Online Web Appendices

Appendix A. Literature Review

Table A.1. Generative AI versus Predictive AI: Differences and Similarities

Categories	Generative AI[a]	Predictive AI[b]
Differences		
Focus/Goal	• Create new content or data by simulating the joint probability distribution of the input data • Open-ended problem-solving	• Predict future trends by learning the conditional probability distribution of the trained data • Routine decision-making
Technical underpinnings	• *Transformer architecture*. Utilize self-attention mechanisms and positional encoding to capture long-range dependencies and generate coherent content • *Fine-tuning and alignment*. Learn with human-annotated data sets and alignment techniques to match human values and expectations	• *Machine learning*. Use supervised learning to learn input-output relationships from labeled datasets for prediction • *Model selection and optimization*. Select suitable models and optimize parameters tailored to scenarios and datasets to enhance prediction
Inputs	• Unlabeled/unstructured data (mostly)	• Labeled/structured data (mostly)
Outputs	• Context-sensitive, prompt-based outputs • Novel/original/creative content • Plausible-sounding but unfaithful outputs	• Rule-based outputs/predictions • Frequently higher accuracy
Application	• Content/dialogue/code generation, abstractive summarization, product design, art creation …	• Risk/credit assessment, market prediction, medical diagnosis, recommendation system …
Algorithm properties	• Algorithms/models are self-adaptive and feedback-dependent, with higher flexibility and generality, tailored to contextual factors and personal perspectives	• Algorithms are often stabilized and follow the learned rules, with formal procedure and established percept
Model Explanation mechanism	• Chain-of-Thought reasoning • Self-reported reasons • Conversational interpretation/interactive explanations	• Feature-based explanations • Rule-based explanations • Local/global explanations

Role of the human	• *Prompting*. Provide LLMs with textual instructions or prompts to elicit/guide outputs • *Feedback and iteration*. Provide feedback according to the output, helping the LLM learn to improve • *Optimizing and customizing workflows*. Control LLMs to fit a self-workflow via prompt-based interactions	• *Data preparation and cleaning*. Provide high-quality data sets by collecting, classifying, cleaning, and labeling data • *Algorithm maintenance*. Choose significant parameters and relevant context during algorithm development
Human cognition mechanism	• *Metacognitive monitor*. Assess one's own thinking to have a clear self-awareness of specific goals and intentions for prompting AI systems • *Metacognitive control*. Guide one's own thinking to adaptively and flexibly change cognitive or prompting strategies	• *Metacognitive monitor*. Assessment of how well a task is, will be, or has been performed by oneself or the AI system • *Metacognitive control*. Frame elaboration, information inference and selection, and asking for additional information
Similarities		
Goal	• Analyze data and implement algorithms to fit specific expectations and achieve target goals.	
Usability challenges	• *Transparency and explainability*. Because the generation/prediction process involves complex algorithms and large data sets, it is difficult to trace the path of their outputs and explain the reasoning in the process. • *Blind trust and overreliance*. People frequently rely on AI's advice without sufficient justification and critical evaluation, ignoring their own judgment and experience, and resulting in suboptimal or even wrong decisions.	
Human-centered strategies	• *Clear awareness*. Assess one's own knowledge and experiences (task-related) and consider subjective beliefs of AI capabilities before decision-making. • *Initiative engagement*. Enhance cognitive effort and analytic thinking in the decision-making process, evaluating the effectiveness and accuracy of generative/predictive AI outputs. • *Monitoring and controlling*. Evaluate the correctness of the decision-making process of both generative/predictive AI and oneself, and make adjustments or corrections when necessary.	

Notes. This summarization was constructed based on the literature review of Vaswani (2017), Jussupow et al. (2021), Dwivedi et al. (2023), Ji et al. (2023), Chung et al. (2024), and Steyvers and Kumar (2024). [a] We focus on generative conversational AI, such as ChatGPT and DeepSeek. These GenAI-powered chatbots do well in content generation and query answering through natural language processing and deep learning. [b] The conceptualization of predictive AI mainly considers decision-support and recommendation systems.

Table A.2. Main Findings on CoT: Technical and Interactive Perspectives

Perspectives	CoT Role	Main Findings	Resource
Technique	*CoT Reasoning*: Serving as the utterance-based reasoning process that externalizes model's decision-making pathways,	CoT-generated stated values (e.g., fairness, altruism, self-interest) statistically predict final actions with high internal consistency, and post-	Leng & Yuan (2026)

	enabling the SUVA framework to extract stated values (V) and understanding (U) from the CoT text.	training alignment methods systematically reshape both CoT content and downstream decisions.	
Interaction	*CoT Prompt Coaching*: Guiding step-by-step reasoning, enhancing diversity, trust, and user control.	CoT prompt coaching increased perceived contingency, user agency, and cognitive elaboration, leading to higher perceptions in terms of trust, trust calibration, and ease of use in image generation tasks.	Chen & Zhao (2025)
Technique & Interaction	*CoT Reasoning*: Serving as a practical intervention to increase semantic breadth and idea diversity.	Using CoT prompting, asking the LLM to follow distinct, ordered steps when generating, expanding, and revising ideas, improves the diversity and quality of generated ideas compared to using a base prompt alone.	De Freitas et al. (2025)
Interaction	*CoT Conjoint*: Assessing LLMs' ability to emulate human intertemporal preferences through CoT reasoning.	LLMs (GPT-3.5/4) show less patience than humans in intertemporal choices, with GPT-4 mitigating discrepancies via CoT conjoint prompts. Language structure influences discount rates, aiding hypothesis generation for preference heterogeneity.	Goli & Singh (2024)
Technique	*CoT Prompting*: Quantifying and optimizing CoT reasoning capabilities in LLMs by defining reasoning boundaries.	Introducing a novel reasoning boundary framework that reveals CoT optimization strategies via boundary promotion and path optimization, demonstrating performance improvements across 27 LLMs and 5 tasks.	Chen et al. (2024)
Technique	*CoT Prompting*: Facilitating step-by-step reasoning in LLMs, aiming to explain model predictions transparently.	CoT explanations can be systematically unfaithful, influenced by input biases, and fail to mention critical biases affecting predictions, posing risks to AI safety and trustworthiness.	Turpin et al. (2023)
Technique	*CoT Prompting*: Enabling stepwise task decomposition and interpretability is a side effect.	AI chains improve interpretability by revealing plausible reasoning paths, aiding debugging, yet they don't fully explain the model's computational process, posing ongoing research challenges.	Wei et al. (2022)
Technique & Interaction	*Chaining LLM*: Decomposing tasks, enhancing transparency, control, and collaboration.	Chaining model steps improves task outcomes, user satisfaction, and transparency. Users leveraged modular steps for debugging, collaboration, and mental support, shifting from passive to active editing.	Wu et al. (2022)
Technique	*CoT Prompting*: Enhancing scientific table understanding via type-guided CoT prompting.	Integrating a fine-tuned model with type-guided CoT prompts significantly enhances the accuracy and efficiency of scientific table processing.	Yin & Wang (2025)

Table A.3. Pertinent Empirical Research on Explanations and Trust

Manipulation of Explanations	Evaluation of Trust	Main mechanisms	Pertinent Findings	Resource
Absent vs. input- vs. group- vs. case- vs. counterfactual-based explanations	Self-reported trust in AI	None	Offering users some types of explanation can positively impact their perceptions of trustworthiness toward an AI decision, but their effectiveness varies across decision-making contexts.	Aoki et al. (2024)
Absent vs. local, feature-based explanations	Self-reported trust in AI	Situational information processing; Mental model adjustments	Note the double-edged sword of AI explanations: they can help users learn new feature-label relationships, but can also risk overconfidence and overreliance due to confirmation biases.	Bauer et al. (2023)
Absent vs. honest vs. deceptive explanations	Self-reported trust in AI: ability, integrity, and benevolence	Perceived persuasion; Critical thinking	Deceptive AI-generated explanations are more persuasive than honest ones or mere results, with higher self-reported knowledge paradoxically predicting increased vulnerability. Yet, logically invalid explanations lose their effects.	Danry et al. (2025)
Absent vs. present explanations	Self-reported confidence in LLM-generated answers	Justification quality; Actionability	While explanations generally increase user reliance on both accurate and inaccurate outputs, inconsistencies within them mitigate this effect on wrong answers. Moreover, LLM-literate users are better at perceiving explanations as mere plausibility, not true reasoning.	Kim et al. (2025)
Sequentially vs. concurrently vs. dynamically presenting explanations	Trust calibration (inferred)	Cognitive load; Information overload	Selectively and adaptively presenting LLM-powered analyses or explanations can decrease users' cognitive load and enhance their appropriate reliance on the AI model's decision recommendations.	Li et al. (2025)
Single explanations vs. recommendation and explanations	Self-reported trust in AI	Cognitive forcing	When human decision-makers are likely to be correct, hiding AI suggestions or offering only explanations to promote independent thought tends to reduce overreliance.	Ma et al. (2023)
Anthropomorphic explanations: process and outcome variation	Self-reported trust in AI: competence, integrity, and benevolence	Perceived anthropomorphism	The co-presence of process and outcome variation can enhance trust through increased perceptions of anthropomorphism. Yet it may also lead users to excuse system failures as exploratory behavior rather than recognizing genuine malfunctions.	Saffarizadeh & Keil (2026)

Explanation style: uncertainty language (low/medium/high confidence) and length (short/long)	Self-reported confidence in LLM-generated answers	Calibration gap; Discrimination gap; Length bias	Default, longer explanations lead users to overestimate LLM accuracy without enhancing discernment. Aligning uncertainty language with model confidence significantly reduces both calibration and discrimination gaps, improving user perceptions of LLM accuracy.	Steyvers et al. (2025)
Absent vs. global explanations	Self-reported cognitive trust in AI	Metacognitive monitoring (calibration and resolution); Metacognitive control (human reasoning and AI logic)	AI explanations can improve effective delegation decisions by helping users better calibrate their confidence about their own ability to perform a task, particularly by reducing their overconfidence when their own reasoning differs from AI logic.	von Zahn et al. (2025)
Absent vs. present explanations	Self-reported trust in AI	Intrinsic cognitive load; Germane cognitive load; Perceived task complexity; Cognitive ability	Even objectively clear explanations can backfire: they increase perceived complexity and reduce compliance, except for users with high cognitive ability, who show improved perceptions of understanding and trust.	Westphal et al. (2023)

Table A.4. Application of the HSM in Human-AI Contexts

Contexts	Heuristic Pathways	Systematic Pathways	Main Findings	Resource
Short-form videos: AI disclosure and user engagement	Heuristic cues associated with *AI novelty* and *technological advancement* enable users' curiosity and positive affect.	Systematic scrutiny of *AI's capabilities* (reliability, flexibility, and integrability) and *perceived content quality*.	AI disclosure has a direct positive effect on user engagement intention. When perceived AI capabilities are high, the negative effect of AI disclosure on content quality is mitigated, and the indirect effect on engagement becomes positive.	Chen et al. (2025)
AI content moderation: AI trust	*Positive machine heuristic* (e.g., objective, accurate, unbiased) versus *negative machine heuristic* (e.g., lack of subjective judgement, inability to detect semantics).	*Interactive transparency*: affording users the option to provide feedback to the AI system by suggesting characteristics to include or exclude when making a classification decision.	In AI-based content moderation, trust depends on users' positive or negative heuristic beliefs regarding machines, while interactive transparency can enhance trust by increasing users' agency.	Molina & Sundar (2022)

Social support chatbots: evaluations and perceived emotional validation	A chatbot's social role cue (*expertise label*: expert vs. non-expert).	A chatbot's social dynamic cue (*interactional contingency*: contingent vs. generic feedback)	While the chatbot's expertise label significantly shaped pre-interaction expectations for sincerity and caring, contingent feedback during the actual interaction overwhelmingly dominated post-interaction evaluations of trust, sincerity, caring, understanding, and emotional validation.	Rheu et al. (2024)
Recommendation systems in travel planning: AI adoption intention	*Anthropomorphism* and *social influence*, enabling emotional trust.	*Performance efficacy* and *perceived personalization*, enabling cognitive trust.	Both heuristic and systematic cues jointly determine trust and adoption intentions (additivity effect). Emotional trust has a stronger impact on delegation intention, while cognitive trust is more influential for decision aid intention (attenuation effect).	Shi et al. (2021)
Algorithm-based systems: AI trust and adoption	*Fairness*, *accountability*, and *transparency* (FAT) serve as mental shortcuts that shape initial trust judgments.	*Personalization* and *accuracy* as systematic cues: perceived personalization influences usefulness, while perceived accuracy affects convenience.	FAT significantly predicts trust, which in turn influences perceived usefulness and convenience. Further, usefulness and convenience shape attitude, and attitude strongly predicts actual use behavior.	Shin et al. (2020)
Online dating platform: AI trust and adoption	Expectation confirmation: Relying on *simple cues* and *low transparency* (concise explanations) enhances perceived understanding.	Expectation disconfirmation: Shifting to systematic processing and *high transparency* (detailed explanations) improves perceived understanding.	Perceived understanding is identified as a key mechanism linking transparency to trust: under disconfirmation, high transparency increases understanding and trust; under confirmation, low transparency is more effective.	Sun et al. (2026)
AI for human resource management (interview, task assignments, performance assessment): AI adoption	*Leaders' AI symbolization*: Applying a simple rule ("if my leader champions it, it is safe and legitimate"), requiring minimal cognitive effort.	*AI transparency* (understandability, explainability, observability): Enabling two parallel mediating pathways: (a) cognitive pathway (AI job role clarity) and (b) affective pathway (reduced discomfort).	High (vs. low) AI transparency enhances change readiness for AI adoption through AI job role clarity and reduced discomfort; yet the two mediating paths are significant only when leaders' AI symbolization is low, supporting that strong heuristic cues can substitute for systematic processing.	Zhang et al. (2026)

Appendix B. Experimental Settings

Table B.1. Definitions of Key Constructs Used in This Study

Construct	Definition and Measurement
Perceived task difficulty	Understood through the cognitive demands it imposes in the context of cognitive load theory, specifically the inherent difficulty of the task itself, including the number of elements and their interactivity (Sweller, 2010). [Measured through three items on seven-point Likert scales.]
Perceived CoT quality	Users' perceptions of CoT's completeness, logical coherence, and alignment with its output (Wu et al., 2022). [Measured through three items on seven-point Likert scales.]
Perceived self-efficacy	Users' belief in their own ability to succeed in specific situations or accomplish particular tasks. It is the confidence a person has in their capacity to execute behaviors, actions, or strategies required to produce a desired outcome (Bandura, 1977). [Measured through three items on seven-point Likert scales.]
Perceived transparency	Users' perceptions of the observable, explainable, and understandable aspects of GenAI's reasoning processes (Shin et al., 2020). [Measured through three items on seven-point Likert scales.]
Perceived diagnosticity	The extent to which people interpret a piece of information, i.e., the AI-generated alone summary in the control condition or CoT-guided summary in the experimental condition, as useful and relevant according to their judgment and decision-making (Pham & Avnet, 2004). [Measured through three items on seven-point Likert scales.]
Germane cognitive load	Associated with the process of constructing new knowledge or skill development in the context of cognitive load theory (Sweller, 2010). [Measured through three items on seven-point Likert scales.]
Perceived trust in GenAI's competence	The extent of users' perceived competence of GenAI in decision-making, including its ability, capability, and good judgment (McKnight et al., 2002). [Measured through four items on seven-point Likert scales.]

Table B.2. Comparisons Between Lower and Higher Difficult Task Conditions

Dimensions	Sub-dimensions	Low-difficulty Task	High-difficulty Task
Brief descriptions of news content		A Sydney woman was caught in a superstorm. A tall, charming gentleman offered her his umbrella. She is now posting "Love Call" notices to find him.	Rod Blagojevich, a former Illinois governor, was imprisoned for corruption. The document details his prison life, his attempt to appeal his conviction, and the legal arguments surrounding his case.
Component complexity	Information cues	Short news text (178 words) Approx. 7-10 key cues: - Entities: MissieSmile21, the gentleman - Location: George Street, Australia Square - Time: Tuesday, 7:10 pm - Events: Superstorm, umbrella offer - Status: Searching, result unknown	Long news text (437 words) Approx. 15-18 key cues: - Entities: Blagojevich, Obama, lawyers, court - Locations: Illinois, FCI Englewood - Times: 2008, 2009, 2014, present - Events: Senate seat sale, impeachment, conviction, appeal, McCutcheon ruling - Status: Serving sentence, appeal pending
	Acts	Approx. 5 acts: - Understand storm scene - Understand umbrella offer - Understand poster action - Infer intention (searching) - Understand outcome (unknown)	Approx. 8 acts: - Understand political background - Understand the "seat sale" charge - Understand impeachment - Understand conviction/sentencing - Understand appeal (McCutcheon) - Understand the government's rebuttal - Understand current prison status - Understand pending appeal
Coordinative complexity	Information relationships	Linear causal chain: Storm → umbrella offer → posters → search	Networked causal chain: Seat sale → impeachment → conviction → prison & appeal → McCutcheon case → Govt. rebuttal → pending
	Information integration	Chronological understanding: information is consistent and non-conflicting.	Requires integrating multiple timelines (2008–2015), multiple actors (defense vs. prosecution), and conflicting viewpoints.
Dynamic complexity		Task is static; information is presented once.	Task is static; information is presented once.

Notes. The manipulation of the task difficulty in this study was based on Wood's (1986) three complexity dimensions: (1) *Component complexity* means the difficult task requires processing nearly double the number of information cues across a wider range of entity types; (2) *Coordinative complexity* means the difficult task shifts from a linear narrative to a networked causal web, requiring users to integrate conflicting information and switch between multiple timelines; and (3) *Dynamic complexity*. Both tasks are externally static in this study.

Table B.3. Experimental Materials

[1] Pre-test (BBC XSum)

Source Document: The 69-year-old's yamaha collided with a Nissan car between handley's corner and barre garroo crossroads at about 17:00 bst on 4 June. Mr. Baker, who was from the island, was airlifted to noble's hospital, where he later died. Insp Dave Dobbie said "thoughts and condolences are with Mr. baker's family". He added that investigations are ongoing in relation to the crash. The car driver, who police say was northern Irish, was treated in hospital but has been discharged. Another motorcyclist who was injured after the crash has also been released from hospital. The collision happened outside of the TT racing, which is taking place on the island.

Summary Choices:

1. On 4 June, Mr. Baker's motorcycle collided with a car, resulting in his death. The car driver and the motorcyclist were injured. Investigations into the crash are done. [Factual hallucination]

2. On 4 June, Mr. Baker's motorcycle collided with a car, resulting in his death. The car driver and the motorcyclist were injured. [Correct summary]

3. Mr. Baker's motorcycle collided with a car, resulting in his death. The car driver and the motorcyclist were injured. [Missing the "date" element]

[2] Low-difficulty Task: Social News (BBC XSum)

Source Document: They say every cloud has a silver lining. At least that's what one Sydney woman, caught out in this week's super storm, is hoping. With the rain clearing, a woman known only by her email address as MissieSmile21, has taped posters to the city's light posts to find the 'tall, charming gentleman' who offered her his umbrella to cross the street. The 'Love Call' notes are posted at traffic lights on Sydney's George Street near Australia Square. Actively looking: Missie Smile is hoping the generous stranger will get in contact. The 'Love Call' notices, posted on George Street at Australia Square in Sydney's CBD, are appealing to the man with 'the most beautiful smile' who was in the area about '7:10 pm-ish' on Tuesday evening. 'I am actively looking for you,' the poster reads. 'If you're single, let's grab a drink together.' People struggled against the heavy rain and strong winds in the Sydney CBD as the storm lashed NSW on Tuesday this week. Most people gave up on their umbrellas in winds of more than 100 km/hr.

High-quality CoT:

1. The important entities in this document are MissieSmile21, the tall, charming gentleman, and George Street at Australia Square.

2. The important dates in this document are Tuesday evening and 7:10 pm.

3. The events happening in this document are the super storm and MissieSmile21's search for the tall, charming gentleman.

4. The result of these events is unknown.

Correct Summary:

On Tuesday evening, a super storm hit Sydney, Australia. In the midst of the storm, a tall, charming gentleman offered his umbrella to a woman known only as MissieSmile21. MissieSmile21 is now actively searching for the gentleman, and has posted "Love Call" notices around George Street at Australia Square in Sydney's CBD in the hopes of finding him. The result of these events is unknown. [7 valid elements; coherent; no redundant information.]

Low-quality CoT:

1. The important entities in this document are the poster and the man she is looking for. [Partially correct. The *poster* serves as a means for the searching effort rather than being a core entity, and the key location, i.e., George Street, is omitted.]

2. The important dates in this document are Tuesday, July 10th. [Seriously incorrect. The news article does not mention a specific month or date, i.e., *July 10th* is fabricated.]

3. The events happening in this document are the heavy rain, and strong winds. [Incomplete. It only describes the background weather conditions while completely omitting the core event, i.e., MissieSmile21's search for the tall, charming gentleman.]

4. The result of these events is many people were unable to use their umbrellas and many people were stranded. [A reasonable inference. It extends from the content of "people struggled" and "gave up umbrellas," though the use of "stranded" is a slight over-interpretation.]

[Note: Overall, many redundant or false facts are shown in gray highlights.]

Incorrect Summary:

Amid the heavy rain and strong winds of the super storm on Tuesday, July 10th, a Sydney woman posted posters to find the tall, charming gentleman who offered her his umbrella when many others were unable to use theirs. [3 valid elements (yellow colored); coherent; redundant information: *heavy rain* and *strong winds* of the *super storm*.]

[3] High-difficulty Task: Policy News (CNN Daily Mail)

Source Document: Once famed for his mop of blacker than black hair, disgraced Democrat Rod Blagojevich, 58, has really let his haircare regime go while he serves his prison time. The former Illinois governor has return to his roots while inside and has been photographed with his still full head of hair a shocking white color rather than the boot polish black that was his trademark as a politician. Blagojevich was infamously caught trying to sell Barack Obama's U.S. Senate seat when he was elected president in 2008. Back in his days as governor of Illinois, Blagojevich was famed for his boot polish black hair. He was impeached and removed from office by the state Legislature in early 2009 following his arrest on federal corruption charges. Blagojevich is currently serving a 14-year sentence at the Federal Correctional Institution Englewood near Denver. The photos, the first of Blagojevich in prison, were first published by the National Enquirer. The photos show the former governor sitting at a desk, reading a book and wearing glasses and sitting outside in shorts and a wife beater shirt. While he continues to awaiting word on a last-ditch appeal, Blagojevich has daily contact with his family and teaches the history of war battles to other inmates, his attorney said. Last July lawyers attempted to argue that a high court ruling in McCutcheon v. the Federal Election Commission in April 2014 bolstered their view that Blagojevich was engaged in legal, run-of-the-mill political horse trading, not corruption. The McCutcheon ruling, they argued, found that soliciting contributions crosses the line into corruption only when a politician makes a clear, explicit promise to take official action in return for a donation. Blagojevich is currently serving a 14-year sentence at the Federal Correctional Institution Englewood near Denver after he was caught trying to sell Barack Obama's U.S. Senate seat when he was elected president. In that context, defense attorneys said the trial judge was wrong to set a lower standard for jurors by telling them, to convict, they only needed to find that Blagojevich sought donations 'knowing or believing' it would be given for some official act. But government attorneys responded saying that Blagojevich's legal team hadn't correctly characterized the Supreme Court's findings in McCutcheon. 'The decision provides no support for Blagojevich's argument on appeal,' the two-page filing said. 'Nothing in the (McCutcheon) decision suggests that an exchange of contributions for specific official acts is quid pro quo corruption only if the arrangement is stated 'explicitly'

or 'expressly.' The Chicago-based appellate court has been considering Blagojevich's appeal for more than a year month. It's unclear when it might issue a decision.

High-quality CoT:

1. The important entities in this document are Rod Blagojevich, the Federal Correctional Institution Englewood, and the National Enquirer.

2. The important dates in this document are 2008, 2009, and 2014.

3. The events in this document are Blagojevich's arrest and imprisonment, and his appeal of his conviction.

4. The result of these events is that Blagojevich is currently serving a 14-year sentence in prison and his appeal is still pending.

Correct Summary:

In 2008, Rod Blagojevich was arrested for trying to sell Barack Obama's U.S. Senate seat. He was impeached and removed from office in 2009, and sentenced to 14 years in prison. He is currently serving his sentence at the Federal Correctional Institution Englewood near Denver. His lawyers attempted to appeal his conviction, arguing that a 2014 Supreme Court ruling supported their view that Blagojevich was engaged in legal, run-of-the-mill political horse trading, not corruption. However, government attorneys responded that the Supreme Court ruling did not support Blagojevich's appeal, and his appeal is still pending. [10 valid elements; coherent; no redundant information.]

Low-quality CoT:

1. The important entities in this document are the photos, Rod Blagojevich and the Federal Correctional Institution Englewood. [Basically correct but omits key entities. The *photo* serves as the news hook, with Blagojevich and the prison being the core elements. However, it fails to mention the National Enquirer (the publisher of the photo) and the McCutcheon case (crucial for the legal appeal), resulting in slightly insufficient completeness.]

2. The events in this document are Blagojevich's arrest and imprisonment. [Incomplete. It only covers the basic events and completely leaves out the *appeal process* and *legal arguments* (which are the focus of the latter part of the news), making the reasoning overly simplistic.]

3. The result of these events is that Blagojevich is currently serving a 14-year sentence in prison. [The facts are correct but not comprehensive enough. While it accurately states his current status ("currently serving a 14-year sentence"), it omits the key outcome that "the appeal is still pending," failing to fully summarize the current situation of the events.]

[Note: Overall, many important dates and events are missing in this CoT.]

Incorrect Summary:

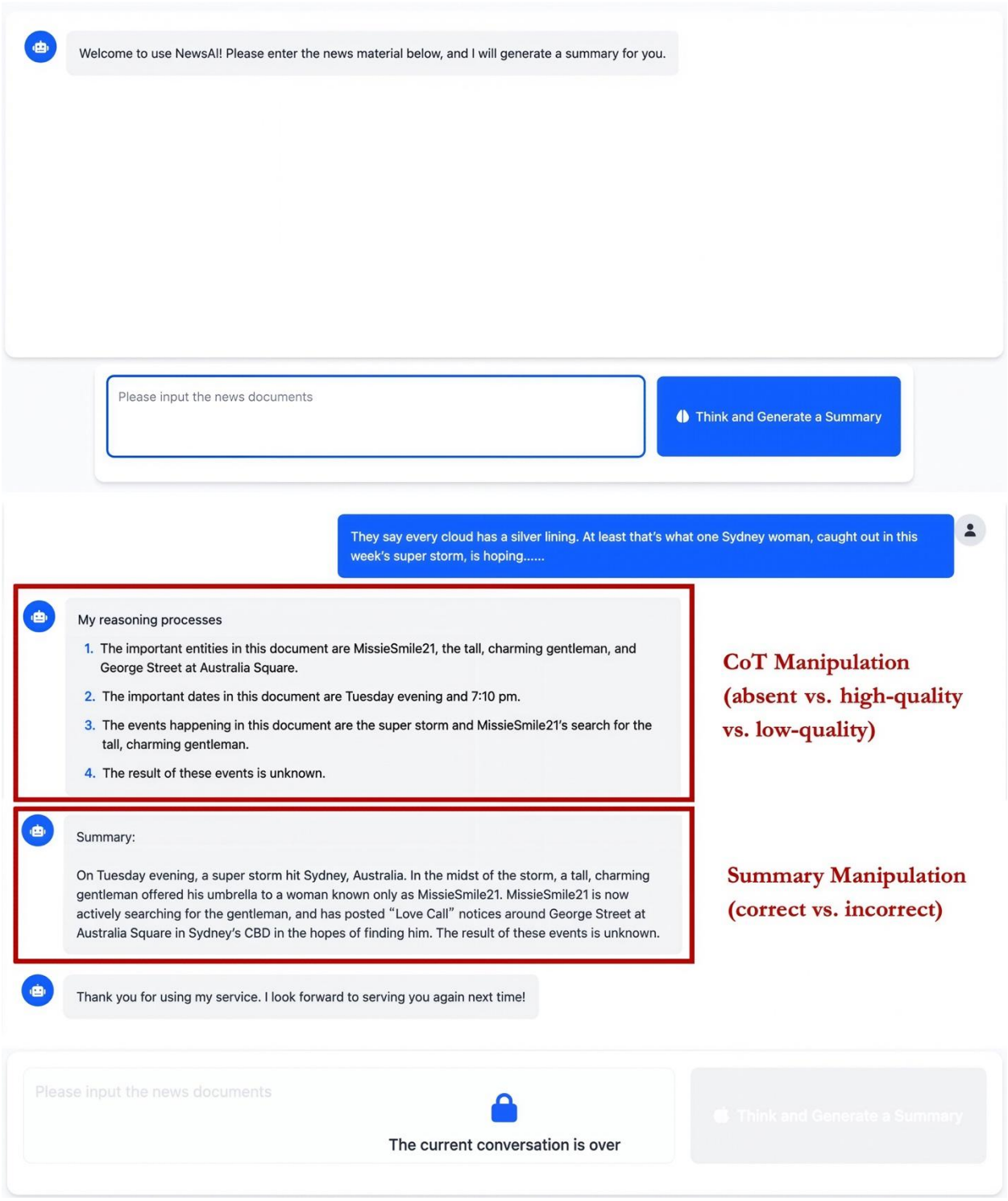
The article discusses how Rod Blagojevich, a former Illinois governor, has let his hair go while serving his prison sentence. The photos show the former governor with white hair, rather than the black hair that was his trademark as a politician. Blagojevich is currently serving a 14-year sentence at the Federal Correctional Institution Englewood near Denver. [4 valid elements (yellow colored); coherent; redundant information: color change of the hair.]

Table B.4. ANOVA Analysis Results on Step-level CoT Score Across Task Difficulty Levels

		Low-difficulty Task	High-difficulty Task
Perceived Task	Overall	3.15	**5.02*****
Difficulty		(1.44)	(1.03)
	Component complexity	3.06	**4.82*****
		(1.60)	(1.20)
	Coordinative complexity	3.14	**4.88*****
		(1.43)	(1.11)
	Mental effort	3.24	**5.37*****
		(1.61)	(1.13)
N		100	100
Low-quality CoT	Overall	3.28	4.01
		(1.06)	(0.68)
	Step 1	4.04	3.82
		(1.09)	(0.96)
	Step 2	2.40	Null
		(1.58)	
	Step 3	3.66	4.18
		(1.41)	(0.85)
	Step 4	3.00	4.02
		(1.64)	(1.15)
N		50	50
High-quality CoT	Overall	**4.44*****	**4.54*****
		(0.44)	(0.32)
	Step 1	**4.42***	**4.24***
		(0.76)	(0.69)
	Step 2	**4.48*****	4.64
		(0.71)	(0.60)
	Step 3	**4.26****	**4.62****
		(0.80)	(0.53)
	Step 4	**4.60*****	**4.66*****
		(0.61)	(0.48)
N		50	50

Notes. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Standard errors are in parentheses.
- Perceived task difficulty was measured from 1 (lower) to 7 (higher). Component complexity: writing a summary for this news material requires processing a large amount of information cues across a wider range of entity types. Coordinative complexity: writing a summary for this news material requires integrating conflicting information and switching between multiple timelines, entities, or events. Mental effort: writing a summary for this news material requires a great deal of effort.
- Steps 1, 2, 3, and 4 correspond to entities, dates, events, and results, respectively. In the high-difficulty task condition, the low-quality CoT omitted Step 2 for dates. The score ranges from 1 (min) to 5 (max).

Figure B.1. Human-AI Chat Interface



Notes. An example of presenting the high-quality CoT and correct summary under the low-difficulty task.

Appendix C. Measurements

Table C.1. Descriptions of Measurement Scales

Dimension	Measures
Pre-Experiment Measures Before Interacting with the GenAI Assistant	
Gender	1. Female; 2. Male
Age	Input
Education	1. Associate degree or below 2. Bachelor degree 3. Master degree or above
GenAI Use Frequency	1. Once a week 2. Several times a week 3. Once a day 4. Several times a day
GenAI Literacy (Laupichler et al. 2023)	(1 (strongly disagree) ---------- 7 (strongly agree)) *Technical understanding*… 1. I can understand how GenAI works. 2. I can understand how GenAI generates responses. *Critical appraisals*… 3. I can evaluate the accuracy of GenAI responses. 4. I can identify biases and errors in GenAI responses.
Initial Trust in GenAI (McKnight et al. 2002)	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. GenAI systems are competent and effective. 2. GenAI systems perform their role very well. 3. Overall, GenAI systems are capable and proficient. 4. In general, GenAI systems have rich and powerful domain knowledge.
Need for Cognition (Cacioppo et al. 1984)	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. I would prefer complex to simple problems.

	2. I like to have the responsibility of handling a situation that requires a lot of thinking. 3. I really enjoy a task that involves coming up with new solutions to problems. 4. I usually end up deliberating about issues even when they do not affect me personally.
Perceived Self-efficacy/Task Confidence (Chen et al., 2001)	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. I know how to generate news summaries. 2. I have mastered the skills necessary for generating news summaries. 3. I am confident about my ability to generate news summaries.
Perceived Task Difficulty (Sweller 1994, 2010)	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. The news content is extremely complex for me. 2. I need to expend a great deal of effort to understand the news content. 3. The concepts involved in the news are very difficult for me to comprehend.
Manipulation Checks	
CoT Presence	1. The GenAI assistant provides reasoning steps in the summary-generating process. (Yes/No)
Open Question	1. How did you understand or make use of the reasoning steps (if any) and summaries generated by the GenAI assistant to reach your final decision?
Perceived CoT Quality (only CoT presence experimental conditions) (Wu et al. 2022)	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. The GenAI assistant provides sufficient details in its reasoning steps to justify its final summary. (Completeness) 2. The GenAI assistant ensures logical consistency among reasoning steps to achieve its final summary. (Logical coherence) 3. I'm satisfied with the GenAI assistant's reasoning steps; they align with its final summary. (CoT-output alignment)
CoT-related Single Choice Questions (Study 4)	*Please choose the most accurate option from the four choices provided.* 1. According to the reasoning steps of the GenAI assistant, what time was identified for the event? 2. In the reasoning steps of the GenAI assistant, which of the following groups are explicitly listed as important entities? 3. In the reasoning steps, what is the GenAI assistant's judgment regarding the outcome of the event?
Post-Experiment Measures After Interacting with the GenAI Assistant	
Perceived AI Transparency (Self-developed based on Shin et al. (2020))	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. The GenAI assistant is transparent about how it arrives at its final summary; I could roughly trace its internal states/reasoning processes. 2. I think the GenAI assistant's final summary is explainable by figuring out its internal mechanics/reasoning processes. 3. I can understand how the GenAI assistant arrives at its final summary even if I don't fully grasp the technical details.

Perceived Diagnosticity (Pham & Avnet 2004)	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. The GenAI assistant helps me effectively evaluate its performance/generated summary. 2. The GenAI assistant provides me a chance to assess its performance/generated summary in a systematic and effective way. 3. Through the GenAI assistant, I can quickly obtain a good understanding of its performance/generated summary.
Germane Cognitive Load (Sweller 1994, 2010)	(1 (strongly disagree) ---------- 7 (strongly agree)) 1. The GenAI assistant helps me better understand the news content. 2. The GenAI assistant makes it easier for me to generate the summary. 3. When using the GenAI assistant, I can connect new knowledge with my existing knowledge. *Note. In CoT-present conditions, we change the "GenAI assistant" into "GenAI's reasoning steps".*
Perceived Trust in GenAI's Competence (McKnight et al. 2002)	*After interacting with the GenAI assistant in this experiment, I think* … (1 (strongly disagree) ---------- 7 (strongly agree)) 1. The GenAI assistant is competent and effective in writing news summary. 2. The GenAI assistant performs its role of writing news summary very well. 3. Overall, the GenAI assistant is a capable and proficient news summarizer. 4. In general, the GenAI assistant is very knowledgeable about news summarization.

Appendix D. Supplementary Results

Table D.1. Descriptive Statistics of Study 1

Correct Summary	Low-difficulty Task			High-difficulty Task		
	No CoT	Low-quality CoT	High-quality CoT	No CoT	Low-quality CoT	High-quality CoT
Main Variables						
Perceived trust in GenAI (Cronbach's α= 0.893)	5.21 (1.00)	5.10 (1.19)	5.79 (0.74)	5.34 (0.75)	5.25 (1.16)	5.73 (0.64)
Perceived CoT quality (Cronbach's α= 0.922)		4.64 (1.38)	5.79 (0.76)		4.99 (1.24)	5.52 (0.90)
Perceived transparency (Cronbach's α= 0.859)	3.94 (1.37)	5.05 (0.98)	5.76 (0.71)	4.25 (1.36)	5.28 (0.96)	5.39 (0.85)
Control Variables						
Genders	48% male	54% male	53% male	58% male	60% male	40% male
Ages	23.77 (4.29)	23.52 (2.91)	23.42 (2.71)	23.90 (4.54)	22.75 (1.87)	23.11 (2.13)
Education	2.10 (0.36)	2.07 (0.33)	2.07 (0.50)	2.04 (0.29)	2.13 (0.39)	2.07 (0.69)
GenAI use frequency	2.79 (0.98)	3.00 (0.97)	3.02 (0.99)	3.13 (0.98)	3.06 (1.05)	2.98 (0.89)
Initial trust in GenAI (Cronbach's α= 0.823)	5.53 (0.78)	5.74 (0.75)	5.89 (0.65)	5.81 (0.65)	5.49 (0.89)	5.65 (0.72)
Need for cognition (Cronbach's α= 0.869)	4.94 (1.27)	4.93 (1.21)	5.15 (1.15)	4.58 (1.52)	4.17 (1.34)	4.28 (1.30)
Perceived self-efficacy (Cronbach's α= 0.848)	5.20 (0.95)	5.13 (0.82)	5.51 (0.81)	5.07 (1.15)	5.13 (0.99)	4.90 (0.88)
Perceived task difficulty (Cronbach's α= 0.934)	3.31 (1.45)	2.80 (0.94)	2.87 (1.00)	4.40 (1.54)	5.01 (0.88)	5.07 (0.90)
Incorrect Summary	No CoT	Low-quality CoT	High-quality CoT	No CoT	Low-quality CoT	High-quality CoT
Main Variables						
Perceived trust in GenAI (Cronbach's α= 0.953)	4.54 (1.41)	4.27 (0.98)	4.93 (1.34)	4.03 (1.63)	5.07 (1.43)	5.02 (1.56)
Perceived CoT quality (Cronbach's α= 0.938)		3.88 (1.58)	5.31 (1.00)		5.25 (1.20)	5.44 (1.03)
Perceived transparency (Cronbach's α= 0.856)	3.87 (1.39)	5.20 (0.85)	5.27 (0.90)	4.01 (1.50)	5.37 (1.01)	5.46 (0.98)
Control Variables						
Genders	47% male	42% male	47% male	55% male	59% male	47% male
Ages	22.80 (3.17)	23.96 (3.55)	23.69 (3.79)	24.59 (4.16)	24.00 (3.75)	23.25 (3.47)
Education	2.08 (0.34)	2.13 (0.44)	2.04 (0.45)	2.10 (0.42)	2.10 (0.41)	2.04 (0.53)
GenAI use frequency	2.82 (0.97)	3.00 (0.99)	2.96 (1.00)	2.88 (0.99)	2.84 (0.95)	2.75 (0.96)
Initial trust in GenAI (Cronbach's α= 0.802)	5.85 (0.70)	5.76 (0.69)	5.71 (0.69)	5.81 (0.66)	5.83 (0.58)	5.86 (0.75)
Need for cognition (Cronbach's α= 0.867)	4.89 (1.30)	4.74 (1.29)	4.87 (1.2)	4.63 (1.41)	4.89 (1.26)	4.72 (1.53)
Perceived self-efficacy (Cronbach's α= 0.874)	5.24 (0.92)	5.33 (0.96)	5.10 (1.03)	5.33 (1.07)	5.23 (0.97)	5.36 (1.21)
Perceived task difficulty (Cronbach's α= 0.947)	3.59 (1.39)	2.88 (1.31)	3.38 (1.69)	4.54 (1.53)	4.50 (1.52)	4.26 (1.65)

Notes. N = 598. GenAI literacy items were dropped in Study 1. Standard errors are in parentheses.

Table D.2. Correlation Matrix of Study 1

Correct Summary	1	2	3	4	5	6	7	8	9	10	11
1. CoT											
2. Perceived trust in GenAI	0.065										
3. Perceived transparency	0.381	0.470									
4. Genders	0.037	0.036	0.064								
5. Ages	0.091	0.048	0.096	0.011							
6. Education	0.029	0.132	0.126	0.009	0.162						
7. GenAI use frequency	0.034	0.005	0.064	0.006	0.014	0.063					
8. Initial trust in GenAI	0.034	0.480	0.278	0.048	0.243	0.045	0.183				
9. Need for cognition	0.081	0.109	0.111	0.150	0.033	0.101	0.048	0.238			
10. Perceived self-efficacy	0.003	0.225	0.230	0.077	0.205	0.070	0.120	0.423	0.305		
11. Perceived task difficulty	0.041	0.105	0.101	0.083	0.106	0.030	0.064	0.049	0.335	0.184	
Incorrect Summary	1	2	3	4	5	6	7	8	9	10	11
1. CoT											
2. Perceived trust in GenAI	0.114										
3. Perceived transparency	0.453	0.640									
4. Genders	0.004	0.126	0.079								
5. Ages	0.032	0.193	0.181	0.066							
6. Education	0.018	0.054	0.030	0.138	0.179						
7. GenAI use frequency	0.030	0.044	0.024	0.071	0.004	0.007					
8. Initial trust in GenAI	0.038	0.245	0.148	0.203	0.062	0.093	0.180				
9. Need for cognition	0.055	0.067	0.069	0.053	0.044	0.074	0.152	0.118			
10. Perceived self-efficacy	0.032	0.043	0.082	0.093	0.140	0.036	0.025	0.370	0.295		
11. Perceived task difficulty	0.090	0.146	0.064	0.015	0.067	0.016	0.070	0.058	0.139	0.165	

Table D.3. Regression Results of Study 1

DV: Perceived trust in GenAI (ANCOVA)	Low-difficulty Task					High-difficulty Task				
	Type III sum of squares	Degrees of freedom	Mean square	*F* value	*p*-value	Type III sum of squares	Degrees of freedom	Mean square	*F* value	*p*-value
Correct Summary										
No vs. *Low*	0.38	1	0.38	0.35	0.557	0.003	1	0.003	0.004	0.949
No vs. *High*	4.43	1	4.43	6.99	0.01	3.48	1	3.48	8.73	0.004
Low vs. *High*	8.74	1	8.74	10.99	0.001	3.70	1	3.70	5.54	0.021
Incorrect Summary										
No vs. *Low*	0.28	1	0.28	0.23	0.631	28.85	1	28.85	12.06	0.001
No vs. *High*	3.68	1	3.68	2.52	0.116	33.75	1	33.75	14.90	0.000
Low vs. *High*	8.31	1	8.31	7.07	0.009	0.20	1	0.20	0.10	0.754

Notes. Control variables were included in the tests.

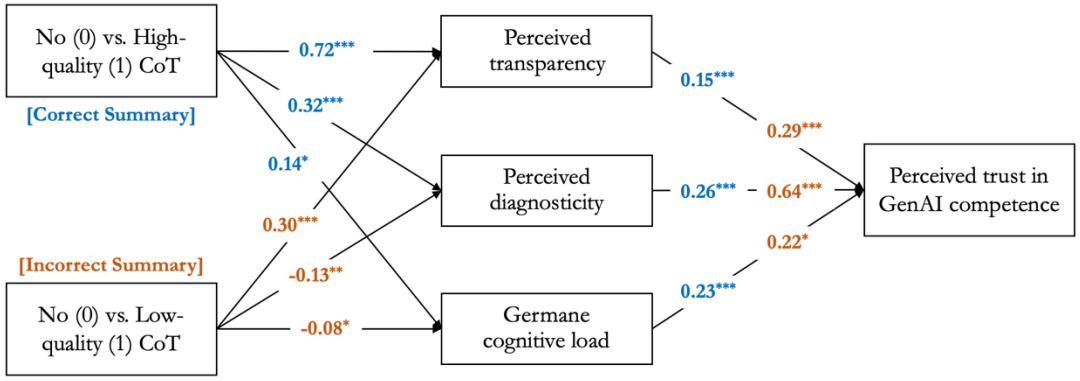
Table D.4. Descriptive Statistics of Study 2

Low-difficulty Task	Correct Summary		Incorrect Summary	
	No CoT	High-quality CoT	No CoT	Low-quality CoT
Main Variables				
Perceived trust in GenAI (Cronbach's α= 0.954)	5.37 (0.94)	5.59 (0.83)	4.56 (1.66)	4.03 (1.62)
Perceived CoT quality (Cronbach's α= 0.945)		5.62 (0.96)		4.36 (1.43)
Perceived transparency (Cronbach's α= 0.881)	4.16 (1.37)	5.60 (0.78)	3.92 (1.55)	4.99 (1.16)
Perceived diagnosticity (Cronbach's α= 0.873)	5.00 (1.27)	5.64 (0.91)	5.03 (1.21)	4.37 (1.33)
Germane cognitive load (Cronbach's α= 0.869)	5.33 (1.01)	5.62 (0.90)	5.45 (1.02)	5.05 (1.15)
Control Variables				
Genders	47% male	43% male	46% male	44% male
Ages	24.04 (4.67)	24.40 (4.69)	24.16 (4.28)	23.41 (3.48)
Education	2.12 (0.40)	2.19 (0.52)	2.15 (0.48)	2.12 (0.40)
GenAI use frequency	2.88 (1.00)	2.95 (1.00)	2.83 (0.99)	3.02 (0.98)
GenAI literacy1 (Cronbach's α= 0.787)	5.41 (0.87)	5.30 (0.91)	5.37 (0.94)	5.32 (0.90)
GenAI literacy2 (Cronbach's α= 0.718)	5.45 (0.94)	5.46 (0.95)	5.33 (0.95)	5.45 (0.92)
Initial trust in GenAI (Cronbach's α= 0.814)	5.74 (0.74)	5.76 (0.72)	5.83 (0.76)	5.62 (0.74)
Need for cognition (Cronbach's α= 0.889)	4.82 (1.40)	4.69 (1.26)	4.79 (1.23)	4.74 (1.29)
Perceived self-efficacy (Cronbach's α= 0.856)	5.12 (1.03)	5.20 (0.92)	5.30 (0.99)	5.25 (0.94)
Perceived task difficulty (Cronbach's α= 0.935)	3.87 (1.47)	3.81 (1.58)	4.03 (1.46)	3.84 (1.47)

Notes. *N* = 464. Standard errors are in parentheses.

Table D.5. Correlation Matrix of Study 2

Low-difficulty Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CoT															
2. Perceived trust in GenAI	0.303														
3. Perceived transparency	0.270	0.497													
4. Perceived diagnosticity	0.218	0.779	0.461												
5. Germane cognitive load	0.135	0.629	0.367	0.799											
6. Genders	0.020	0.110	0.068	0.071	0.036										
7. Ages	0.062	0.142	0.156	0.173	0.136	0.077									
8. Education	0.011	0.058	0.027	0.030	0.028	0.137	0.113								
9. GenAI use frequency	0.062	0.023	0.070	0.025	0.018	0.005	0.019	0.026							
10. GenAI literacy1	0.034	0.166	0.151	0.190	0.256	0.097	0.121	0.081	0.212						
11. GenAI literacy2	0.019	0.091	0.063	0.119	0.204	0.055	0.184	0.043	0.183	0.521					
12. Initial trust in GenAI	0.087	0.348	0.222	0.356	0.347	0.132	0.169	0.093	0.197	0.544	0.382				
13. Need for cognition	0.022	0.062	0.025	0.090	0.110	0.145	0.037	0.040	0.024	0.249	0.318	0.155			
14. Perceived self-efficacy	0.038	0.072	0.091	0.202	0.265	0.084	0.200	0.038	0.112	0.542	0.589	0.444	0.485		
15. Perceived task difficulty	0.024	0.094	0.089	0.105	0.086	0.093	0.101	0.107	0.073	0.108	0.189	0.031	0.161	0.213	



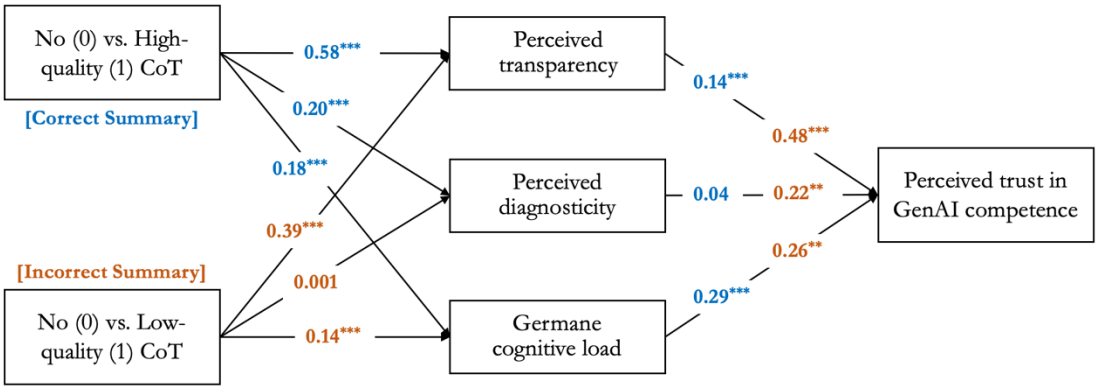
Notes. Control variables were included in the tests. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure D.1. Mediating Paths of Study 2 in Low-Difficulty Task

Table D.6. Descriptive Statistics of Study 3

High-difficulty Task	Correct Summary		Incorrect Summary	
	No CoT	High-quality CoT	No CoT	Low-quality CoT
Main Variables				
Perceived trust in GenAI (Cronbach's α= 0.954)	5.55 (0.88)	5.85 (0.70)	3.89 (1.62)	4.66 (1.50)
Perceived CoT quality (Cronbach's α= 0.917)		5.78 (0.80)		4.86 (1.27)
Perceived transparency (Cronbach's α= 0.893)	4.40 (1.34)	5.66 (0.78)	3.60 (1.39)	5.11 (1.04)
Perceived diagnosticity (Cronbach's α= 0.900)	5.26 (1.03)	5.72 (0.90)	4.76 (1.48)	4.74 (1.30)
Germane cognitive load (Cronbach's α= 0.838)	5.47 (0.70)	5.86 (0.70)	4.76 (1.27)	5.31 (1.09)
Control Variables				
Genders	51% male	47% male	47% male	43% male
Ages	24.69 (4.18)	25.02 (4.12)	25.68 (4.41)	25.14 (3.89)
Education	2.15 (0.51)	2.16 (0.53)	2.17 (0.53)	2.10 (0.47)
GenAI use frequency	3.14 (0.98)	3.06 (0.93)	3.11 (0.99)	2.99 (0.99)
GenAI literacy1 (Cronbach's α= 0.776)	5.22 (0.97)	5.45 (0.84)	5.27 (0.98)	5.24 (0.90)
GenAI literacy2 (Cronbach's α= 0.726)	5.38 (0.98)	5.48 (0.94)	5.28 (0.82)	5.33 (0.94)
Initial trust in GenAI (Cronbach's α= 0.783)	5.86 (0.70)	5.85 (0.68)	5.69 (0.74)	5.67 (0.68)
Need for cognition (Cronbach's α= 0.889)	4.74 (1.40)	4.78 (1.24)	4.87 (1.18)	4.60 (1.31)
Perceived self-efficacy (Cronbach's α= 0.841)	5.15 (1.03)	5.38 (0.86)	5.31 (0.88)	5.08 (1.04)
Perceived task difficulty (Cronbach's α= 0.934)	4.41 (1.48)	4.30 (1.66)	4.25 (1.45)	4.29 (1.48)

Notes. N = 559. Standard errors are in parentheses.



Notes. Control variables were included in the tests. $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure D.2. Mediating Paths of Study 3 in High-Difficulty Task

Table D.7. Correlation Matrix of Study 3

High-difficulty Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CoT															
2. Perceived trust in GenAI	0.092														
3. Perceived transparency	0.297	0.607													
4. Perceived diagnosticity	0.117	0.583	0.506												
5. Germane cognitive load	0.056	0.674	0.572	0.795											
6. Genders	0.054	0.069	0.078	0.041	0.033										
7. Ages	0.012	0.076	0.033	0.027	0.036	0.086									
8. Education	0.043	0.072	0.066	0.025	0.036	0.064	0.078								
9. GenAI use frequency	0.057	0.017	0.013	0.036	0.031	0.037	0.125	0.089							
10. GenAI literacy1	0.033	0.094	0.240	0.154	0.125	0.068	0.097	0.103	0.169						
11. GenAI literacy2	0.007	0.123	0.145	0.170	0.200	0.126	0.076	0.062	0.184	0.603					
12. Initial trust in GenAI	0.085	0.306	0.149	0.227	0.259	0.107	0.150	0.040	0.210	0.475	0.328				
13. Need for cognition	0.060	0.070	0.041	0.068	0.043	0.139	0.047	0.050	0.031	0.295	0.325	0.174			
14. Perceived self-efficacy	0.051	0.118	0.165	0.170	0.201	0.121	0.169	0.075	0.138	0.553	0.504	0.434	0.371		
15. Perceived task difficulty	0.017	0.139	0.063	0.138	0.132	0.059	0.094	0.052	0.023	0.063	0.106	0.077	0.068	0.150	

Table D.8. Descriptive Statistics of Study 4

Correct Summary	Low-difficulty Task		High-difficulty Task	
	No CoT	High-quality CoT	No CoT	High-quality CoT
Main Variables				
Perceived trust in GenAI (Cronbach's α= 0.905)	4.85 (1.49)	5.78 (0.55)	5.24 (1.05)	5.78 (0.64)
Perceived CoT quality (Cronbach's α= 0.949)		5.77 (0.62)		5.86 (0.57)
Perceived transparency (Cronbach's α= 0.900)	3.56 (1.47)	5.78 (0.65)	4.09 (1.41)	5.84 (0.62)
Control Variables				
Genders	60% male	48% male	43% male	47% male
Ages	27.71 (4.05)	26.57 (3.32)	27.43 (4.13)	27.25 (3.48)
Education	2.13 (0.42)	2.06 (0.54)	2.20 (0.48)	2.14 (0.53)
GenAI use frequency	3.32 (0.92)	3.17 (0.91)	3.38 (0.90)	3.20 (0.96)
GenAI literacy1 (Cronbach's α= 0.729)	5.41 (0.93)	5.34 (0.86)	5.27 (0.87)	5.56 (0.92)
GenAI literacy2 (Cronbach's α= 0.822)	5.40 (0.96)	5.29 (0.95)	5.34 (0.97)	5.52 (0.83)
Initial trust in GenAI (Cronbach's α= 0.724)	5.89 (0.70)	5.75 (0.59)	5.90 (0.70)	5.92 (0.55)
Need for cognition (Cronbach's α= 0.888)	5.18 (1.09)	4.94 (1.19)	4.93 (1.15)	5.21 (1.18)
Perceived self-efficacy (Cronbach's α= 0.794)	5.24 (0.97)	5.20 (1.08)	5.22 (0.87)	5.42 (0.94)
Perceived task difficulty (Cronbach's α= 0.943)	2.45 (1.02)	2.93 (1.22)	4.32 (1.48)	4.31 (1.52)
Incorrect Summary	No CoT	Low-quality CoT	No CoT	Low-quality CoT
Main Variables				
Perceived trust in GenAI (Cronbach's α= 0.970)	4.60 (1.72)	3.19 (1.61)	3.08 (1.61)	4.12 (1.64)
Perceived CoT quality (Cronbach's α= 0.940)		4.11 (1.56)		4.42 (1.60)
Perceived transparency (Cronbach's α= 0.923)	4.05 (1.74)	4.55 (1.55)	3.16 (1.45)	4.74 (1.25)
Control Variables				
Genders	45% male	43% male	46% male	44% male
Ages	27.05 (4.17)	27.34 (3.79)	27.44 (3.92)	26.34 (4.27)
Education	2.16 (0.54)	2.18 (0.53)	2.19 (0.47)	2.10 (0.43)
GenAI use frequency	3.34 (0.91)	3.39 (0.86)	3.17 (0.98)	3.26 (0.90)
GenAI literacy2 (Cronbach's α= 0.795)	5.51 (0.86)	5.36 (0.72)	5.39 (0.89)	5.54 (0.88)
Initial trust in GenAI (Cronbach's α= 734)	5.61 (0.84)	5.75 (0.56)	5.87 (0.68)	5.80 (0.63)
Need for cognition (Cronbach's α= 0.910)	5.15 (1.25)	5.09 (1.16)	5.08 (1.35)	4.99 (1.29)
Perceived self-efficacy (Cronbach's α= 0.825)	5.26 (0.83)	5.16 (0.90)	5.25 (0.82)	5.27 (0.94)
Perceived task difficulty (Cronbach's α= 0.950)	2.83 (1.35)	2.44 (0.94)	4.34 (1.46)	4.31 (1.35)

Notes. N = 500. GenAI literacy1 items dropped in the incorrect-summary condition of Study 4. Standard errors are in parentheses.

Table D.9. Correlation Matrix of Study 4

Correct Summary	1	2	3	4	5	6	7	8	9	10	11	12	13
1. CoT													
2. Perceived trust in GenAI	0.168												
3. Perceived transparency	0.337	0.500											
4. Genders	0.040	0.035	0.079										
5. Ages	0.087	0.050	0.034	0.064									
6. Education	0.061	0.052	0.067	0.142	0.167								
7. GenAI use frequency	0.087	0.043	0.035	0.183	0.043	0.205							
8. GenAI literacy1	0.073	0.055	0.054	0.101	0.059	0.148	0.255						
9. GenAI literacy2	0.059	0.070	0.037	0.219	0.100	0.112	0.116	0.534					
10. Initial trust in GenAI	0.080	0.123	0.057	0.084	0.046	0.076	0.101	0.484	0.276				
11. Need for cognition	0.041	0.040	0.066	0.202	0.059	0.126	0.119	0.445	0.455	0.163			
12. Perceived self-efficacy	0.043	0.073	0.092	0.185	0.016	0.199	0.238	0.426	0.522	0.286	0.602		
13. Perceived task difficulty	0.084	0.073	0.197	0.023	0.023	0.009	0.023	0.074	0.098	0.104	0.274	0.141	
Incorrect Summary	1	2	3	4	5	6	7	8	9	10	11	12	
1. CoT													
2. Perceived trust in GenAI	0.082												
3. Perceived transparency	0.106	0.663											
4. Genders	0.026	0.016	0.028										
5. Ages	0.050	0.020	0.060	0.146									
6. Education	0.035	0.059	0.017	0.119	0.163								
7. GenAI use frequency	0.036	0.021	0.031	0.057	0.027	0.121							
8. GenAI literacy2	0.023	0.056	0.027	0.167	0.038	0.051	0.133						
9. Initial trust in GenAI	0.034	0.102	0.196	0.087	0.099	0.178	0.029	0.240					
10. Need for cognition	0.043	0.047	0.025	0.194	0.031	0.008	0.058	0.550	0.146				
11. Perceived self-efficacy	0.038	0.153	0.040	0.115	0.033	0.096	0.080	0.561	0.320	0.578			
12. Perceived task difficulty	0.066	0.070	0.076	0.055	0.085	0.027	0.141	0.183	0.044	0.232	0.183		

Table D.10. Regression Results of Study 4

	Low-difficulty Task					High-difficulty Task				
Correct Summary	Type III sum of squares	Degrees of freedom	Mean square	*F* value	*p*-value	Type III sum of squares	Degrees of freedom	Mean square	*F* value	*p*-value
DV: Perceived trust in GenAI (ANCOVA)										
No vs. *High*	22.36	1	22.36	19.12	0.000	7.06	1	7.06	10.69	0.001
Incorrect Summary	Type III sum of squares	Degrees of freedom	Mean square	*F* value	*p*-value	Type III sum of squares	Degrees of freedom	Mean square	*F* value	*p*-value
DV: Perceived trust in GenAI (ANCOVA)										
No vs. *Low*	40.84	1	40.84	17.40	0.000	26.07	1	26.07	11.54	0.001

Notes. Control variables were included in the tests.

References

Aoki, N., Tatsumi, T., Naruse, G., & Maeda, K. (2024). Explainable AI for government: Does the type of explanation matter to the accuracy, fairness, and trustworthiness of an algorithmic decision as perceived by those who are affected?. *Government Information Quarterly*, *41*(4), 101965.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191.

Bauer, K., Von Zahn, M., & Hinz, O. (2023). Expl (AI) ned: The impact of explainable artificial intelligence on users' information processing. *Information Systems Research*, *34*(4), 1582-1602.

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307.

Chen, C., & Zhao, X. (2026). Let's think step by step: Effects of chain-of-thought prompt coaching on users' perceptions and trust in image generative AI tools. *International Journal of Human–Computer Interaction*, *42*(5), 2943-2964.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, *4*(1), 62-83.

Chen, H., Wang, P., & Hao, S. (2025). AI in the spotlight: The impact of artificial intelligence disclosure on user engagement in short-form videos. *Computers in Human Behavior*, *162*, 108448.

Chen, Q., Qin, L., Wang, J., Zhou, J., & Che, W. (2024). Unlocking the capabilities of thought: A reasoning boundary framework to quantify and optimize chain-of-thought. *Advances in Neural Information Processing Systems*, *37*, 54872-54904.

Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., ... & Wei, J. (2024). Scaling instruction-finetuned language models. *Journal of Machine Learning Research*, *25*(70), 1-53.

Danry, V., Pataranutaporn, P., Groh, M., & Epstein, Z. (2025, April). Deceptive explanations by large language models lead people to change their beliefs about misinformation more often than honest explanations. In *Proceedings of the 2025*

CHI Conference on Human Factors in Computing Systems (pp. 1-31).

De Freitas, J., Nave, G., & Puntoni, S. (2025). Ideation with generative AI—in consumer research and beyond. *Journal of Consumer Research*, *52*(1), 18-31.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, *27*(1), 51-90.

Goli, A., & Singh, A. (2024). Frontiers: Can large language models capture human preferences?. *Marketing Science*, *43*(4), 709-722.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1-38.

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*, *32*(3), 713-735.

Kim, S. S., Vaughan, J. W., Liao, Q. V., Lombrozo, T., & Russakovsky, O. (2025, April). Fostering appropriate reliance on large language models: The role of explanations, sources, and inconsistencies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy"–An exploratory factor analysis. *Computers in Human Behavior Reports*, *12*, 100338.

Leng, Y., & Yuan, Y. (2026). SUVA: A probabilistic framework for auditing LLMs with an application to social preferences. *Information Systems Research*. https://doi.org/10.1287/isre.2024.0857

Li, Z., Zhu, H., Lu, Z., Xiao, Z., & Yin, M. (2025, April). From text to trust: Empowering AI-assisted decision making with adaptive LLM-powered analysis. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).

Ma, S., Lei, Y., Wang, X., Zheng, C., Shi, C., Yin, M., & Ma, X. (2023, April). Who should I trust: AI or myself? Leveraging human and AI correctness likelihood to promote appropriate trust in AI-assisted decision-making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334-359.

Molina, M. D., & Sundar, S. S. (2022). When AI moderates online content: effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication*, *27*(4), zmac010.

Pham, M. T., & Avnet, T. (2004). Ideals and oughts and the reliance on affect versus substance in persuasion. *Journal of Consumer Research*, *30*(4), 503-518.

Rheu, M., Dai, Y., Meng, J., & Peng, W. (2024). When a chatbot disappoints you: Expectancy violation in human-chatbot interaction in a social support context. *Communication Research*, *51*(7), 782-814.

Saffarizadeh, K., & Keil, M. (2026). Conversational AI agents: The effect of process and outcome variation on anthropomorphism and trust. *Information Systems Journal*, *36*(4), 597-623.

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences?. *International Journal of Information Management*, *52*, 102061.

Shi, S., Gong, Y., & Gursoy, D. (2021). Antecedents of trust and adoption intention toward artificially intelligent recommendation systems in travel planning: A heuristic–systematic model. *Journal of Travel Research*, *60*(8), 1714-1734.

Steyvers, M., & Kumar, A. (2024). Three challenges for AI-assisted decision-making. *Perspectives on Psychological Science*, *19*(5), 722-734.

Steyvers, M., Tejeda, H., Kumar, A., Belem, C., Karny, S., Hu, X., Mayer, L. W., & Smyth, P. (2025). What large language models know and what people think they know. *Nature Machine Intelligence*, *7*(2), 221-231.

Sun, Y., Liao, M. M., Sundar, S. S., & Walther, J. B. (2026). Does transparency matter when an AI system meets performance expectations? An experiment with an online dating site. *Computers in Human Behavior*, 108875.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295-312.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123-138.

Turpin, M., Michael, J., Perez, E., & Bowman, S. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems*, *36*, 74952-74965.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NIPS'17)* (Curran Associates Inc., Red Hook, NY), 6000–6010.

Von Zahn, M., Liebich, L., Jussupow, E., Hinz, O., & Bauer, K. (2025). Knowing (not) to know: Explainable artificial intelligence and human metacognition. *Information Systems Research*. https://doi.org/10.1287/isre.2024.1431

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824-24837.

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B., & Rafaeli, A. (2023). Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. *Computers in Human Behavior*, *144*, 107714.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*(1), 60-82.

Wu, T., Terry, M., & Cai, C. J. (2022, April). AI chains: Transparent and controllable human-AI interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-22).

Yin, Z., & Wang, S. (2025). Enhancing scientific table understanding with type-guided chain-of-thought. *Information Processing & Management*, *62*(4), 104159.

Zhang, Z., Lin, M., & Liu, X. (2026). Enhancing employees' change readiness for artificial intelligence (AI) adoption: The role of AI transparency. *International Journal of Hospitality Management*, *134*, 104620.