

Supplementary Materials

Data collection methods

The primary data consisted of open-ended written responses to a survey item that asked participants to define trust in artificial intelligence. An initial pilot phase ($n \approx 49$) informed the refinement of the analytic procedures and the development of the codebook. Pilot responses were retained and included in the final dataset because no substantive modifications were made to the survey items. Pilot data were collected in May 2025. Prolific data collection was completed in June. Undergraduate data were collected in July and August 2025. Formal coding of the Prolific dataset began in late June and for the second recruitment source in September 2025. After expanding the sample, the final dataset was analyzed using qualitative content analysis. The survey instrument remained unchanged throughout the recruitment phases. The iterative recruitment strategy enabled greater conceptual breadth and response diversity. This strengthened the robustness of the identified themes. Formal coding of the finalized dataset ($N = 204$) began after the data collection was completed. This staged approach ensured sufficient variability and analytical robustness in identifying recurring conceptual dimensions of trust in AI.

Data processing

The responses were exported in text format for analysis. As the responses were provided in written form, no transcription was required. Prior to the analysis, the data were screened for completeness and substantive content. Responses identified as non-substantive (e.g., repeated characters, unintelligible entries) or clearly non-original AI-

generated text were excluded from the analysis.

List of excluded responses

For the item measuring trust in general, seven participants' responses were excluded because they were identified as AI-generated and all began with the same verbatim text; two responses were excluded because they were off-topic (e.g., "hhhhhhh"); two responses were removed because they were duplicates; and 18 responses had missing values. In total, 29 participants were excluded from the analysis. Therefore, the final sample for the general trust item consisted of 175 responses.

For the item measuring trust in AI, two responses were excluded because they were identified as AI-generated, and 10 responses were excluded because they did not answer the target behavior (for example, "yuou xjgjbk.jm,.", "no"), and 23 were missing values. In total, 35 participants were excluded from the analysis. Therefore, the final sample for the Trust in AI item consisted of 169 responses.

The cleaned data were imported into NVivo for qualitative content analysis.

Data analysis

Data were analyzed using an inductive qualitative content analysis. This approach was selected to systematically identify recurring conceptual dimensions within the participants' definitions of trust in artificial intelligence while remaining grounded in textual data.

The analysis proceeded in multiple stages. During the pilot phase, the first author conducted exploratory coding to familiarize themselves with response variability and generate preliminary codes. These preliminary codes informed the development of a draft codebook that was used to train additional coders.

Following the expansion of the dataset, a team of four coders independently applied the draft codebook to a subset of responses. The coders were encouraged to refine the existing codes, merge overlapping categories, and introduce new codes as needed. Through iterative discussions, the codebook was revised to reflect the patterns emerging across the full dataset. Once a consensus was reached on the code definitions, the final codebook was applied to all responses. The codes were subsequently grouped into higher-order conceptual dimensions based on semantic similarity and thematic coherence. These dimensions represent recurring definitional patterns in the participants' conceptualization of trust in AI.

Qualitative content analysis was appropriate for this study because the research objective was to identify recurring conceptual elements in short, definition-focused survey responses, rather than to develop formal theory or analyze lived experience narratives.

This study was conducted within a post-positivist qualitative framework. While recognizing that textual interpretation involves researcher judgment, we assumed that systematic coding procedures and multi-coder reliability assessment could approximate recurring conceptual patterns in the participants' definitions. (see the Data Availability Statement).

Units of study

At the outset, 204 participants were recruited via the online survey platform Prolific ($N=100$) and from the undergraduate participant pool ($N=104$). Responses were voluntary, and participants could skip questions. Prolific participants were required to be at least 18 years old, have English as their first language, have completed at least 100 Human Intelligence Tasks (HITs), and maintain an HIT approval rate above 97%. They were compensated with £0.80, whereas student participants received 0.5 course credits for their participation.

After exclusions (due to failure to complete or nonsensical responses), the final analytic sample consisted of 169 participants, with responses ranging from 1 to 101 words ($M = 15.32$ words), yielding a total corpus of 2,589 words. Approximately 14% of the responses contained fewer than five words, 63% ranged between five and 20 words, and 23% exceeded 20 words, indicating variability in response depth. Demographic information on the sample (age, gender, and education level) is presented in the results section.

The final analytic sample for the general trust item consisted of 175 responses ranging from 1 to 85 words ($M = 12.65$ words), yielding a total corpus of 2,215 words. Approximately 17% of the responses contained fewer than five words, 69% ranged between five and 20 words, and 14% exceeded 20 words, indicating variability in response depth.

Techniques to enhance data analysis trustworthiness

Several procedures were implemented to enhance the analysis trustworthiness. First, a multi-coder approach was used to reduce individual interpretive biases. A structured codebook was developed iteratively and refined through team discussions. Inter-coder reliability was assessed using Cohen's kappa, with values ranging from 0.68 to 0.78, which revealed substantial agreement.

Second, analytic decisions were documented throughout the coding process, creating an audit trail of the code revisions and category development. Discrepancies were resolved through consensus discussions to ensure consistent code application. These documents are available for review (see the Data Availability Statement).

Planned analyses

This content analysis was conducted in three primary phases: preparation, organization, and reporting of the results (see Fig. 1). The preparation phase began with the selection of the unit of analysis. In this study, the unit of analysis consisted of meaningful units (or semantic units) extracted from participants' written responses. For example, if a participant wrote, 'I trust ChatGPT because it is transparent, accurate, and unbiased,' we would not code the entire response as one unit; rather, we would code it in three units (transparency, accuracy, and unbiased).

The coders reviewed the raw responses multiple times and recorded their initial categories. Qualitative inductive content analysis was employed, enabling the creation of new categories that emerged from participants' responses when the data did not fit existing frameworks. After the initial conceptual open coding, inter-coder reliability checks were conducted to ensure the consistent application of codes by all four coders.

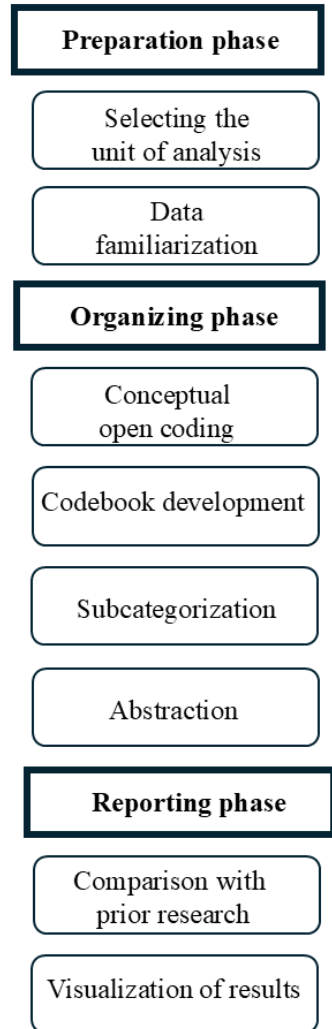


Fig. 1 Workflow of the qualitative content analysis procedure used in this study. The process included preparation, organizing, and reporting phases

The coding process was iterative, enabling the refinement of both predefined and emerging categories as the analysis progressed. Following the initial conceptual coding, the categories were further refined by identifying their internal dimensions. For instance, the category ‘Transparency’ was subdivided into areas such as transparency related to sources, data usage, and information-sharing policies. Related categories were grouped

into broader themes through iterative categorization, sub-categorization (Fig. 2), and abstraction (Fig. 3). The results are presented in hierarchical charts.

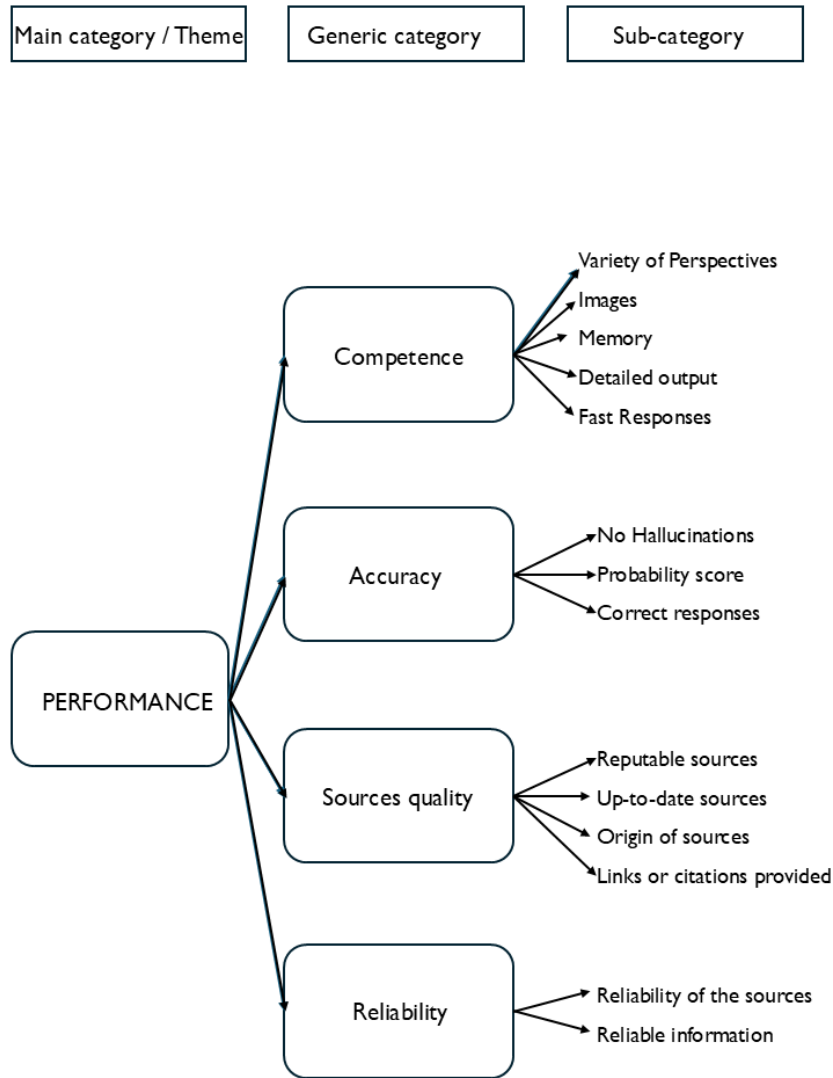


Fig. 2 An example of the sub-categorization process

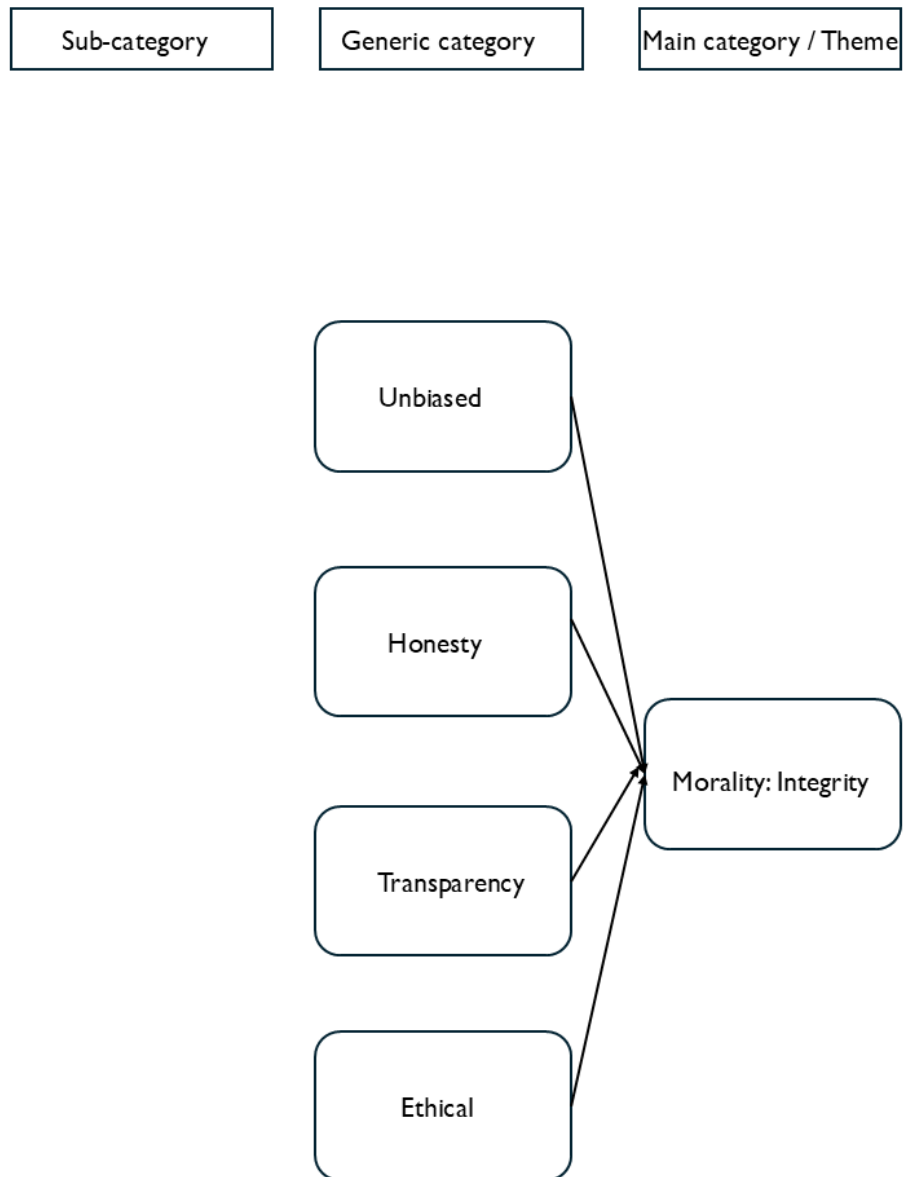


Fig. 3 An example of the abstraction process

The coding team consisted of the principal researcher and three volunteer research assistants. Coders were introduced to NVivo through a review of the starter codebook (available in the Supplementary Materials) and instructional videos on creating parent codes, child codes, and uncode selection in NVivo. The initial round of coding focused on

the item *general trust* (“How do you define the word trust in general?”) to familiarize the team with coding. The starter codebook and its existing concepts were utilized; at this stage, the inductive, individual development of new codes was permitted to capture emergent concepts not present in the starter codebook. A subsequent review of the initial coding round included revisions to the coding procedure and a discrepancy analysis, resulting in the development of the initial formal codebook. This document standardizes the general coding procedure through uniform code names, operational definitions, and instructions for code use. Using the refined formal codebook, the item *general trust* was recoded; this round did not permit the development of new codes. Inter-rater reliability and agreement analyses demonstrated significant improvement and strong agreement among the coders.

Further amendments and the introduction of new codes to the codebook were made to capture novel concepts, as additional questions (i.e. “What does trust mean to you specifically when interacting with an AI assistant?”) were coded. The agreement between coders was assessed using coding comparison queries that merged the encoded data. This macro-level assessment led to major revisions in the codebook, primarily at the interpretive level, including updates to code definitions, thematic structure, inclusion and exclusion criteria, and the development of subcodes. Following the codebook revisions, full dataset coding was conducted, with all participants’ responses assigned to individual assistants for independent coding. In the final analytic phase, a detailed review of individual participant responses across all datasets was conducted to ensure rigor and alignment between the coded data, final codebook definitions, and rater interpretations.

To enhance transparency, an audit trail and peer debriefing were conducted. Specifically, the audit trail generated a report detailing the decision-making process for each step of coding, data analysis, and interpretation. The research team also maintained detailed analytic and reflective memos to document the process and progress. These strategies increased the dependability of the findings by employing a systematic documentation approach.

Unlike previous studies that have examined trust in AI using predefined theoretical dimensions or structured questionnaires, this study adopts an inductive bottom-up qualitative approach. This approach enabled the identification of trust dimensions that may have not been captured by traditional deductive models.