suggests that items share less common variance than ideal. This pattern is common for early-stage child instruments and likely reflects the developmental diversity in how children read trust cues. Overall, the scale is reliable and shows partial evidence of convergent validity, appropriate for a first-iteration tool.

We also assessed discriminant validity between dispositional trust (PTT) and situational trust as measured by K-AI. The correlation between the trust scores was small and non-significant, r(83) = −.14, p = .22, with a wide confidence interval (−.34, .08), indicating no stable association. This supports the interpretation that K-AI captures a distinct, interaction-shaped judgment that is not reducible to children's baseline attitudes toward computers. Thus, K-AI demonstrates clear discriminant validity relative to an adapted dispositional trust measure.

Measurement Invariance Measurement invariance across gender (female vs. male) and age groups (9 vs. 10–11) was tested using multigroup CFA with robust estimation (Satorra-Bentler corrections).

Gender Invariance. Chi-square difference testing showed no significant differences across configural, metric, scalar, and strict models ($\Delta\chi^2(6) = 5.95$, $p = .43$ for metric).

Measurement invariance across gender was supported. The configural model showed acceptable though modest fit (CFI = .79, RMSEA = .15, SRMR = .10), and fit indices remained stable for the metric (CFI = .78, RMSEA = .14, SRMR = .11), scalar (CFI = .79, RMSEA = .13, SRMR = .12), and strict models (CFI = .81, RMSEA = .11, SRMR = .12).

Age Group Invariance. Measurement invariance was also supported across age groups ($\Delta\chi^2(12) = 7.70$, $p = .81$ for metric).

Similarly, measurement invariance was supported across age groups. The configural model indicated adequate fit (CFI = .78, RMSEA = .16, SRMR = .11), which improved incrementally across the metric (CFI = .80, RMSEA = .13, SRMR = .14), scalar (CFI = .83, RMSEA = .11, SRMR = .14), and strict models (CFI = .86, RMSEA = .09, SRMR = .14).

The final K-AI Trust questionnaire, with seven items, is shorter, more focused, and developmentally tailored. Items were rephrased from abstract ethical reflections to concrete, experience-based trust judgments, ensuring children could understand and answer meaningfully (see Appendix D).

4 Discussion

This study set out to develop and validate a child-centred tool to assess trust in intelligent systems, addressing a gap in HCI and child–AI research. Although "trustworthy AI" features prominently in policy frameworks such as the EU AI Act, we still lack empirical tools that capture children's own perspectives. Our mixed-method, iterative process combined psychometric refinement and thematic analysis to address this gap. Below we discuss the results through our two guiding research questions. We interpret trust as operating at two related levels. Dispositional trust, measured with the child-adapted PTT in Study 1, reflects children's general, trait-like tendency to trust technology. By contrast, the K-AI measures in Studies 2 and 3 tap post-interaction, domain-specific trust in intelligent systems: children answer immediately after engaging with an AI system, so their judgments about safety, usefulness, fairness, and data use are coloured by that specific experience. In this way, K-AI captures situational aspects of trust that children can readily report, for example whether, in this context, the system felt safe, fair, and worth relying on. This trait–state inspired framing helps position our work within child-centred HCI and provides tools for evaluating trust grounded in children's lived experience with AI.

RQ1: Which dimensions best capture the trust construct when creating a child-centred scale within the questionnaire?

Across the three studies, trust levels remained moderate, with mean scores of $M = 2.83$ (Revised PTT), $M = 2.72$ (Preliminary K-AI Trust), and $M = 3.03$ (Final K-AI Trust). Children did not over-trust or distrust the AI; most adopted a cautious and balanced stance, mixing practical experience with an emerging ethical sense.

The adapted PTT produced moderate trust scores but low internal consistency ($\alpha = .59$). Abstract, adult-centric items (e.g. *Q4 "My tendency to trust computers is high"*), were poorly understood and displayed weak factor loadings. This stressed the clear mismatch between adult metrics and how children reason about technology, which often centres on relational and emotional signals like friendliness or tone.

The preliminary K-AI Trust Questionnaire introduced items covering performance and ethical aspects, yielding a tentative two-factor structure: (1) *Usefulness and Reliability* and (2) *Transparency and Critical Awareness*. Although conceptually consistent with current HCI debates, the two dimensions were moderately correlated ($r = .58$, $p < .001$), and cross-loadings suggested that children evaluated AI more holistically, less as separate components and more as an integrated experience [3]. The first factor mainly captured helpfulness and consistency, while the second reflected data use and system understanding. This twofold structure resonates with the idea of calibrated trust [32] and recent work on children's multidimensional evaluations of AI [23], but it offered limited statistical advantage.

In response, we refined the scale, simplified language, removed redundancy, and grounded items in children's everyday lived experiences with AI. The resulting nine-item K-AI supported a unidimensional structure and showed improved reliability ($\alpha = .75$, $\lambda_6 = .77$). Developmental psychology suggests that children this age often integrate multiple cues when judging objects, so a single, coherent factor is theoretically aligned and more parsimonious [4, 16, 49].

Qualitative data confirm this interpretation. Children described AI as "helpful," "friendly," or "understanding," mixing functionality with how the interaction felt. Many linked trust to personal use, schoolwork help, searching for information, or playful exploration. A smaller group described the AI in more relational terms, sometimes as a companion. These accounts show that trust is shaped not only by performance but also by social and emotional meaning-making, raising ethical questions about the roles these systems play in children's lives.

PTT and K-AI are related but not interchangeable. PTT offers only a rough baseline of general trust and still includes terminology that does not reflect how children encounter AI today. K-AI captures trust built during the interaction itself and aligns more closely with children's lived experiences. Correlations between the two were small to moderate, consistent with the idea that they assess different levels of trust. Younger children may require alternative formats (e.g., visual scales), and future work should explore how trust develops across age. The adapted PTT likely needs substantial redesign, whereas K-AI provides a clearer foundation for assessing situational trust.

RQ2: How do children interpret and respond to items related to ethical AI principles (e.g., consent, data use) within a trust questionnaire context?

Ethical concerns, especially consent and data use, emerged in both the questionnaire and children's explanations. Although one consent-related item performed poorly, the underlying construct is central to children's digital agency and should remain part of future iterations, with clear wording. Qualitatively, many children expressed discomfort about systems "taking things without asking," showing intuitive awareness of personal data and ownership. Others framed AI as a "memory helpner," linking support with questions of fairness or privacy. Children clearly think about these issues in concrete, experience-based ways rather than abstract policy terms. These findings highlight why situational measures matter.

When children judge trust, they consider fairness, data practices, and ease of understanding during the actual interaction, factors that trait-based tools like PTT do not capture.

Overall, children's trust evaluations were holistic and ethically aware. They recognised principles such as consent, fairness, and privacy, but only when phrased concretely. This stresses the need for child-centred questionnaires that reflect how children include ethical principles into their lived experiences of AI.

Our framework translates child-rights principles (e.g., fairness, agency, privacy) and classical trust constructs (competence, relational confidence) [39, 40] into measurable items. This strengthens construct validity and clarifies how ethical and cognitive concepts can be expressed in child-appropriate ways.

For RQ1, the evidence supports a developmentally coherent unidimensional construct of trust, reflecting children's integrated evaluations. For RQ2, children recognised ethical principles like consent and fairness, though items require careful framing to ensure comprehension. The final K-AI Trust Questionnaire demonstrates both psychometric stability and ecological validity, while highlighting areas, particularly digital agency, for future refinement.

Design Implications. Designers should view children's trust in AI as a dynamic, relational process shaped by developmental context. This means designing for emotional resonance, transparency, and agency in addition to functionality. The validated K-AI Trust questionnaire can be used within iterative design cycles to evaluate how child-facing AI systems, such as educational tutors or well-being apps, are perceived. Importantly, items on ethical principles show that children can engage with concepts like privacy and consent, but only when framed concretely (e.g., "asking before taking" rather than "data sharing"). Designers should therefore translate abstract ethical requirements into relatable interactions to foster comprehension and trust.

Research Implications. Beyond addressing current limitations, our findings suggest several directions for advancing research on children's trust in AI:

- **Cross-cultural theorisation** - Future studies should not only validate the scale in diverse contexts but also examine how cultural and educational norms shape children's trust constructs.
- **Methodological expansion** - Combining self-report with behavioural and observational methods can generate richer, multi-layered accounts of how trust unfolds in practice.
- **Developmental pathways** - Longitudinal research can trace how trust trajectories evolve with age, digital literacy, and repeated exposure to AI, clarifying when multi-factor models may emerge.
- **Ethical constructs in focus** - Items capturing digital agency (e.g., consent, data ownership) should be refined through co-design and cognitive interviewing, ensuring that ethical salience is retained in developmentally accessible ways.

Our findings highlight the importance of moving beyond adult-centric assumptions. The K-AI Trust questionnaire represents a meaningful step toward child-informed, developmentally grounded evaluation of AI systems. The two-level approach—dispositional PTT plus situational K-AI—offers a simple strategy for understanding children's baseline attitudes and experience-based trust. At the same time, the limits of the adapted PTT show the need for renewed dispositional measures that reflect how children actually encounter AI today.

5 Future work

While this study advances the development of child-centred trust measures for intelligent systems, several limitations and future directions should be noted. Our main aim here was to ensure that the construct was clear, developmentally appropriate, internally consistent, and usable in child-centred AI settings. In future work, we plan to further refine K-AI to capture additional facets of children's trust and to administer it alongside other child–AI measures to examine convergent and divergent relations, positioning the scale more clearly within the wider trust-measurement landscape. The sample was limited to children aged 9–11 from a relatively homogenous socio-cultural context, which may constrain generalisability. Children's trust in AI is likely shaped by linguistic, cultural, and socio-economic factors. Future studies should therefore include more diverse participants (ages and cultures) to assess robustness and ensure inclusivity. Some participants took part in more than one study phase, which may have introduced familiarity effects. Although Study 3 involved a different scenario and a new set of items, trust judgments are closely tied to the immediate interaction, and repeated participation might still have shaped children's comfort or expectations. Future work should therefore aim to replicate these findings with fully independent samples to rule out subtle carryover effects. We chose this design trade-off to maintain comparability across studies in an iterative validation process, while recognising that replication with separate samples is an important next step.

As platforms and interfaces continue to change quickly, future uses of the K-AI scale may also need to refresh the concrete examples, for instance, by substituting child-familiar systems (e.g., "a voice assistant such as ...") without altering the underlying constructs. This will help keep the measure ecologically valid as technologies evolve.

Finally, the instrument is currently scoped to everyday conversational and educational AI. Future studies should examine how well it transfers to other domains, such as clinical or safety-critical settings. The final questionnaire showed acceptable psychometric properties, yet refinement is still needed. Retaining Q4 reflects a commitment to developmental and ethical completeness, even if the item performed weakly. Future work will focus on rephrasing through child interviews and co-design, aiming to improve both statistical robustness and developmental clarity.

The reliance on self-report measures, though adapted, raises concerns about social desirability bias and children's limited metacognitive capacities. Complementary methods such as trust games or observational studies could enhance ecological validity and capture trust in situ.

The cross-sectional design offers only a snapshot of children's responses. As trust evolves with exposure, context, and developmental change, longitudinal research is needed to trace trajectories and examine influences such as digital literacy and negative experiences.

Our findings also show that children's reasoning about ethical principles, transparency, fairness and consent is emerging yet through meaningful relational narratives. Overall, children are capable and reflective users of AI. The proposed questionnaire provides a developmentally grounded tool for evaluation, but continued iteration, combining psychometric rigour with child-led design, is essential to ensure measures remain meaningful, inclusive, and rights-aligned.

Another limitation regards the three trust facets proposed in the conceptual framework. They are grounded in prior work on trust in automation and child-rights principles; however, our data did not support a stable multidimensional structure. Specifically, transparency and agency, as well as emotional comfort, were not captured reliably with the small number of items adopted in this first iteration, and may be more difficult for younger children to articulate after a short interaction. Future work should expand and refine these elements, test age-adapted items, and examine whether additional dimensions such as benevolence or integrity emerge more clearly in longer or repeated interactions. Strengthening these links will help align the proposed facets more closely with existing theoretical models while ensuring they remain developmentally appropriate.

Finally, although we examined the convergent and discriminant properties of both scales, these assessments are preliminary. AVE values for K-AI indicate that additional item refinement is needed to strengthen shared variance at the construct level. Similarly, the weak relation between PTT and K-AI aligns with theoretical expectations but also reflects the limitations of using an adult-designed dispositional trust measure with children. Further work should extend these analyses with larger samples, age-tailored dispositional measures, and comparisons against other child-centred frameworks to consolidate convergent and discriminant validity evidence.

6 Conclusion

Do children trust AI, and should they? Our findings suggest that children aged 9–11 place trust in intelligent systems, but this trust is expressed holistically, blending functional, relational, and ethical cues.

This article introduced the development and validation of a novel tool to assess children's trust in intelligent systems. Through iterative design and testing, we adapted and refined existing frameworks using age-appropriate language and ethically grounded constructs, addressing a critical gap in HCI and child–AI research.

Using a mixed-methods approach, we showed that children's trust can be reliably measured as a unidimensional construct, despite early indications of multidimensionality. The final questionnaire demonstrated acceptable psychometric properties and measurement invariance across age and gender, making it a promising tool for educational, research, and design contexts.

Importantly, our results highlight that children are not passive users of AI. Rather, they evaluate intelligent systems through a combination of social cues, perceived utility, and ethical concerns, grounded in everyday interactions rather than abstract reasoning. Trust thus emerges as a developmental and relational process, shaped by emotional resonance, data transparency, and perceived agency.

Demographic analysis further revealed gendered differences in AI familiarity and system use, underscoring the need for inclusive and context-sensitive design. More broadly, our findings highlight the limitations of adapting adult-oriented measures for child populations and advocate for tools that reflect how children actually experience and make sense of AI technologies.

By combining empirical rigour with developmental sensitivity, this work contributes to the growing movement toward rights-based, child-centred AI design. The proposed questionnaire provides a concrete method for evaluating trust and fostering more meaningful, empowering interactions between children and intelligent systems.

Acknowledgments

This work is partially supported by the co-funding of the European Union—Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 – Partnerships extended to universities, research centres, companies and research D.D. MUR n. 341, 15.03.2022 – Next Generation EU (PE0000013 – "Future Artificial Intelligence Research – FAIR"—CUP: H97G22000210007)

We sincerely thank the Istituto Comprensivo San Francesco d'Assisi in Altamura, Italy, for hosting this research. Our gratitude goes to the students who participated, the teachers for their collaboration, and the headteacher, Pasqua Loviglio, whose support made this study possible.

Appendices

A Prior Knowledge of AI Questionnaire

(1) Have you ever heard of Artificial Intelligence (AI)?
- Yes
- No
- I'm not sure

(2) Have you ever used any of these tools? (You can choose more than one answer)
- Siri (Apple)
- Alexa (Amazon)
- ChatGPT or similar
- Educational Robot (e.g. Cozmo, BeeBot, etc.)
- None of these

(3) When you use these tools, do they seem intelligent to you?
- Yes
- A little
- No
- Not sure

(4) What do you like to use technology for? (you can choose more than one answer)
- To learn
- To play games
- To listen to music
- To talk to friends or family
- Other:

(5) Do you think that what a voice assistant (like Alexa or Siri) says is always true?
- Yes
- No
- Sometimes
- I've never thought about it

B Modified Propensity To Trust Questionnaire

The modified Propensity to Trust questionnaire is shown below. Each question was rated using a 5-point Likert scale, ranging from 'Strongly disagree' to 'Strongly agree'.

(1) I usually trust computers until they make mistakes.
Reflects a conditional form of trust, where the child places initial trust in technology unless a failure occurs.
(2) For the most part, I distrust computers or systems like Alexa, Siri.
A reverse-coded item assessing baseline skepticism toward intelligent systems.
(3) In general, I would rely on a computer to assist me.
Captures the child's openness to depending on technological assistance in various tasks.
(4) My tendency to trust computers is high.
Measures the child's self-perceived dispositional trust in computers.
(5) It is easy for me to trust computers to do their job.
Assesses how naturally or comfortably the child feels when delegating tasks to computers.
(6) I am likely to trust a computer even when I have little knowledge about it.
Explores trust in unfamiliar or opaque technologies, where understanding is limited.

C K-AI Trust Questionnaire (Preliminary version)

Questions 1 to 7 used a five-point ordered response scale with the following choices: 1) Not at all, 2) A little, 3) I don't know, 4) Enough, 5) A lot)

(1) Do you think that smart systems like Alexa or platforms like TikTok are safe to use?
(2) Do you think your personal information (name, home address) is protected when using smart systems?
(3) Do you think that smart systems like ChatGPT are useful?
(4) Would you trust intelligent systems more if you knew exactly how they worked?
(5) Would you trust intelligent systems more if you knew exactly how they were using your data?
(6) Would you feel comfortable if intelligent systems made important decisions about you, such as which school to attend, as long as they respected human rights?
(7) Would you feel comfortable if intelligent systems made important decisions about you, such as which games to play, as long as they respected human rights?

Questions 8 to 11 used a five-point ordered response scale with the following options: (1) never, (2)sometimes, (3) I don't know, (4) often, (5) always

(8) Do you think intelligent systems can make mistakes?
(9) Do you think intelligent systems can harm people?
(10) Do you think that intelligent systems treat all people equally, regardless of gender, age, or skin colour?
(11) Do you think that intelligent systems should ask your permission before using your personal data such as name, surname, address?
(12) Have you used ChatGPT? **(yes/no)**
(13) What have you used ChatGPT for? **(open-ended, thematic analysis)**
(14) Do you think these systems are intelligent? **(used for qualitative coding of perceived system intelligence)**
(15) Would you like to add anything? **(optional comment box for final thoughts)**

D Final K-AI Trust Questionnaire

Questions 1-3 measure Trust / Usefulness on a five-point ordered response scale (not at all, a little, I don't know, enough, a lot).
Questions 4-6 measure Transparency / Awareness / Control on a five-point ordered response scale (never, sometimes, I don't know, often, always).
Questions 7-9 measure trustworthiness on a five-point ordered response scale (strongly disagree, disagree, don't know, agree, strongly agree).

(1) Do you think that smart systems like Alexa or platforms like Tik Tok are safe to use?
(2) Do you think your personal information (name, home address) is protected when using smart systems?
(3) Do you think that smart systems like ChatGPT are useful?
(4) Do you think intelligent systems should ask your permission before using your personal data such as your name, surname, address?
(5) Would you trust intelligent systems more if you knew exactly how they worked?
(6) Would you trust intelligent systems more if you knew exactly how they were using your data?
(7) Do you think intelligent systems can make mistakes?
(8) Would you rely on a computer to help you?
(9) Are you likely to trust a computer even if you don't know how to use it?

References

[1] 5Rights Foundation. 2025. Children & AI Design Code. https://5rightsfoundation.com/children-and-ai-code-of-conduct/ Accessed: 2025-05-08.
[2] Ann M. Arthur, Michelle Howell Smith, Andrew S. White, Leslie Hawley, and Natalie A. Koziol. 2017. *Age-Sensitive Instrument Design for Youth: A Developmental Approach.* Technical Report. Nebraska Center for Research on Children, Youth, Families and Schools.
[3] Ayça Atabey, Cara Wilson, Lachlan D Urquhart, and Burkhard Schafer. 2025. Fairness by Design: Cross-Cultural Perspectives from Children on AI and Fair Data Processing in their Education Futures. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25).* Association for Computing Machinery, New York, NY, USA, Article 1067, 20 pages. doi:10.1145/3706598.3714402
[4] A Bandura and P Hall. 2018. Albert bandura and social learning theory. *Learning theories for early years* 78 (2018), 35–36.
[5] Michael S Bernath and Norma D Feshbach. 1995. Children's trust: Theory, assessment, development, and research directions. *Applied and Preventive Psychology* 4, 1 (1995), 1–19.
[6] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems.* ACM, Montreal, QC, Canada, 1–14.
[7] Raymond R Bond, Felix Engel, Michael Fuchs, Matthias Hemmje, PM McKevitt, Michael McTear, Maurice Mulvenna, Paul Walsh, and Huiru Zheng. 2019. Digital empathy secures Frankenstein's monster. In *Collaborative european research conference.* Collaborative European Research Conference, Springer, Munich, Germany, 335–349.

[8] Miriana Calvano, Antonio Curci, Giuseppe Desolda, Andrea Esposito, Rosa Lanzilotti, and Antonio Piccinno. 2025. Building Symbiotic AI: Reviewing the AI Act for a Human-Centred, Principle-Based Framework. arXiv preprint arXiv:2501.08046. https://arxiv.org/abs/2501.08046
[9] Ki-Taek Chun and John B Campbell. 1974. Dimensionality of the Rotter interpersonal trust scale. *Psychological Reports* 35, 3 (1974), 1059–1070.
[10] Cinzia Di Dio, Federico Manzi, Giulia Peretti, Angelo Cangelosi, Paul L Harris, Davide Massaro, and Antonella Marchetti. 2020. Shall I trust you? From child–robot interaction to trusting relationships. *Frontiers in psychology* 11 (2020), 469.
[11] Digital Future Commission and 5Rights Foundation. 2023. Child Rights by Design. https://5rightsfoundation.com/resource/child-rights-by-design/ Accessed: 2025-05-09.
[12] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How smart are the smart toys? Children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM conference on interaction design and children*. ACM, Trondheim, Norway, 231–240.
[13] KVVB Durgaprasad, Hassan Khalid Abozibid, Jawdat Nasri Hawas, Pusala Bhuvan Sai Krishna, Kanumuri Bhavya Sri, and Gali Pranay Reddy. 2024. AI Agents and Conversation System. In *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)*. IEEE, 1–7.
[14] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O Riedl. 2022. Human-Centered Explainable AI (HCXAI): beyond opening the black-box of AI. In *CHI conference on human factors in computing systems extended abstracts*. ACM, 1–7.
[15] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing transparency design into practice. In *Proceedings of the 23rd international conference on intelligent user interfaces*. ACM, 211–223.
[16] Erik Erikson. 1959. Theory of identity development. *E. Erikson, Identity and the life cycle. Nueva York: International Universities Press. Obtenido de http://childdevpsychology. yolasite. com/resources/theory% 20of% 20ident ity% 20erikson. pdf* (1959).
[17] Erik H. Erikson. 1950. *Childhood and Society*. W. W. Norton & Company, New York, NY.
[18] European Parliament and Council of the European Union. 2023-04-13. Regulation (EU) 2016/679 of the European Parliament and of the Council. https://data.europa.eu/eli/reg/2016/679/oj
[19] Tianna Gadbaw. 2016. Legislative Update: Children's Online Privacy Protection Act of 1998. *Child. Legal Rts. J.* 36 (2016), 228.
[20] Omri Gillath, Ting Ai, Michael S Branicky, Shawn Keshmiri, Robert B Davison, and Ryan Spaulding. 2021. Attachment and trust in artificial intelligence. *Computers in human behavior* 115 (2021), 106607.
[21] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527.
[22] Paul L Harris. 2012. *Trusting what you're told: How children learn from others*. Harvard University Press.
[23] Dagmar Mercedes Heeg and Lucy Avraamidou. 2024. Young Children's Understanding of AI. *Education and Information Technologies* (2024), 1–24. Advance online publication.
[24] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.
[25] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.
[26] Robert Hoffman, Shane Mueller, Gary Klein, and Jordan Litman. 2021. Measuring trust in the XAI context.
[27] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. ACM, 1–16.
[28] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. ACM, 624–635.
[29] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics* 4, 1 (2000), 53–71.
[30] Xiangqi Kong, Yang Xing, Antonios Tsourdos, Ziyue Wang, Weisi Guo, Adolfo Perrusquia, and Andreas Wikander. 2024. Explainable interface for human-autonomy teaming: a survey.
[31] Moritz Körber. 2018. Theoretical considerations and development of a questionnaire to measure trust in automation. In *Congress of the International Ergonomics Association*. Springer, 13–30.
[32] Markus Langer, Daniel Oster, Timo Speith, Holger Hermanns, Lena Kästner, Eva Schmidt, Andreas Sesing, and Kevin Baum. 2021. What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial intelligence* 296 (2021), 103473.
[33] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.
[34] Zhixin Li, Trisha Thomas, Chi-Lin Yu, and Ying Xu. 2024. " I Said Knight, Not Night!": Children's Communication Breakdowns and Repairs with AI Versus Human Partners. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*. ACM, 781–788.
[35] Sonia Livingstone, Mariya Stoilova, and Rishita Nandagiri. 2020. Data and privacy literacy: The role of the school in educating children in a datafied society. *The handbook of media education research* (2020), 413–425.
[36] John Long. 2021. *Approaches and frameworks for HCI research*. Cambridge University Press.
[37] Tambiama Madiega. 2021. Artificial intelligence act.
[38] Bahar Mahmud, Guan Hong, and Bernard Fong. 2023. A study of human–ai symbiosis for creative work: Recent developments and future directions in deep learning. *ACM Transactions on Multimedia Computing, Communications and Applications* 20, 2 (2023), 1–21.
[39] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.
[40] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research* 13, 3 (2002), 334–359.
[41] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The journal of strategic information systems* 11, 3-4 (2002), 297–323.
[42] Stephanie M Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human factors* 55, 3 (2013), 520–534.
[43] Stephanie M Merritt and Daniel R Ilgen. 2008. Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human factors* 50, 2 (2008), 194–210.
[44] Hiroyuki Miki. 2014. User experience evaluation framework for human-centered design. In *International Conference on Human Interface and the Management of Information*. Springer, 602–612.
[45] Office of the High Commissioner for Human Rights. 1989. Convention on the Rights of the Child. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child Accessed: April 2, 2025.
[46] United Nations Convention on the Rights of the Child Committee. 2021. General Comment No. 25 (2021) on children's rights in relation to the digital environment. https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation Accessed 04 April 2025.
[47] Celso Cancela Outeda. 2024. The EU's AI act: a framework for collaborative governance. *Internet of Things* (2024), 101291.
[48] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human factors* 39, 2 (1997), 230–253.
[49] Jean Piaget. 1976. Piaget's theory. In *Piaget and his school: A reader in developmental psychology*. Springer, 11–23.
[50] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and Measuring Model Interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 237, 52 pages. doi:10.1145/3411764.3445315
[51] Grazia Ragone, Zhen Bai, Judith Good, Arzu Guneysu, and Elmira Yadollahi. 2025. Child-centered Interaction and Trust in Conversational AI. In *Proceedings of the 24th Interaction Design and Children*. 1235–1238.
[52] Grazia Ragone, Sara Tibidò, Paolo Buono, Rosa Lanzilotti, Azzurra Ragone, and Maria Teresa Baldassarre. 2025. Towards a Rights-Based Metric Framework for Symbiotic AI Systems Designed for Children. In *Proceedings of the 2025 International Conference on Information Technology for Social Good (GoodIT '25)*. ACM, Antwerp, Belgium. doi:10.1145/3748699.3749816
[53] Christian Remy, Oliver Bates, Alan Dix, Vanessa Thomas, Mike Hazas, Adrian Friday, and Elaine M Huang. 2018. Evaluation beyond usability: Validating sustainable HCI research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 1–14.
[54] Ken J Rotenberg, Claire Fox, Sarah Green, Louise Ruderman, Kevin Slater, Kelly Stevens, and Gustavo Carlo. 2005. Construction and validation of a children's interpersonal trust belief scale. *British Journal of Developmental Psychology* 23, 2 (2005), 271–293.
[55] Julian B Rotter. 1971. Generalized expectancies for interpersonal trust. *American psychologist* 26, 5 (1971), 443.
[56] Julian B Rotter. 1980. Interpersonal trust, trustworthiness, and gullibility. *American psychologist* 35, 1 (1980), 1.

[57] Kristin E Schaefer. 2016. Measuring trust in human robot interactions: Development of the "trust perception scale-HRI". In *Robust intelligence and trust in autonomous systems*. Springer, 191–218.

[58] Nicolas Scharowski, Sebastian AC Perrig, Lena Fanya Aeschbach, Nick von Felten, Klaus Opwis, Philipp Wintersberger, and Florian Brühlmann. 2024. To trust or distrust trust measures: Validating questionnaires for trust in AI. *arXiv preprint arXiv:2403.00582* (2024).

[59] Patrick Shafto, Baxter Eaves, Daniel J Navarro, and Andrew Perfors. 2012. Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental science* 15, 3 (2012), 436–447.

[60] Matthias Sutter and Martin G Kocher. 2007. Trust and trustworthiness across different age groups. *Games and Economic behavior* 59, 2 (2007), 364–382.

[61] James L Szalma and Grant S Taylor. 2011. Individual differences in response to automation: the five factor model of personality. *Journal of experimental psychology: Applied* 17, 2 (2011), 71.

[62] Małgorzata Szcześniak, Melusina Colaço, and Gloria Rondón. 2012. Development of interpersonal trust among children and adolescents. *Polish Psychological Bulletin* 43, 1 (2012), 50–58.

[63] JMF ten Berge. 1995. Review J. C. Nunnally and I. H. Bernstein. Psychometric Theory. *Psychometrika* 60, 2 (June 1995), 313–315. doi:10.1007/BF02301419

[64] Isabel Thielmann and Benjamin E Hilbig. 2015. The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin* 41, 11 (2015), 1523–1536.

[65] Daniel Ullman and Bertram F Malle. 2018. What does it mean to trust a robot? Steps toward a multidimensional measure of trust. In *Companion of the 2018 acm/ieee international conference on human-robot interaction*. 263–264.

[66] UNICEF. 2024. Developing Global Guidance for Child Rights Impact Assessments in Relation to the Digital Environment. https://www.unicef.org/reports/CRIA-responsibletech

[67] UNICEF. 2025. *TBest Interests of the Child in Relation to the Digital Environment. Working Paper.* https://www.unicef.org/innocenti/innocenti/reports/best-interests-child-relation-digital-environment

[68] UNICEF Innocenti. 2021. Policy Guidance on AI and Children. https://www.unicef.org/innocenti/reports/policy-guidance-ai-children Accessed: April 2, 2025.

[69] Jessica Van Brummelen, Mingyan Claire Tian, Maura Kelleher, and Nghi Hoang Nguyen. 2023. Learning affects trust: Design recommendations and concepts for teaching children—and nearly anyone—about conversational agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. ACM, 15860–15868.

[70] Ming-Yan Wang, Peng-Zhu Zhang, Cheng-Yang Zhou, and Neng-Ye Lai. 2019. Effect of emotion, expectation, and privacy on purchase intention in WeChat health product consumption: The mediating role of trust. *International journal of environmental research and public health* 16, 20 (2019), 3861.

[71] William M. Webb and Philip Worchel. 1986. Trust and Distrust. In *Psychology of Intergroup Relations*, S. Worchel and G. Austin, W. (Eds.). Nelson-Hall, Chicago, 334–362.

[72] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. 2021. "Let me explain!": exploring the potential of virtual agents in explainable AI interaction design. *Journal on Multimodal User Interfaces* 15, 2 (2021), 87–98.

[73] Ying Xu, Trisha Thomas, Chi-Lin Yu, and Echo Zexuan Pan. 2025. What makes children perceive or not perceive minds in generative AI? *Computers in Human Behavior: Artificial Humans* 4 (2025), 100135.

[74] Toshio Yamagishi, Satoshi Akutsu, Kisuk Cho, Yumi Inoue, Yang Li, and Yoshie Matsumoto. 2015. Two-component model of general trust: Predicting behavioral trust from attitudinal trust. *Social Cognition* 33, 5 (2015), 436–458.

[75] Yi Zeng, Enmeng Lu, and Kang Sun. 2025. Principles on symbiosis for natural life and living artificial intelligence. *AI and Ethics* 5, 1 (2025), 81–86.