

Opening the black box: How reasoning-enabled AI agents influence user perceptions and behavior in sustainable consumption

Pejman Ebrahimi^{a,*}, Stefan Hoffmann^b, Johannes Schneider^a

^a Department of Information Systems & Computer Science, University of Liechtenstein, Liechtenstein

^b Department of Marketing & Kiel Institute for Responsible Innovation, Kiel University, Germany

ARTICLE INFO

Dataset link:

[COT-Sustainable-Consumption-Behavior](#)

Keywords:

Chain-of-Thought Reasoning
Conversational Agents
Explainable AI
Sustainable Consumption
Technology Acceptance

ABSTRACT

Recent advances in chain-of-thought (CoT) reasoning in large language models are expected to improve the transparency and coherence of AI-generated recommendations, yet evidence on how CoT affects users' perceptions and downstream behavior in real decision contexts remains limited. To address this gap, we develop and evaluate a reasoning-enabled conversational agent in the domain of sustainable consumption, where uncertainty, value-laden trade-offs, and multi-criteria decisions make reasoning transparency especially salient. We employ a three-phase research design that combines model development with empirical evaluation. In Phase 1, we fine-tune a conversational agent using a multi-agent framework to produce high-quality CoT data, leveraging Group Relative Policy Optimization (GRPO) and Low-Rank Adaptation (LoRA). Phase 2 tests the agent in a between-subjects experiment (N = 417) comparing a CoT-enabled chatbot with a standard chatbot. Grounded in the Technology Acceptance Model, IS Success Model, Cognitive Load Theory, and Privacy-Calculus Theory, we examine how user-friendliness, usefulness, personalization, trust, transparency, cognitive load, inefficiency, and privacy concerns mediate CoT effects on perceived knowledge and behavioral intentions. Results show that CoT significantly increases perceived knowledge and behavioral intentions. Specifically, PLS-SEM supports our mediation model, demonstrating that CoT acts through promoters such as enhanced trust, perceived usefulness, and personalization, which outweigh inhibitors like cognitive load and privacy concerns. Phase 3 complements these findings through qualitative content analysis, indicating that CoT improves user experience by providing clearer, more helpful reasoning, whereas standard chatbots are often perceived as verbose, vague, or technically unreliable. Overall, the study provides empirical evidence on how reasoning-enabled conversational systems shape user perceptions and behavioral outcomes, offering actionable guidance for designing decision-support transparent AI in complex domains.

1. Introduction

Sustainable consumption has become an urgent global priority as climate change, resource depletion, and waste generation increasingly threaten ecological and social systems worldwide (IPCC, 2023). In response, the United Nations' Sustainable Development Goal (SDG) 12 calls for responsible consumption and production patterns, emphasizing the need to decouple economic growth from environmental degradation while ensuring long-term societal well-being (United Nations, 2015).

However, despite rising awareness, a persistent gap remains between consumer attitudes and actual behavior (Carrington et al., 2010; ElHaffar et al., 2020). From an information management perspective, this gap is not merely motivational but fundamentally cognitive: sustainable choices are reasoning-intensive and place substantial demands on consumers' information-processing capabilities (Ries & Rese, 2025). Consumers must integrate environmental, social, and economic considerations, often under conditions of limited knowledge, uncertainty, and competing goals (Trudel, 2019; White et al., 2019).

Abbreviations: SDG, Sustainable Development Goal; LLMs, Large Language Models; CoT, Chain-of-Thought; XAI, Explainable Artificial AI; TAM, Technology Acceptance Model; PLS-SEM, Partial Least Squares Structural Equation Modeling; GRPO, Group Relative Policy Optimization; LoRA, Low-Rank Adaptation; KL, Kullback-Leibler; RLHF, Reinforcement Learning from Human Feedback; AVE, Average Variance Extracted; HTMT, Heterotrait–Monotrait ratio of Correlations; VIFs, Variance Inflation Factors; PLSc, Consistent PLS; GSPO, Group Relative Sequence Optimization.

* Correspondence to: Fürst-Franz-Josef-Strasse, Vaduz 9490, Liechtenstein.

E-mail addresses: pejman.ebrahimi@uni.li, pejman.ebrahimi77@gmail.com (P. Ebrahimi), stefan.hoffmann@bwl.uni-kiel.de (S. Hoffmann), johannes.schneider@uni.li (J. Schneider).

<https://doi.org/10.1016/j.ijinfomgt.2026.103075>

Received 26 January 2026; Received in revised form 2 May 2026; Accepted 3 May 2026

Available online 12 May 2026

0268-4012/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

In the past, digital information systems developed to address these cognitive barriers have evolved from static information provision, such as simple eco-labels (Gorton et al., 2021) via feedback-based carbon footprint calculator apps (Hoffmann et al., 2022a; Hoffmann et al., 2024a; Lasarov et al., 2024) to early interactive tools like augmented reality (Hoffmann et al., 2022b; Joerss et al., 2021) that enable more user interaction. Recently, the field has progressed toward conversational recommender systems based on large language models (LLMs) that allow human-like conversations and adapt interactively to user preferences (Bi et al., 2024; Jannach et al., 2021). However, a critical limitation remains: Although the system responds to the user and provides adapted recommendations, most existing systems operate as 'black boxes'. They rely on simplified signals offering limited insight into why a particular option is more sustainable. In high-stakes, value-laden domains like sustainability, accuracy alone is insufficient (Rai, 2020; Zhang & Chen, 2020). Looking toward the field's future development, information systems must evolve from simple retrieval tools into reasoning-oriented decision aids to fully bridge the gap between algorithmic complexity and human understanding. To bridge the gap between algorithmic complexity and human understanding, information systems must evolve from simple retrieval tools into reasoning-oriented decision aids. Recent advances in Chain-of-Thought (CoT) reasoning offer a novel mechanism to achieve this "glass box" transparency. By explicitly generating intermediate reasoning steps, CoT-enabled LLMs can articulate the causal and normative justifications underlying a recommendation (Wei et al., 2022). This capability mimics a human advisor, transforming the interaction from a static output into an explainable process (Miller, 2019). Yet, despite the technical potential of CoT, empirical evidence remains scarce regarding its actual impact on user psychology. It remains unclear whether exposing the "reasoning" of an AI agent effectively improves consumer knowledge or translates into stronger sustainable behavioral intentions.

Rather than merely identifying an unexplored empirical gap (Alvesson & Sandberg, 2011), this study problematizes the foundational assumption prevailing in Explainable AI (XAI) and Information Systems research: the premise that exposing the internal logic of an AI inherently improves user understanding and technology acceptance. We argue that in complex, value-laden domains like sustainable consumption, achieving 'glass-box' transparency through CoT reasoning creates a critical intellectual puzzle. Specifically, while explicit reasoning may reduce uncertainty, it simultaneously risks triggering cognitive overload or privacy concerns by revealing the depth of data processing involved. Thus, we introduce a necessary narrative tension: does the cognitive relief provided by transparent AI reasoning outweigh the cognitive friction and privacy anxieties it may induce? By navigating this tension, our study investigates how and through which competing psychological mechanisms CoT-based transparency affects users' knowledge and behavioral intentions.

To answer these questions, we adopt a three-phase research design that integrates technical model development with empirical user studies. Phase 1 involves the development of a specialized LLM fine-tuned to generate high-quality CoT reasoning data for sustainable consumption scenarios. Phase 2 implements a between-subjects experiment (N = 417) comparing this reasoning-based chatbot against a standard, non-reasoning chatbot. To comprehensively map the psychological mechanisms through which CoT reasoning affects user outcomes, we integrate multiple complementary theoretical perspectives that capture the dual nature of AI transparency. Building on the technology adoption literature, the Technology Acceptance Model (TAM, Davis, 1989) and the IS Success Model (DeLone & McLean, 2003) provide foundational frameworks for understanding how specific system characteristics—such as explicit reasoning transparency and personalization—shape users' perceived usefulness and trust. Conversely, to account for the potential negative psychological impacts of detailed explanations, Cognitive Load Theory (Sweller, 1988) explains how structured reasoning may either reduce or exacerbate information

processing demands, while Privacy Calculus Theory (Culnan & Bies, 2003) addresses potential concerns arising from explicit data usage disclosure. Together, these frameworks enable us to theoretically ground and test the competing psychological mechanisms—promoters and inhibitors—that may simultaneously operate when users interact with reasoning-enabled AI. This multi-theoretical approach is not a lack of focus, but rather a necessary architecture for understanding the complex, potentially contradictory psychological responses to AI transparency. Partial Least Squares Structural Equation Modeling (PLS-SEM) is used to test these mediating mechanisms. Phase 3 complements these findings through qualitative content analysis of open-ended user feedback. The triangulation of the theory-driven quantitative analysis and the exploratory qualitative investigation of specific drivers provides a novel theoretical contribution to understanding how consumers process conversations with explainable AI and offers a foundation for future research and practical applications.

The novelty and originality of this research lie in bridging the gap between technical AI engineering and behavioral information systems research. While prior computer science studies have demonstrated the computational accuracy of CoT algorithms (Alghamdi, 2025; Kojima et al., 2022; Wei et al., 2022), this study originally shifts the analytical lens to evaluate the human-centric psychological mechanisms—testing how these 'glass-box' models influence user cognition, trust, and behavior in high-stakes, value-laden contexts. By doing so, this study makes several contributions to information management research. First, it advances explainable AI research (Bauer et al., 2023; Shin, 2021) by providing empirical evidence on whether and how exposing reasoning processes affects user psychology beyond task performance metrics. While technical AI research has established CoT's computational benefits, we demonstrate its impact on knowledge acquisition, and behavioral intentions—outcomes central to IS adoption but rarely examined in AI engineering studies. Second, it extends the sustainable consumption literature (Laukkanen et al., 2022; White et al., 2019) by evaluating how "glass box" AI support affects perceived knowledge and behavioral change in a domain where information deficits are a primary barrier. Third, the study contributes methodologically by bridging advanced AI engineering (Jiang et al., 2026) and user-centered information management research (Foroughi et al., 2025; Lasarov et al., 2026; Ma et al., 2025), demonstrating how fine-tuned reasoning agents can be empirically evaluated not just for accuracy but for their psychological and behavioral effects. Finally, by explicating the mediating mechanisms through which reasoning transparency operates, the findings offer actionable design guidance applicable beyond sustainability to any domain requiring justification-based decision support—including healthcare, financial planning, and policy decisions where users need to understand not just *what* to do but *why* certain choices are recommended.

2. Related works

2.1. Sustainable consumption

Sustainable consumption refers to patterns of consumption that minimize negative environmental and social impacts while maintaining individual well-being. It encompasses behaviors such as choosing environmentally friendly products and reducing resource use across all phases of consumption from acquisition via usage to disposal. White et al. (2019, p. 3) define sustainable consumer behavior as "actions that result in decreases in adverse environmental impacts as well as decreased utilization of natural resources across the lifetime of the product, behavior, or service".

Despite widespread awareness of sustainability issues, research consistently demonstrates a pronounced attitude-behavior gap: many consumers who express positive attitudes toward sustainable consumption fail to translate these intentions into actual behavior (Carrington et al., 2010; Munro et al., 2023). This gap remains a central

challenge in sustainable consumption research (White et al., 2019). Consumers often endorse sustainability in principle but default to convenience, habit, or price considerations in concrete consumption situations. Beyond motivational barriers, this discrepancy is increasingly attributed to cognitive limitations, information deficits, and the complexity of sustainability-related decisions (Hoffmann et al., 2024a; Trudel, 2019).

The SHIFT framework proposed by White et al. (2019) identifies five psychological drivers that can promote sustainable consumer behavior: social influence, habit formation, individual self, feelings and cognition, and tangibility. Within this framework, information plays a particularly important role in strengthening the cognition component by shaping consumers' understanding of consequences and trade-offs. According to the SHIFT framework, consumers are more likely to translate positive sustainability attitudes into action when they have access to clear, comprehensible information that reduces uncertainty and when they trust the source and credibility of that information (White et al., 2019). Consumers need information about the desired and the undesired behaviors and which consequences they have. Consumers frequently lack accurate, decision-relevant information about the environmental impacts of their consumption, such as carbon emissions, energy use, or resource intensity. Empirical evidence shows that individuals systematically underestimate the environmental impact of high-impact activities and products while overestimating the impact of low-impact behaviors (Attari et al., 2010; van Bussel et al., 2022). They must have easy access to information without the amount of information being too complex, confusing and overwhelming (Hoffmann et al., 2022b).

Traditional approaches to informing consumers, such as eco-labels, sustainability scores, or carbon footprint indicators, aim to reduce this gap by simplifying complex information or by providing personalized feedback (Gorton et al., 2021; Hoffmann et al., 2022a; Hoffmann et al., 2024a). However, sustainable consumption decisions typically involve multidimensional trade-offs across environmental, social, and economic criteria, many of which are abstract, uncertain, or difficult to verify (Trudel, 2019). These characteristics make sustainability information cognitively demanding to process and often difficult to trust. As a result, consumers may feel ill-equipped to evaluate the actual sustainability implications of available options, leading them to rely on heuristics, defaults, or convenience-oriented choices rather than deliberative evaluation. The low-cost hypothesis further supports the need for such systems by suggesting that individuals are more likely to engage in pro-environmental behavior when the required effort is low (Diekmann & Preisendörfer, 2003). Applied to sustainable consumption, this implies that consumers need easy, low-effort access to reliable, interpretable sustainability information at the point of decision.

Against this background, we argue that digital information tools are essential for addressing the persistent information gap in sustainable consumption (Grinstein et al., 2018). Digital tools, such as particularly conversational agents and chatbots, offer a promising avenue in this respect. By providing personalized, interactive, and on-demand explanations, conversational agents can reduce search effort, lower cognitive barriers, and support reasoning about complex sustainability trade-offs. In doing so, they have the potential to bridge the gap between sustainability intentions and actual behavior by aligning consumers' values with informed, confident decision-making. However, consumers require decision support systems that not only present sustainability information but also explain causal relationships, justify recommendations, and make the reasoning and trade-offs explicit.

2.2. Recommender systems and conversational agents

Digital decision support systems such as recommender systems, conversational agents, and chatbots have become central components of contemporary information environments. Recommender systems are algorithmic tools designed to predict users' preferences and suggest relevant items, information, or actions, thereby reducing decision

complexity (Ricci et al., 2022). Traditionally embedded in e-commerce, media platforms, and online services, these systems rely on techniques such as collaborative filtering, content-based filtering, and hybrid approaches to personalize recommendations at scale (Adomavicius & Tuzhilin, 2005). More recently, conversational agents, often referred to as chatbots when implemented in text-based interfaces, have emerged as an interactive extension of recommender systems. Conversational agents enable natural language interaction, allowing users to articulate preferences, ask questions, and receive adaptive responses in real time. By shifting from one-shot recommendation delivery to dialog-based interaction, these systems enhance user engagement, lower interaction barriers, and support iterative preference construction, positioning them as powerful information management tools in complex decision contexts (Jannach et al., 2021; Jannach, 2023).

Recent advances in generative artificial intelligence (GenAI), particularly LLMs such as ChatGPT, have substantially expanded the capabilities of conversational agents beyond predefined intents and scripted dialog flows. Unlike earlier rule-based or retrieval-based chatbots, GenAI-powered agents can generate context-sensitive, coherent, and linguistically rich responses, allowing for open-ended interaction, dynamic personalization, and flexible information provision across domains (Duan et al., 2019; Schneider, 2025). Studies show that consumers believe AI recommendations in particular in utilitarian rather than in hedonic consumption contexts (Longoni & Cian, 2022).

In the context of sustainable consumption, recommender systems and conversational agents can serve as digital interventions to support more environmentally and socially responsible consumer choices (Joerss et al., 2021). Recommender systems can incorporate sustainability-related attributes, such as eco-labels, carbon footprints, or ethical certifications, into product rankings and suggestions, thereby nudging consumers toward more sustainable alternatives without eliminating choice freedom. Conversational agents further extend this potential by providing personalized sustainability information on demand, responding to user queries, and embedding sustainability considerations into everyday decision processes (Bi et al., 2024; Pizzi et al., 2021).

2.3. Reasoning-oriented language models and fine-tuning for decision support

As digital decision environments become increasingly complex, there is growing recognition that effective decision support systems must go beyond presenting outcomes. Consumers require systems that not only provide sustainability-related information but also explain causal relationships, justify recommendations, and make trade-offs transparent across competing criteria. In the context of recommender systems, this objective is reflected in the literature on explainable recommendation, which seeks to answer the question of "why" by providing explanations that help users understand the rationale behind algorithmic suggestions, thereby increasing transparency, trust, and perceived usefulness (Zhang & Chen, 2020).

Recent advances in LLMs have opened new avenues for addressing these challenges through reasoning-oriented architectures, particularly chain-of-thought models that explicitly generate intermediate reasoning steps before producing a final output (Wei et al., 2022). By externalizing the reasoning process, such models enable users to follow how information is weighted, how causal links are inferred, and how conclusions are derived, rather than merely observing a recommendation or score. Importantly, these capabilities can be further enhanced through fine-tuning, which adapts general-purpose language models to domain-specific data, objectives, and reasoning patterns, thereby improving their relevance, consistency, and alignment with application contexts (Liu et al., 2023). Reasoning-based and fine-tuned language models thus represent a significant shift from opaque prediction systems toward interactive decision support tools that can actively scaffold users' understanding and confidence in complex decision situations.

Chain-of-thought models and fine-tuned language models are particularly well suited to stimulating sustainable consumption because they directly address the cognitive, informational, and trust-related barriers of sustainable consumption. CoT models explicitly articulate intermediate reasoning steps, which can help consumers structure their thinking, understand why certain options are more sustainable, and reduce ambiguity associated with sustainability claims. Fine-tuning further enhances this potential by aligning language models with domain-specific sustainability knowledge, decision heuristics, and contextual constraints, thereby increasing relevance and perceived competence. Despite these theoretically compelling advantages, empirical evidence on whether and how reasoning-oriented and fine-tuned language models actually influence sustainable consumption-related outcomes remains scarce.

2.4. Explainable AI, reasoning, and user impact

The call for XAI has intensified as AI systems become more pervasive in high-stakes decision contexts (Longo et al., 2024; Schneider, 2024). Explainability refers to the degree to which humans can understand the causes of AI decisions and predictions (Miller, 2019). Traditional machine learning models, particularly deep neural networks, operate as "black boxes" whose internal logic remains opaque to users, undermining trust and accountability (Rudin, 2019). XAI emerged to address this limitation by developing techniques that make AI decision-making transparent and interpretable, thereby enabling users to assess whether outputs are reasonable, identify potential biases, and maintain human oversight (Adadi & Berrada, 2018; Doshi-Velez & Kim, 2017).

CoT reasoning represents a distinctive approach to explainability in LLMs. Unlike post-hoc explanation techniques that attempt to reverse-engineer decisions after they are made (e.g., LIME, SHAP; Ribeiro et al., 2016), CoT generates explanations *intrinsically* by articulating intermediate reasoning steps as part of the generation process itself (Wei et al., 2022; Yao et al., 2024). This approach aligns more closely with human explanatory norms: people typically justify decisions by describing their reasoning process rather than by decomposing internal cognitive mechanisms (Miller, 2019). Recent technical research demonstrates that CoT improves model performance on complex reasoning tasks by 20–50% compared to direct answer generation (Kojima et al., 2022; Wei et al., 2022), and that models trained on CoT reasoning exhibit greater logical consistency and fewer factual errors (Chen et al., 2025; Chen et al., 2025; Lightman et al., 2023).

However, technical performance metrics do not directly answer the critical IS question: Do users benefit from CoT transparency? A growing body of research distinguishes between *system explainability* (technical interpretability) and *user understanding* (psychological outcomes; Hoffman et al., 2018; Meske et al., 2022). Studies show that technically "interpretable" models do not always lead to better user comprehension or trust (Kaur et al., 2020; Poursabzi-Sangdeh et al., 2021). In some cases, exposing too much complexity overwhelms users (Chromik et al., 2021), while in others, explanations fail to align with users' mental models or information needs (Liao & Sundar, 2022). This research underscores that explainability must be evaluated not only by computational metrics but by its impact on user psychology and behavior (Hoffman et al., 2018; Ribeiro et al., 2016).

Recent empirical studies on XAI and user outcomes reveal mixed findings. On one hand, explanations have been shown to increase trust (Kizilcec, 2016; Shin & Park, 2019), perceived usefulness (Pu & Chen, 2008), and decision confidence (Zhang & Chen, 2020) in recommender systems and decision support contexts. Tintarev and Masthoff (2015) identify multiple goals of explanations, including transparency (helping users understand system logic), trust (increasing confidence), and effectiveness (helping users make better decisions). Conversely, existing research highlights several unintended drawbacks of explainability. When explanations become overly complex, they can impose additional cognitive load on users, making systems harder rather than easier to

understand. At the same time, revealing how data is utilized may heighten privacy concerns, as users become more aware of the extent of data collection and processing (Chromik et al., 2021; Knijnenburg et al., 2013; Najafian et al., 2021). Paradoxically, explanations can also erode trust, particularly when they expose system limitations or uncertainties that users might otherwise overlook (Schilke et al., 2025; Bansal et al., 2021).

The impact of CoT reasoning on user outcomes remains largely unexplored in information systems research. While computer science studies have validated CoT's computational advantages, empirical evidence on its psychological effects is scarce. Recent work suggests that reasoning explanations may be particularly effective for complex, multi-step problems where users need to follow logical progression (Chen et al., 2025; Chen et al., 2025; Kovari, 2025), but these findings are preliminary and often based on expert evaluations rather than end-user studies. Critically, no research has systematically examined whether CoT-based transparency affects knowledge acquisition and behavioral intentions in real-world decision contexts—particularly in domains like sustainable consumption where recommendations involve normative judgments and value trade-offs, not just factual accuracy.

This gap is significant because sustainable consumption decisions differ fundamentally from typical AI tasks. Unlike factual queries (e.g., "What is the capital of France?") or optimization problems (e.g., "Find the shortest route"), sustainability decisions require balancing competing values (environmental vs. economic), interpreting uncertain information (carbon footprint estimates), and justifying normative choices (why ethical sourcing matters). In such contexts, CoT reasoning may serve not merely as a technical explanation but as a decision scaffold that helps users structure their thinking, weigh trade-offs, and build confidence in complex choices. However, it may also overwhelm users with excessive detail or trigger concerns about data usage. Understanding these competing effects requires empirical investigation of the psychological mechanisms through which CoT transparency operates—an investigation that technical AI research alone cannot provide.

The present paper addresses the outlined gap by empirically testing the effects of CoT-based and fine-tuned conversational agents in a controlled experimental setting. We draw on the following constructs and theories. We expect that features of CoT and fine-tuned language models improve perceived usefulness and ease of use of the conversational agents. These constructs are core determinants of technology adoption according to the TAM (Davis et al., 1989; Davis, 1989). Simultaneously, these features increasing information quality, which is conceptualized in the IS Success Model (DeLone & McLean, 1992; DeLone & McLean, 2003). At the same time the additional information may increase the cognitive processing and the anxiety that private data is misused. Therefore, we also consider these effects from the perspective of the Cognitive Load Theory (Sweller, 1988) and the Privacy-Calculus-Theory (Culnan & Bies, 2003).

2.5. Review of empirical literature

Although prior empirical research in information systems, information management, marketing, and sustainability has informed our study, no existing work has directly addressed our research question at the intersection of reasoning transparency, conversational AI, and sustainable consumption.

A first stream of research examines how digital recommendations influence sustainable consumption behavior. For example, Satinet et al. (2025) compare traditional recommendations, sustainability-oriented recommender systems, and no recommendations in a controlled online shopping environment. Their findings show that sustainability-oriented recommendations increase the purchase of sustainable products by enhancing user interest and reducing search effort. Similarly, von Zahn et al. (2024) demonstrate how online retailers can reduce environmentally harmful behavior through digital interventions that combine green nudges with causal machine learning-based personalization. In a

related vein, Majid et al. (2025) study conversational AI chatbots as pro-social nudging mechanisms that can trigger pro-environmental behavioral spillovers beyond the immediate decision context. Furthermore, Silalahi, (2025) identify trust as a key driver of user adoption of GenAI-based sustainability recommendations.

However, this stream of literature primarily focuses on whether and under which conditions digital interventions influence sustainable behavior, while largely treating AI systems as black-box recommenders. In particular, these studies do not examine how the structure and transparency of AI-generated reasoning shape user perceptions and behavioral outcomes.

A second stream of research, primarily situated in marketing and human–AI interaction, investigates how users respond to explanations in conversational recommendation systems. Chen et al. (2024) show that post-hoc explanations increase perceived interpretability, which in turn enhances trust, purchase intention, and click-through behavior. Similarly, Chen, Lin, et al. (2025), Chen, Deng, et al. (2025) demonstrate that different explanation types (e.g., expert validation vs. consensus validation) increase perceived transparency and trust in conversational recommendations.

While this literature highlights the importance of explainability, it remains limited in two important respects. First, prior studies focus predominantly on post-hoc explanations, rather than intrinsic, process-based reasoning transparency such as CoT. Second, these studies are largely conducted in conventional consumption contexts and do not account for the unique characteristics of sustainability-related decisions, which are typically characterized by uncertainty, moral trade-offs, and multi-criteria evaluation.

Moreover, prior research indicates counterarguments regarding the value of explainability. Some studies suggest that increased transparency may impose cognitive costs, leading to information overload or reduced usability, and that users may prefer simple recommendations over complex explanations in low-involvement contexts (Awad et al., 2006; Dinev & Hart, 2006; Hoffmann et al., 2022b). This raises an unresolved question as to whether more detailed reasoning enhances or hinders decision-making.

Against this backdrop, our study addresses three critical gaps. First, we move beyond black-box recommendation systems by examining how reasoning transparency in the form of CoT affects user perceptions and behavior. Second, we integrate model development and empirical user evaluation, thereby linking technical advances in AI with behavioral outcomes. Third, we focus on sustainable consumption as a high-relevance context, where reasoning transparency is expected to play a particularly important role in resolving uncertainty and supporting informed decision-making.

To the best of our knowledge, this is the first study to combine the development of a reasoning-enabled conversational agent with a causal empirical evaluation of its effects on user perceptions, knowledge acquisition, and behavioral intentions in the domain of sustainable consumption.

3. Research design

This study is organized into three comprehensive studies (phases) (Fig. 1) that collectively address the development, evaluation, and user experience of CoT enhanced LLMs for sustainable consumption behavior. Study 1 (Phase 1) centers on the technical fine-tuning of the Qwen2.5-3B-Instruct¹ language model, leveraging a sophisticated multi-agent system to generate high-quality CoT reasoning examples within the sustainable consumption domain. These data² are publicly shared via Hugging Face to ensure transparency and reproducibility. The fine-

tuning process employs advanced techniques such as Group Relative Policy Optimization (GRPO) and Low-Rank Adaptation (LoRA), with training progress rigorously monitored through metrics including reward signals, Kullback-Leibler (KL) divergence, and loss functions. Study 2 (Phase 2) implements a robust quantitative experimental design involving IT specialists, utilizing a between-subjects framework to compare the CoT-enabled chatbot against a standard baseline. PLS-SEM with 5000 bootstraps is applied to test eight mediating mechanisms—spanning trust, transparency, cognitive load, and personalization—that influence knowledge acquisition and behavioral intentions. This phase provides strong causal evidence supporting the efficacy of CoT reasoning in enhancing user outcomes. Study 3 (Phase 3) complements these findings through a qualitative content analysis of open-ended feedback from participants, employing inductive coding to identify six thematic categories that capture nuanced user perceptions of the chatbot's reasoning transparency, response length, and overall experience. Together, these phases offer a holistic and rigorous evaluation of CoT-enhanced language models, bridging technical innovation with behavioral validation and user-centered insights.

4. Study 1 (Phase 1): Model fine-tuning

4.1. Methodology

We utilized the Qwen2.5-3B-Instruct model (Yang et al., 2025) as our base model due to its strong instruction-following capabilities and efficiency.

4.1.1. GRPO for reasoning

To enhance the model's reasoning capabilities, we employed GRPO, an advanced reinforcement learning technique designed to improve logical consistency in LLMs. Unlike standard reinforcement learning which scores answers in isolation, GRPO evaluates a group of responses relative to each other to encourage diverse and accurate reasoning paths (Shao et al., 2024). By generating a cohort of multiple potential reasoning chains for a single sustainability query and ranking them within that group, the model learns to distinguish between superficial answers and those that offer robust, step-by-step justification.

This approach is particularly valuable for the domain of sustainable consumption, where decision-making often involves navigating complex trade-offs rather than retrieving a single static fact. GRPO eliminates the need for a separate value function approximation—common in traditional Reinforcement Learning from Human Feedback (RLHF)—thereby reducing computational overhead while stabilizing the training process. This ensures the model effectively learns to structure its output as a coherent "chain of thought," providing the transparency required for the subsequent user experiments (Guo et al., 2025).

4.1.2. Data for fine-tuning

The development of high-quality training data represented a critical component in the fine-tuning process (Ebrahimi & Schneider, 2025). We implemented a multi-agent synthetic data generation pipeline (Fig. 2) specifically designed to create CoT reasoning examples focused on sustainable consumption behavior. This pipeline consisted of four specialized agents working in sequence to ensure comprehensive coverage of the domain knowledge while maintaining reasoning quality.

The data generation process began with a "Question Answer Generator Agent" that created domain-specific queries related to sustainable consumption behavior. These questions were designed to cover diverse aspects including consumer decision-making, green marketing strategies, policy implications, and behavioral economics principles within sustainability contexts. The generated questions underwent evaluation by an "Evaluator Agent" that assessed their relevance, complexity, and alignment with sustainable consumption research frameworks.

Questions passing the evaluation criteria were then processed by a "Reasoning Steps/CoT Generator Agent", which produced detailed step-

¹ Link to model: <https://huggingface.co/Qwen/Qwen2.5-3B-Instruct>

² Link to data: <https://huggingface.co/datasets/arad1367/COT-Sustainable-Consumption-Behavior>

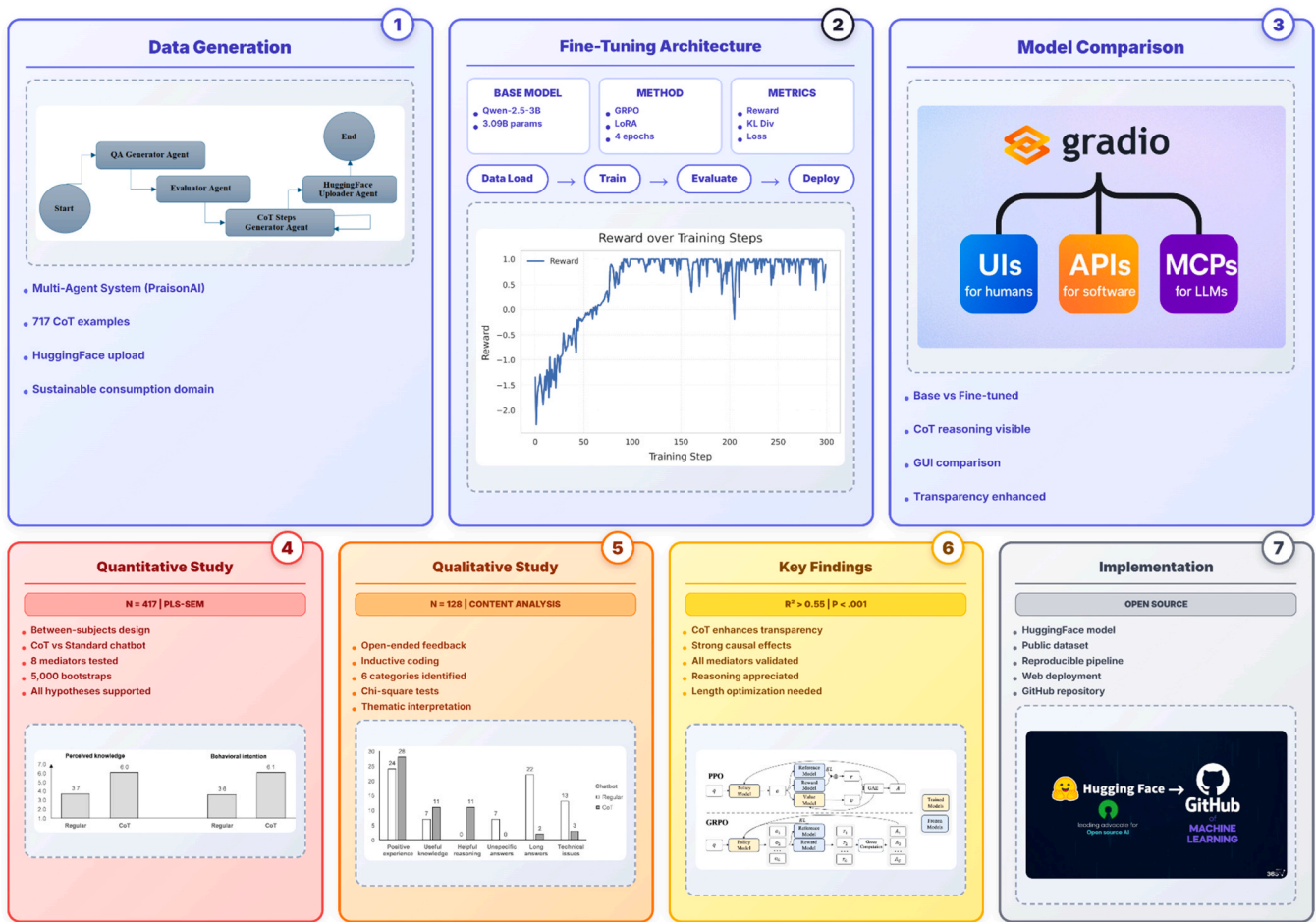


Fig. 1. Research methodology workflow. (Source: Authors' own work)

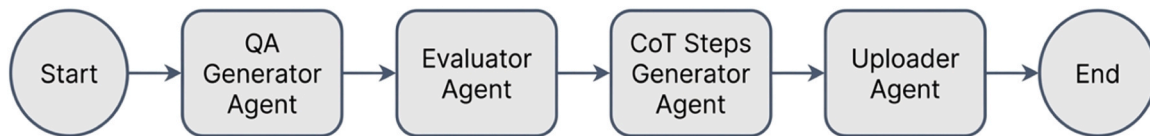


Fig. 2. Multi-agent synthetic data generation pipeline with PraisonAI (n.d.). (Source: Authors' own work using PraisonAI)

by-step reasoning paths leading to comprehensive answers. This agent was specifically designed to articulate intermediate reasoning steps that demonstrate logical progression, consideration of multiple factors, and integration of domain knowledge. The reasoning chains explicitly modeled the cognitive processes underlying sustainable consumption decisions, providing exemplars of desired reasoning patterns for the model to learn from.

The final component of the pipeline, the “Uploader Agent”, standardized the format of the generated data and published it to a public repository (Fig. 3). The resulting dataset comprised 717 question-answer pairs with corresponding reasoning chains, covering a comprehensive range of sustainable consumption topics. Each entry contained a question, the reasoning process, and the final answer, structured to facilitate the model's learning of both content knowledge and reasoning methodology.

This synthetic data generation approach offered several advantages over manual annotation, including scalability, consistency in reasoning structure, and comprehensive domain coverage. The multi-agent

architecture ensured quality control at each stage of the generation process, resulting in training data that effectively captured the complexity of sustainable consumption reasoning while maintaining coherence and accuracy.

4.2. Results

The fine-tuning process was conducted over 300 steps using the GRPO methodology. To validate that the model successfully learned to generate reasoning chains without losing its foundational knowledge, we monitored key training metrics including reward scores and KL divergence.

As detailed in Appendix A (see Fig. A1 and Table A1), the training progression followed a characteristic S-curve. Initial steps showed negative reward scores as the model explored reasoning patterns, then rapidly improved and stabilized around step 150. The KL divergence metric confirmed that while the model adapted to the specific reasoning format required for sustainable consumption advice, it remained within

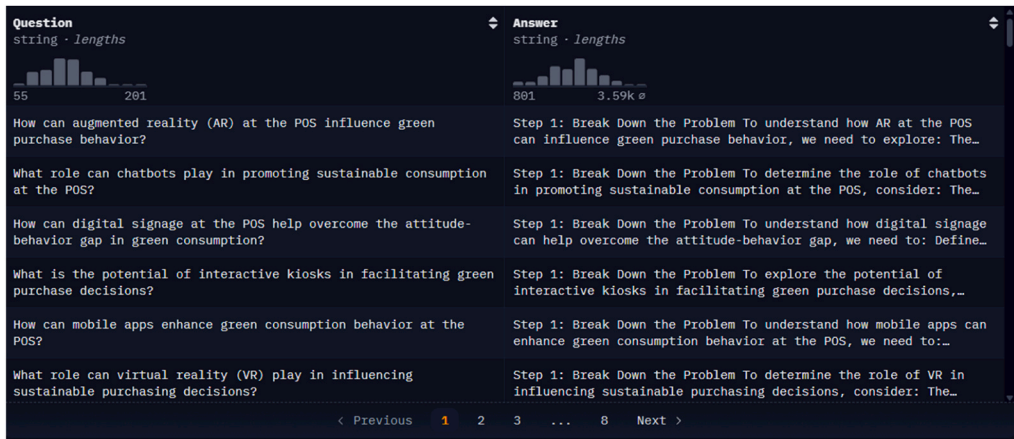


Fig. 3. Display of some rows of the CoT sustainable consumption behavior dataset from Hugging Face. (Source: Authors' own work)

a safe probability distribution relative to the base model, preventing "catastrophic forgetting" of general knowledge.

Beyond numerical metrics, comparative analysis confirmed that the fine-tuned model successfully transitioned from providing direct, unexplained answers to generating structured, step-by-step CoT explanations. Table 1 illustrates this shift, comparing a response from the baseline model against the fine-tuned reasoning agent.

While standard performance metrics such as format adherence showed a quantitative improvement in the fine-tuned model (see Appendix A), the primary contribution of the fine-tuning process was a fundamental qualitative shift in decision support capabilities. The baseline model, despite its general competence, operated as an opaque "black box," typically delivering direct recommendations without exposing the underlying rationale. This lack of transparency forces users to rely blindly on the system's authority.

In contrast, the fine-tuned agent successfully adopted a "glass box" approach. By consistently generating explicit CoT reasoning paths, the model demonstrated the capability to externalize its decision-making process. This is not merely an incremental improvement in accuracy, but a structural change in how information is presented: users are no longer just given an answer; they are guided through the trade-offs,

Table 1
Comparative performance analysis (Base vs. CoT).

| Evaluation Metric | Base Model | Fine-Tuned Model (Final Phase) | Improvement |
|-----------------------------|---------------------|--------------------------------|---------------|
| Reasoning Structure | Direct answer | Step-by-step CoT | Qualitative |
| Structured Format Adherence | 0% (0.00) | 45.6% (0.456) | + 45.6 pp |
| Response Consistency (1/SD) | 3.79 (1/0.264) | 3.21–10.53 | Variable |
| Average Response Length | 150–200 tokens | 207.8 ± 28.4 tokens | + 10–40% |
| Reasoning Transparency | Implicit | Explicit with justification | Qualitative |
| Multi-factor Analysis | Limited | Comprehensive trade-offs | Qualitative |
| Policy Divergence (KL) | 0.00 (reference) | 0.088 ± 0.022 | + 0.088 |
| Training Efficiency | 100% parameters | 3.99% parameters (LoRA) | 96% reduction |
| Inference Speed | Baseline | ~Equivalent (vLLM) | Maintained |
| Domain Adaptation | General instruction | Sustainable consumption | Specialized |

Note: Base model evaluated on 717 sustainable consumption queries; Fine-tuned metrics from final phase (steps 151–300). Structured Format Adherence measures CoT template compliance; KL divergence < 0.1 indicates controlled adaptation without catastrophic forgetting; pp - percentage points.

causal links, and normative considerations that lead to that answer. This capability is critical for sustainable consumption, where trust relies less on the "correctness" of a prediction and more on the justification of complex, value-laden choices. This technical validation ensures that the artifact used in the subsequent user study (Phase 2) effectively represents a "reasoning-based" system capable of transparent decision support.

5. Study 2 (Phase 2): Survey-based study

5.1. Theoretical background & hypotheses development

Transparency, understood as the degree to which a system's internal logic and decision-making processes are made visible and comprehensible to users, has been extensively studied as a key driver of perceived ease of use in AI systems (Chen et al., 2025; Chen et al., 2025; Kovács & Horváth, 2025; Schilke & Reimann, 2025). Recent research emphasizes that transparent systems reduce uncertainty, facilitate mental model formation, and lower the cognitive burden associated with evaluating system outputs (Joshi et al., 2024). In the rapidly evolving landscape of large language models and conversational agents, transparency has emerged as a critical factor distinguishing trustworthy systems from opaque "black boxes" (Chen et al., 2025; Chen et al., 2025). Studies demonstrate that when users understand how AI systems arrive at conclusions—such as a chatbot articulating the reasoning steps behind a sustainability recommendation—they experience reduced cognitive friction and enhanced confidence in the system's outputs (Kovács & Horváth, 2025).

The emergence of CoT reasoning in LLMs represents a paradigm shift in how transparency can be operationalized (Chen et al., 2025; Chen et al., 2025; Wei et al., 2022). Unlike traditional post-hoc explanations appended to recommendations, CoT models generate intermediate reasoning steps as an integral part of the response, mimicking human problem-solving processes. Recent evidence confirms that CoT-based reasoning enhances model interpretability and transparency by allowing users to inspect the reasoning path, making it easier to diagnose errors and understand decision-making processes (Chen et al., 2025; Chen et al., 2025). Research demonstrates that CoT disclosure builds trust and accountability by transforming AI from an incomprehensible oracle into a collaborative partner whose logical steps can be followed and evaluated by users (Chen et al., 2025; Chen et al., 2025). This capability is particularly valuable in contexts requiring justification and normative reasoning, where users need to understand not just *what* the system recommends but *why* and *how* it arrived at that conclusion.

In the context of sustainable consumption, this mechanism is particularly relevant. Traditional "black box" recommender systems

present outcomes (e.g., "Product A is more sustainable") without revealing the underlying logic, forcing users to mentally reconstruct the decision criteria—a process that widens the "gulf of evaluation" (Norman, 2013) and increases cognitive load. When users receive an unexplained recommendation, they must infer whether it accounts for carbon footprint, ethical sourcing, local production, or other sustainability dimensions. This inference process is cognitively demanding and error-prone, particularly in high-complexity domains like sustainability where multiple conflicting criteria must be weighed (Trudel, 2019; White et al., 2019).

CoT-enabled chatbots directly address this cognitive bottleneck by externalizing the reasoning process. By explicitly articulating intermediate steps (e.g., "Given your preference for organic products and low carbon footprint, I compared local non-organic options versus imported organic options based on..."), CoT reduces the interpretive effort required to understand *why* a recommendation is made and *how* it aligns with user values. Recent research on cognitive load in conversational AI confirms that reducing information processing complexity through structured, concise explanations significantly lowers cognitive load and improves task completion (Zhao et al., 2025). A 2025 study on visually impaired assistance agents demonstrated that delivering brief, well-sequenced information via a "Less is More" principle reduced cognitive load by 39.9% and improved usability scores, confirming that structured reasoning outputs minimize extraneous cognitive processing (Zhao et al., 2025).

This ease of understanding is conceptually distinct from general interface usability or user-friendliness. While interface design, response time, and navigation structure influence overall user-friendliness, perceived ease of understanding specifically captures the cognitive accessibility of system outputs (Venkatesh & Bala, 2008). CoT's contribution operates at this informational level: it makes the *content* easier to process by providing structured reasoning scaffolds, independent of other interface factors. Research on explainable AI chatbots confirms that transparency in decision-making—showing users the logic behind recommendations—dramatically improves comprehension and reduces the "black box" problem, making AI systems more intuitive and less complex to interpret (Kovari, 2025). A comprehensive 2025 review of explainable AI chatbots emphasizes that comprehensibility and human understanding focus on the ease with which explanations can be understood and the clarity with which they convey the model's reasoning to humans, which are central to effective human-AI interaction (Kovari, 2025).

The TAM (Davis et al., 1989; Davis, 1989) and its refinements (e.g., UTAUT, Venkatesh et al., 2003) identify perceived ease of use as a central determinant of technology adoption. In the context of AI-based decision support systems, perceived ease of use encompasses not only interface usability but also the cognitive effort required to interpret and act upon system outputs (Venkatesh & Bala, 2008). When systems are perceived as easy to use, users are more likely to engage with them, process information effectively, and translate this understanding into behavioral outcomes (Gefen et al., 2003; Shin, 2021).

Given this theoretical foundation, we expect that CoT-enabled chatbots will enhance perceived ease of understanding by reducing the cognitive effort required to interpret recommendations and evaluate their alignment with sustainability goals. This enhanced ease of understanding, in turn, should facilitate knowledge acquisition—users can more effectively process and retain sustainability-related information—and strengthen behavioral intentions by increasing confidence in decision-making (Davis, 1989; Venkatesh et al., 2003). Hence:

H1. Perceived user friendliness will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

According to the TAM, perceived usefulness, the degree to which a user believes a system supports task achievement, is a key determinant of behavioral intentions to use technology (Davis, 1989). Recent

research indicates that perceived usefulness is also critical for the adoption of generative AI tools (Lasarov et al., 2026; Ma et al., 2025). In the context of sustainable consumption, reasoning-based chatbots can enhance perceived usefulness by explaining not only what actions to take but also why certain options are more sustainable (Wei et al., 2022). By revealing intermediate reasoning steps and clarifying causal and normative relationships, these chatbots make the system's instrumental value more salient, support users' goal attainment, and facilitate learning. Enhanced perceptions of usefulness are expected to improve knowledge acquisition and increase intentions to engage in sustainable behaviors, as users better understand both the recommendations and their underlying rationale.

H2. Perceived usefulness of the chatbot will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

Next, we propose that perceived personalized information will mediate the influence of a reasoning-based chatbot. The IS Success Model by DeLone and McLean, 1992; DeLone and McLean, 2003, which provides a comprehensive framework for evaluating the effectiveness of information systems, identifies six interrelated dimensions of IS success, in terms of use and user satisfaction: system quality, information quality, service quality, use, user satisfaction, and net benefits. For the present study information quality is of particular relevance. Information quality assesses the accuracy, relevance, and timeliness of the information produced by the system. Personalization increases the relevance of information, thereby strengthening user engagement and persuasion (Kapoor & Kumar, 2025; Kim et al., 2025; Tam & Ho, 2006). Personalization is a defining characteristic of effective recommender systems and has been shown to enhance perceived relevance and engagement (Kocaballi et al., 2019) and reduce distress (Wu et al., 2026). Reasoning-based chatbots can tailor explanations to users' queries, values, and contexts, thereby increasing perceived personalization. Users who experience personalized reasoning are more likely to perceive the recommendations as meaningful and actionable, which should increase knowledge acquisition and sustainable behavioral intentions.

H3. Perceived personalization of information will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

The answers generated by GenAI often appear logical and they are often presented as facts. However, sometimes such answers are "hallucinations", they are wrong or misleading (Dale, 2021; Kim & Lee, 2026). Users value accuracy when they search for AI guidance (Wu et al., 2026). Therefore, trust is a central antecedent of acceptance and reliance on AI systems (Gefen et al., 2003; Glikson & Woolley, 2020; Kang et al., 2025). In particular, trust in the truthfulness of answer is an important aspect for conversational agents (Amrollahi et al., 2026; Lasarov et al., 2026; Shin, 2021; Yang et al., 2026). Reasoning-based chatbots that clearly articulate the rationale behind recommendations can enhance perceived trustworthiness by making the decision process more transparent (Wei et al., 2022). When users trust the system, they are more likely to accept its guidance, improve their understanding, and adopt recommended sustainable behaviors.

H4. Trust in the chatbot will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

Transparency refers to the clarity with which a system communicates its internal logic, operations, or reasoning processes (Rai, 2020). In AI-enabled conversational systems, transparency can mitigate uncertainty and increase user confidence in decision-making, which is a critical antecedent of reliance on recommender systems (Gefen et al., 2003; Shin & Park, 2019; Shin, 2021; Wei et al., 2022). CoT-enabled reasoning explicitly exposes intermediate steps in the recommendation process, allowing users to understand how trade-offs and sustainability

assessments are derived. This increased transparency is expected to strengthen perceived knowledge and foster sustainable behaviors. The relationship between transparency and user experience is often described as a "double-edged sword" in IS literature (Kizilcec, 2016). While transparency fosters trust, excessive detail can lead to information overload, potentially overwhelming the user and reducing satisfaction (Hibbeln et al., 2017). However, the domain of sustainable consumption is characterized by high ambiguity and conflicting values (e.g., local non-organic vs. imported organic). In such high-uncertainty contexts, the risk of "blind trust" in a black box is perceived as greater than the cost of processing additional information. Users require the "why" to validate that the system has considered the complex trade-offs inherent in sustainability. Therefore, we argue that in this specific context, the utility of justification provided by CoT reasoning outweighs the potential negative effects of increased information length.

H5. Perceived transparency of the chatbot will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

While explanations can facilitate learning, excessive or poorly structured reasoning may increase cognitive effort to process the information. Cognitive Load Theory posits that the amount of mental effort required to process information influences learning outcomes (Sweller, 1988). While reasoning-based chatbots provide structured explanations, excessive or poorly organized information can increase cognitive load, potentially diminishing knowledge acquisition and behavioral intentions (Hoffmann et al., 2022b). Therefore, we assume that CoT reduces cognitive load, which finally leads to higher levels of perceived knowledge and behavioral intentions.

H6. Perceived cognitive load will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

Extending the length of the answers of a chatbot without providing good explanations may also increase users' impression that the conversational agent provides ineffective or irrelevant information (Longoni et al., 2019). When CoT reasoning appears tentative, confusing, or unconvincing, users may perceive the chatbot as less effective. Such inefficacy perceptions can undermine persuasion of the gathered knowledge and weaken intentions to behave in the recommended manner. Reasoning-based chatbots may reduce the user's perception of inefficiency.

H7. Perceived inefficiency of the chatbot will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

Privacy concerns represent a well-documented barrier to adoption of AI-based systems, particularly when personalization requires user data (Awad and Krishnan, 2006; Dinev & Hart, 2006) and in particular with regard to GenAI (Lasarov et al., 2026; Puntoni et al., 2021). In recommender systems, a well-established body of research has documented that explanations can paradoxically heighten privacy awareness by making data processing more visible to users (Knijnenburg et al., 2013; Najafian et al., 2021). According to the Privacy Calculus Theory, individuals evaluate the trade-off between the perceived benefits of a service, such as personalized recommendations or enhanced decision support, and the perceived risks associated with disclosing personal information (Culnan & Bies, 2003). Knijnenburg et al. (2010) demonstrated that despite privacy concerns arising from increased awareness of personalization, higher system effectiveness can still increase users' willingness to engage with recommender systems, suggesting a complex privacy-utility trade-off. This phenomenon is particularly pronounced when explanations reveal inferences about personal attributes or detailed profiling that users were previously unaware of (Gedikli et al., 2014; Tintarev & Masthoff, 2015). While traditional "black box" systems collect and use data opaquely, explanation-enhanced systems explicitly

reveal *which* user characteristics, preferences, or behaviors are being analyzed. CoT chatbots explicitly articulate reasoning steps that may reference user-specific information (e.g., "Based on your preference for organic products and concern about carbon footprint..."), making data usage more transparent. For privacy-sensitive users, this visibility of data usage may reduce their fears of unintended data usage. Fig. 4 shows the conceptual model of the research.

H8. Privacy concerns will mediate the effect of interacting with the reasoning-based (vs. regular) chatbot on (a) perceived knowledge and (b) behavioral intentions.

5.2. Methodology

5.2.1. Procedure

This study employed an experimental between-subjects design in which participants were randomly assigned to one of two chatbot conditions, contrasting a CoT reasoning chatbot (1) with a standard chatbot without CoT (0). Both versions are identical except for the fact that (1) reports the CoT. In both groups, participants were invited to use a Sustainable Consumer Behavior App, which included an interactive chatbot designed to provide information and guidance on sustainable consumption. They were instructed to engage in a conversation with the chatbot, asking questions about topics such as selecting environmentally friendly products, reducing waste, energy, or emissions in daily life, understanding the environmental impact of specific habits or products, and obtaining practical tips for adopting sustainable behaviors. Participants interacted with the chatbot through a free-text chat window, allowing them to ask any questions of interest. To ensure sufficient engagement, participants were asked to pose at least three questions and maintain the conversation for a minimum of three minutes, although they were free to continue for longer if desired. This procedure ensured that participants had a meaningful and interactive experience with the chatbot, allowing the study to capture responses to the experimental manipulation effectively.

Figs. 5 and 6 illustrate the core distinction between black-box and glass-box approaches across two sustainability contexts (See Appendix B for more examples). In both figures, the black-box approach provides direct answers without exposing the underlying reasoning process, leaving the decision logic opaque to users. In contrast, the glass-box approach explicitly surfaces the analytical steps—such as problem analysis, trade-off evaluation, and justification (Fig. 6), or certification verification, carbon footprint assessment, and greenwashing detection (Fig. 5)—before presenting the final recommendation. By revealing how conclusions are derived rather than merely stating them, the glass-box design operationalizes transparent decision support and enables informed consumer decision-making.

Appendix C presents screenshots of our web-based experiment, hosted on GitHub.io, which used Google Forms to collect user data after participants interacted with the chatbots.

5.2.2. Sample

Data were collected using an online experiment distributed to participants recruited from a community of IT specialists. The final sample consisted of $N = 417$ participants (38.1% female, 61.9% male). Appendix D presents the sample profile.

5.2.3. Analytical approach

We used PLS-SEM to test the model. This method is appropriate due to (a) the predictive orientation of the research model, (b) the inclusion of multiple mediators, and (c) the presence of a dichotomous experimental variable (Hair et al., 2022; Henseler et al., 2009). In recent years, PLS has been frequently applied in information management research (e.g., Hubert et al., 2025; Koohang et al., 2022; Li et al., 2021; Liu et al., 2018; Matheus et al., 2023; Zheng et al., 2019). All analyses were conducted using SmartPLS 4 (Ringle et al., 2024) with 5000 bootstrapping

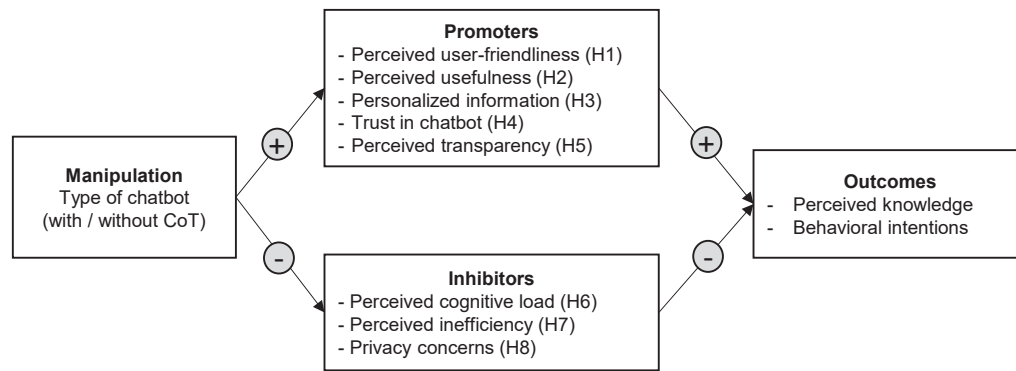
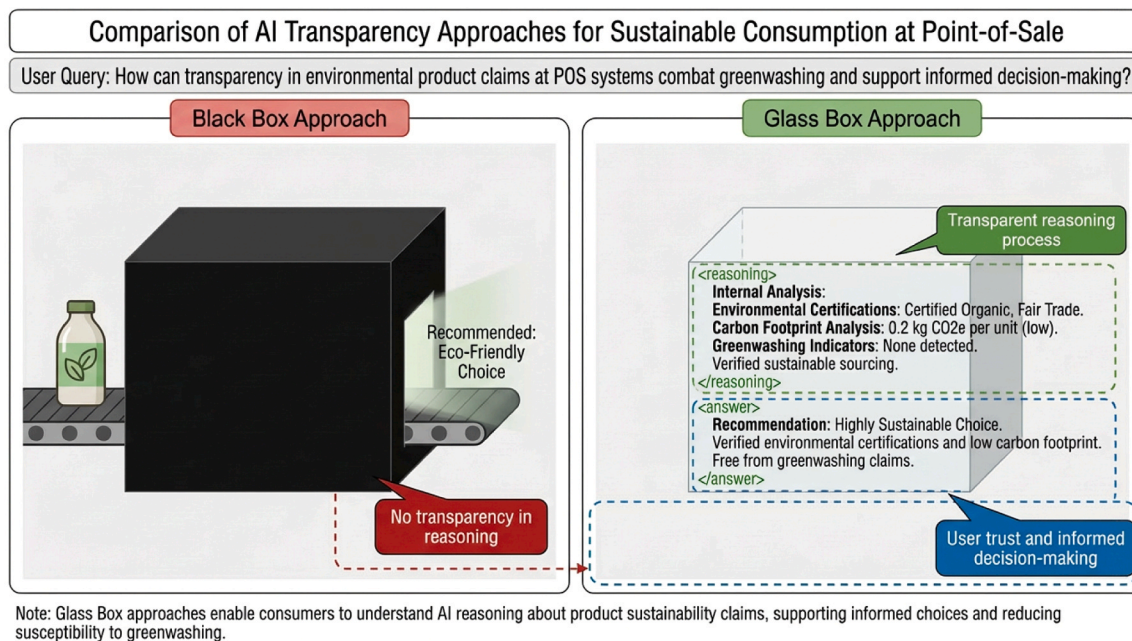


Fig. 4. Conceptual model of consumer processing. (Source: Authors' own work)



Note: Glass Box approaches enable consumers to understand AI reasoning about product sustainability claims, supporting informed choices and reducing susceptibility to greenwashing.

Fig. 5. Black Box vs. Glass Box AI approaches across sustainability domains (POS). (Source: Authors' own work)

subsamples. We report our model relying on typical guidelines, such as Hair et al., 2019; Hair et al., 2022.

In evaluating our theoretical model, we explicitly acknowledge the ongoing methodological discourse regarding the appropriateness of variance-based SEM techniques in Information Systems research (Evermann & Rönkkö, 2023). Recent debates highlight the limitations of traditional PLS-SEM in consistently estimating common factor models without bias. To ensure methodological rigor and directly address these concerns, we utilize Consistent PLS (PLSc-SEM), which applies a correction for attenuation to provide consistent estimates for reflective measurement models. As demonstrated by Henseler and Schubert (2025), PLSc serves as a robust and mathematically sound method to overcome traditional PLS limitations. Furthermore, we align with Henseler and Schubert (2025) by relying on established PLSc rather than adopting premature factor-based PLS (PLSF-SEM) techniques, ensuring that our complex, multi-mediator framework is evaluated with the highest degree of established statistical validity.

5.2.4. Measures

We used multi-items scales with reflective specification to operationalize the promoting and inhibiting mediator variables and the

dependent variables. Three items to measure the perceived user-friendliness as well as three items to measure perceived usefulness were adapted from Hoffmann et al., 2022a; Hoffmann et al., 2024b. We adapted three items to measure personalized information and three items to measure trust in the chatbot from Hoffmann et al. (2024b). We created three items for perceived transparency on bases of Mao et al. (2024). We developed three items for perceived cognitive load and three items for perceived inefficiency. Three items to measure privacy concerns were taken from Hoffmann et al. (2022b). We created scales for perceived knowledge (three items) and behavioral intentions (four items). All items were measured on seven-point rating scales. A complete list of measurement items is presented in Appendix E.

5.2.5. Measurement model

In assessing the measurement models of the PLSc-SEM analysis, we followed established reliability and validity criteria. All relevant statistics are reported in detail in Table 3. Internal consistency reliability was examined using Cronbach's alpha and composite reliability, both of which exceeded the recommended thresholds (> 0.70), indicating satisfactory reliability. Convergent validity was evaluated through the average variance extracted (AVE). While most constructs reached values

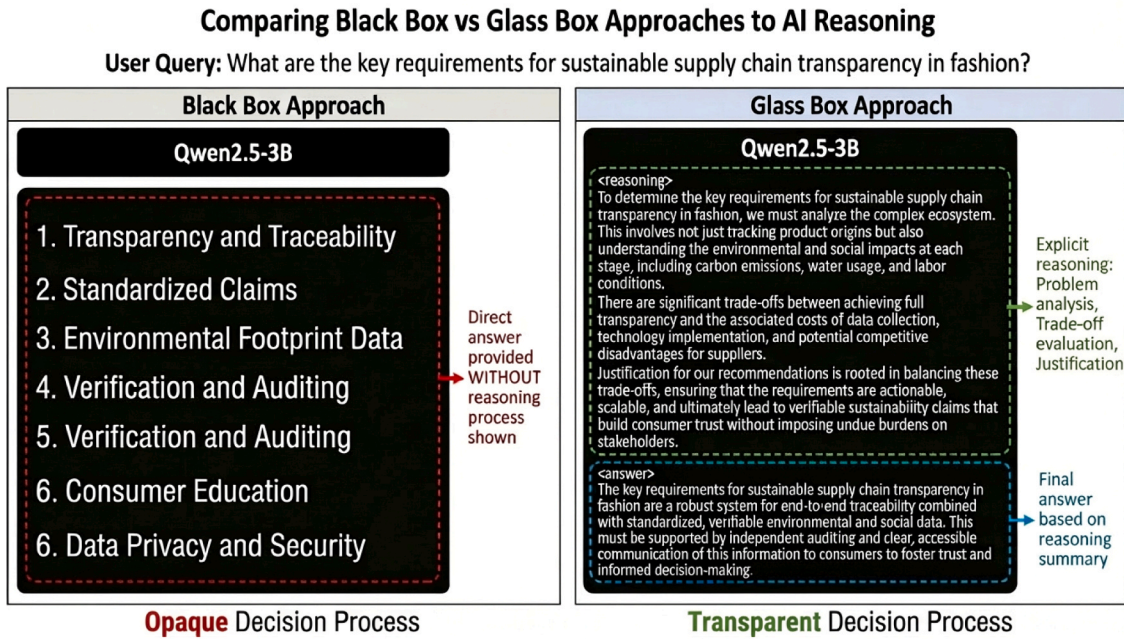


Fig. 6. Black Box vs. Glass Box AI approaches across sustainability domains (fashion). (Source: Authors' own work)

Table 3
Measurement model validation and correlation matrix.

| | M | SD | CA | CR | AVE | | KNOW | BI | UF | PU | PI | TR | TRN | CL | INEFF | PRIV |
|-------|------|------|------|------|------|---|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| KNOW | 4.76 | 1.57 | 0.77 | 0.77 | 0.52 | – | 0.72 | 0.83 | 0.41 | 0.48 | 0.49 | 0.53 | 0.44 | 0.52 | 0.54 | 0.54 |
| BI | 4.72 | 1.57 | 0.79 | 0.79 | 0.48 | | 0.83 | 0.70 | 0.41 | 0.55 | 0.55 | 0.52 | 0.45 | 0.47 | 0.50 | 0.46 |
| UF | 4.94 | 1.35 | 0.70 | 0.71 | 0.44 | | 0.41 | 0.41 | 0.67 | 0.10 | 0.09 | 0.21 | 0.20 | 0.16 | 0.13 | 0.16 |
| PU | 4.99 | 1.32 | 0.75 | 0.76 | 0.51 | | 0.48 | 0.55 | 0.10 | 0.71 | 0.31 | 0.22 | 0.15 | 0.25 | 0.21 | 0.25 |
| PI | 4.76 | 1.36 | 0.75 | 0.76 | 0.50 | | 0.49 | 0.55 | 0.08 | 0.31 | 0.71 | 0.12 | 0.23 | 0.10 | 0.27 | 0.21 |
| TR | 4.93 | 1.43 | 0.78 | 0.79 | 0.54 | | 0.53 | 0.52 | 0.21 | 0.22 | 0.13 | 0.74 | 0.14 | 0.16 | 0.33 | 0.29 |
| TRN | 4.97 | 1.27 | 0.75 | 0.76 | 0.51 | | 0.44 | 0.45 | 0.20 | 0.14 | 0.24 | 0.14 | 0.71 | 0.25 | 0.23 | 0.12 |
| CL | 3.41 | 1.42 | 0.73 | 0.73 | 0.48 | | -0.52 | -0.47 | -0.13 | -0.25 | -0.10 | -0.16 | -0.25 | 0.69 | 0.26 | 0.27 |
| INEFF | 3.43 | 1.38 | 0.70 | 0.70 | 0.44 | | -0.54 | -0.50 | -0.12 | -0.21 | -0.27 | -0.32 | -0.23 | 0.26 | 0.66 | 0.18 |
| PRIV | 3.49 | 1.39 | 0.74 | 0.75 | 0.49 | | -0.55 | -0.46 | -0.16 | -0.25 | -0.20 | -0.28 | -0.11 | 0.27 | 0.18 | 0.70 |

Note: PLS-SEM. KNOW = perceived knowledge, BI = behavioral intention, UF = user friendliness, PU = perceived usefulness, PI = personalized information, TR = trust in chatbot, TRN = perceived transparency, CL = perceived cognitive load, INEFF = perceived inefficiency, PRIV = privacy concerns, SD = Standard deviation, CA = Cronbach's alpha, CR = Composite reliability (rho_a), AVE = Average variance extracted (boldfaced diagonal = square root; lower part = standardized construct correlations; upper part = heterotrait-monotrait-ratio (HTMT))

above the commonly accepted cutoff (> 0.50), the AVE for user friendliness (0.44) and inefficiency (0.44) fell marginally below this threshold. However, recent methodological advancements emphasize that the 0.50 cutoff should not be interpreted as a strict, isolated decision rule (Cheung et al., 2024). Aligning with the foundational work of Fornell and Larcker (1981, p. 46)—who explicitly state that AVE is a conservative measure and that convergent validity can be deemed adequate based on composite reliability alone—we retained these constructs to preserve the theoretical integrity of our model. Discriminant validity was assessed using both the Fornell and Larcker (1981) criterion and the heterotrait–monotrait ratio of correlations (HTMT; Henseler et al., 2015). The Fornell–Larcker criterion assesses discriminant validity by testing whether a construct shares more variance with its own indicators than with other constructs. The HTMT criterion assesses discriminant validity by comparing correlations between constructs with correlations within constructs; low HTMT values indicate that the constructs are empirically distinct. The square roots of the AVE values were higher than the inter-construct correlations, fulfilling the Fornell–Larcker requirement, and all HTMT values remained below the recommended thresholds (< 0.90), confirming discriminant validity.

Finally, for full transparency and to allow for methodological comparison, the traditional PLS-SEM results are provided in the [supplementary material](#) (see [Supplementary Material](#)).

5.3. Results

Initially, we ran *t*-tests to analyze the effect of the type of chatbot on the dependent variables. We found strong statistically significant differences (Fig. 7). Participants who used the CoT-chatbot reported higher perceived knowledge ($M_{reg} = 3.676$, $M_{CoT} = 6.049$, $t(415) = 23.445$, $p < 0.001$) and behavioral intentions ($M_{reg} = 3.607$, $M_{CoT} = 6.061$, $t(415) = 25.469$, $p < 0.001$) that those using a regular chatbot.

Next, we estimated the structural model in SmartPLS 4 using bootstrapping with 5000 resamples. All variance inflation factors (VIFs) were below 2, indicating no multicollinearity issues in the model (Hair et al., 2019). The structural model explained substantial portions of variance, with perceived knowledge showing an R^2 of 0.832 ($R^2_{adj} = 0.829$) and behavioral intention an R^2 of 0.817 ($R^2_{adj} = 0.813$). According to Cohen (1988), R^2 values around 0.02 indicate small effects, values around 0.15 indicate medium effects, and values above 0.35 indicate large effects,

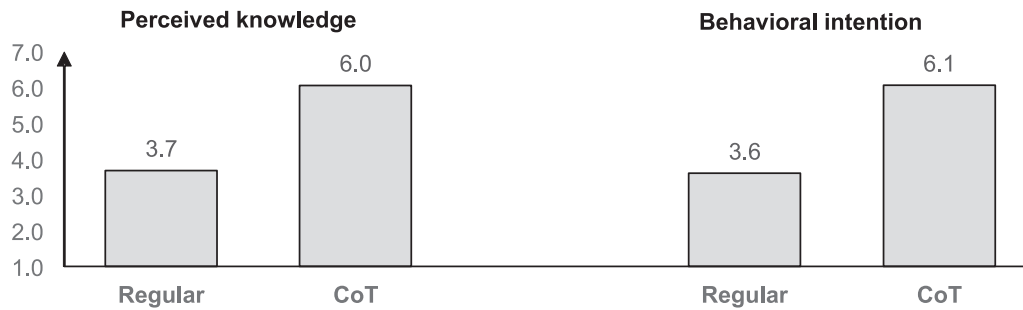


Fig. 7. Means of perceived knowledge and behavioral intention depending on type to chatbot.

suggesting that the model demonstrates large explanatory effects for both endogenous constructs.

The results of the PLS path modeling are summarized in Table 4. In the upper section of the table, the effects of the experimental manipulation of the chatbot on the mediator variables are presented. Here, all five hypothesized promoters show positive path coefficients, while the three hypothesized inhibitors exhibit negative coefficients. All effects are statistically significant and of substantial magnitude. The middle

section of Table 4 reports the influences of the mediators on the two dependent variables, perceived knowledge and behavioral intentions. The sign of each path coefficient aligns with the promoter or inhibitor classification, and all effects are again statistically significant.

Finally, the lower section of Table 4 presents the results of the mediation analyses. All indirect effects are significant, confirming that the proposed mediation mechanisms hold. These results provide full support for all of the hypothesized relationships.

Table 4
Structural model and hypotheses testing.

| | β (original) | β (mean) | SD | t | p | CI-2.5 | CI-97.5 | Hypotheses |
|--|-----------------------|-------------------|-------|--------|---------|--------|---------|---------------|
| Chatbot -> Mediator | | | | | | | | |
| Chatbot -> UF | 0.786 | 0.791 | 0.100 | 7.837 | < 0.001 | 0.588 | 0.987 | |
| Chatbot -> PU | 0.906 | 0.908 | 0.086 | 10.547 | < 0.001 | 0.734 | 1.071 | |
| Chatbot -> PI | 0.875 | 0.879 | 0.092 | 9.527 | < 0.001 | 0.697 | 1.055 | |
| Chatbot -> TR | 1.105 | 1.104 | 0.078 | 14.197 | < 0.001 | 0.945 | 1.253 | |
| Chatbot -> TRN | 0.832 | 0.835 | 0.090 | 9.234 | < 0.001 | 0.651 | 1.008 | |
| Chatbot -> CL | -0.753 | -0.753 | 0.101 | 7.465 | < 0.001 | -0.948 | -0.551 | |
| Chatbot -> INEFF | -0.945 | -0.949 | 0.098 | 9.634 | < 0.001 | -1.139 | -0.753 | |
| Chatbot -> PRIV | -0.882 | -0.884 | 0.099 | 8.918 | < 0.001 | -1.069 | -0.686 | |
| Mediator -> DV | | | | | | | | |
| UF -> KNOW | 0.205 | 0.206 | 0.053 | 3.877 | < 0.001 | 0.104 | 0.309 | |
| UF -> BI | 0.208 | 0.210 | 0.052 | 4.012 | < 0.001 | 0.107 | 0.310 | |
| PU -> KNOW | 0.154 | 0.155 | 0.053 | 2.922 | 0.003 | 0.048 | 0.257 | |
| PU -> BI | 0.245 | 0.245 | 0.052 | 4.754 | < 0.001 | 0.143 | 0.345 | |
| PI -> KNOW | 0.231 | 0.231 | 0.052 | 4.445 | < 0.001 | 0.128 | 0.330 | |
| PI -> BI | 0.306 | 0.308 | 0.054 | 5.647 | < 0.001 | 0.204 | 0.417 | |
| TR -> KNOW | 0.231 | 0.230 | 0.051 | 4.485 | < 0.001 | 0.129 | 0.328 | |
| TR -> BI | 0.238 | 0.239 | 0.050 | 4.795 | < 0.001 | 0.139 | 0.334 | |
| TRN -> KNOW | 0.155 | 0.154 | 0.053 | 2.921 | 0.004 | 0.048 | 0.257 | |
| TRN -> BI | 0.163 | 0.163 | 0.052 | 3.116 | 0.002 | 0.058 | 0.264 | |
| CL -> KNOW | -0.239 | -0.238 | 0.053 | 4.483 | < 0.001 | -0.341 | -0.132 | |
| CL -> BI | -0.198 | -0.198 | 0.052 | 3.781 | < 0.001 | -0.300 | -0.094 | |
| INEFF -> KNOW | -0.205 | -0.206 | 0.054 | 3.805 | < 0.001 | -0.310 | -0.099 | |
| INEFF -> BI | -0.145 | -0.145 | 0.054 | 2.675 | 0.007 | -0.249 | -0.035 | |
| PRIV -> KNOW | -0.244 | -0.247 | 0.051 | 4.812 | < 0.001 | -0.347 | -0.149 | |
| PRIV -> BI | -0.142 | -0.142 | 0.051 | 2.764 | 0.006 | -0.242 | -0.041 | |
| Mediation (Total indirect effect) | | | | | | | | |
| Chatbot -> UF -> KNOW | 0.161 | 0.164 | 0.051 | 3.167 | 0.002 | 0.075 | 0.274 | H1a supported |
| Chatbot -> UF -> BI | 0.163 | 0.168 | 0.051 | 3.202 | 0.001 | 0.076 | 0.276 | H1b supported |
| Chatbot -> PU -> KNOW | 0.140 | 0.141 | 0.050 | 2.778 | 0.005 | 0.043 | 0.241 | H2a supported |
| Chatbot -> PU -> BI | 0.222 | 0.223 | 0.057 | 3.932 | < 0.001 | 0.119 | 0.339 | H2b supported |
| Chatbot -> PI -> KNOW | 0.202 | 0.204 | 0.056 | 3.622 | < 0.001 | 0.103 | 0.319 | H3a supported |
| Chatbot -> PI -> BI | 0.268 | 0.272 | 0.063 | 4.228 | < 0.001 | 0.159 | 0.408 | H3b supported |
| Chatbot -> TR -> KNOW | 0.255 | 0.254 | 0.063 | 4.063 | < 0.001 | 0.137 | 0.382 | H4a supported |
| Chatbot -> TR -> BI | 0.264 | 0.265 | 0.062 | 4.224 | < 0.001 | 0.147 | 0.390 | H4b supported |
| Chatbot -> TRN -> KNOW | 0.129 | 0.130 | 0.049 | 2.613 | < 0.001 | 0.038 | 0.232 | H5a supported |
| Chatbot -> TRN -> BI | 0.136 | 0.138 | 0.050 | 2.708 | 0.007 | 0.046 | 0.241 | H5b supported |
| Chatbot -> CL -> KNOW | 0.180 | 0.180 | 0.051 | 3.560 | < 0.001 | 0.091 | 0.287 | H6a supported |
| Chatbot -> CL -> BI | 0.149 | 0.151 | 0.049 | 3.050 | 0.002 | 0.063 | 0.254 | H6b supported |
| Chatbot -> INEFF -> KNOW | 0.194 | 0.196 | 0.056 | 3.467 | < 0.001 | 0.089 | 0.309 | H7a supported |
| Chatbot -> INEFF -> BI | 0.137 | 0.138 | 0.055 | 2.494 | 0.013 | 0.033 | 0.249 | H7b supported |
| Chatbot -> PRIV -> KNOW | 0.215 | 0.220 | 0.059 | 3.681 | < 0.001 | 0.118 | 0.342 | H8a supported |
| Chatbot -> PRIV -> BI | 0.125 | 0.127 | 0.051 | 2.431 | 0.015 | 0.034 | 0.234 | H8b supported |

Note: PLS-SEM. KNOW =perceived knowledge, BI = behavioral intention, UF = user friendliness, PU = perceived usefulness, PI = personalized information, TR = trust in chatbot, TRN = perceived transparency, CL = perceived cognitive load, INEFF = perceived inefficiency, PRIV = privacy concerns, β (original) = path coefficient PLS modelling, β (sample M) = mean path coefficient via bootstrapping with 5000 samples, SD = standard deviation. CI = confidence interval 2.5%-97.5%.

In sum, the results strongly support the hypothesized model, showing that the experimental manipulation of the chatbot effectively influenced the mediators as expected. The significant mediation effects confirm the proposed process mechanisms. Promoters and inhibitors had significant and directionally consistent effects, which in turn significantly impacted perceived knowledge and behavioral intentions.

While these results provide robust quantitative evidence for the effectiveness of the manipulation and the underlying process mechanisms, they do not fully illuminate how users subjectively experienced the interaction with the two chatbot versions. To deepen our understanding of these experiential aspects and to contextualize the statistical findings, we conducted a qualitative follow-up study.

6. Study 3 (Phase 3): Qualitative study

6.1. Methodology

6.1.1. Objective

The primary objective of this qualitative study is to gain a deeper understanding of participants' experiences when interacting with chatbots that differ in their reasoning capabilities. Specifically, the study seeks to explore how the presence of CoT reasoning in a chatbot affects users' perceptions of usefulness, clarity, and overall interaction quality. While the quantitative study assessed perceived knowledge and behavioral intentions as outcomes through structured measures, this qualitative investigation focuses on capturing the nuanced, subjective experiences of participants.

6.1.2. Procedure

This study employed the same experimental procedure as the quantitative study. Participants were randomly assigned to one of two chatbot conditions, contrasting a CoT reasoning chatbot with a standard chatbot without CoT. To ensure comparable interaction depth, participants were instructed to engage in a conversation with the chatbot, ask at least three questions, and continue the interaction for a minimum of three minutes. They were free to continue for longer if desired. After completing the interaction, they answered an open-ended question prompting them to describe their overall experience with the chatbot, including perceived usefulness, clarity, and any issues encountered.

6.1.3. Sample

Data were collected via the online panel provider Prolific. In total, 128 participants interacted with the chatbot for more than three minutes and provided narrative feedback. Of the participants, 64 (50.0%) identified as female, 61 (47.7%) as male, 1 (0.8%) as diverse, and 2 (1.6%) preferred not to disclose their gender. The mean age was 27.6 years ($SD = 7.3$), ranging from 18 to 55 years. In total, 70 participants (54.7%) were assigned to the regular chatbot, whereas 58 (45.3%) used the CoT chatbot.

6.1.4. Data processing

We conducted an inductive qualitative content analysis following established principles of iterative coding development (Krippendorff, 2018; Neuendorf, 2017). First, we thoroughly reviewed the material to identify recurring themes. Initial categories were refined through repeated comparisons between statements and preliminary codes, resulting in six final categories. Three categories reflected positive reactions, including positive experiences, useful knowledge, and helpful reasoning, while three reflected negative reactions, including unspecific answers, long answers, and technical issues.

To ensure coding transparency and replicability, we formulated precise category definitions and documented representative anchor examples, which are provided in Appendix F. The full dataset was then systematically coded based on this scheme.

6.2. Results

We first examined the frequency of categories across the full sample. Because participants were not prompted to comment on specific features of the chatbot, the observed frequencies likely represent conservative estimates of the underlying perceptions. Overall, 52 participants (40.8%) reported positive experiences, 18 (14.1%) indicated that they gained useful knowledge, and 11 (8.6%) explicitly emphasized the helpfulness of the chatbot's reasoning process. On the negative side, 7 participants (5.5%) noted unspecific answers, 24 (18.8%) found the responses too long or confusing, and 16 (12.5%) reported technical or design issues, such as incomplete answers or a rudimentary interface.

Next, we examined whether category frequencies differed between chatbot conditions. The descriptive results are displayed in Fig. 8. Although the descriptive patterns suggest more favorable evaluations of the CoT chatbot, the Chi² tests revealed no statistically significant differences for general positive evaluations ($\chi^2(1) = 2.574$, $p = 0.109$) or useful knowledge ($\chi^2(1) = 2.110$, $p = 0.146$). However, there was a strong and statistically significant effect regarding the evaluation of the reasoning process ($\chi^2(1) = 14.524$, $p < 0.001$). Moreover, the CoT chatbot significantly reduced negative evaluations, including unspecific answers ($\chi^2(1) = 6.136$, $p = 0.013$), long or confusing answers ($\chi^2(1) = 16.300$, $p < 0.001$), and technical or design issues ($\chi^2(1) = 5.206$, $p = 0.023$).

In sum, when comparing the two chatbot designs qualitatively, the CoT chatbot received more favorable evaluations regarding its reasoning structure and demonstrated significantly fewer negative reactions across. These findings suggest that participants value the CoT-based reasoning and this may enhance perceived clarity, transparency, and answer quality while simultaneously reducing frustrations.

7. Discussion

This research integrates three complementary studies to explore the development, evaluation, and user experience of CoT enhanced language models in the context of sustainable consumption. In doing so, the study introduces an innovative research protocol that systematically links model optimization, experimental validation, and qualitative user insights. This approach bridges the gap between technical AI development and user-center behavioral research.

Study 1 demonstrated that fine-tuning the Qwen-2.5-3B-Instruct model using GRPO significantly improved the model's ability to generate structured, transparent reasoning chains. The technical metrics—specifically the stabilization of reward scores and controlled KL divergence—confirm that the model successfully adapted to the domain-specific reasoning tasks without catastrophic forgetting. This provides important evidence that parameter-efficient fine-tuning approaches can enhance reasoning capabilities without compromising model stability, which is critical for real-world deployment in resource-constrained environments.

Study 2 provided causal evidence that interacting with this CoT-enabled chatbot leads to higher perceived knowledge and stronger behavioral intentions compared to a standard chatbot. The PLS-SEM analysis supported all hypothesized mediating mechanisms, revealing that CoT promotes effects of trust, transparency, usefulness, and personalization, while simultaneously reducing cognitive load, perceptions of inefficiency, and privacy concerns. These findings demonstrate that CoT does not merely improve perceptions of system performance but fundamentally alters how users process, evaluate, and act upon information.

Finally, Study 3 contextualized these findings, showing that users qualitatively perceive the CoT reasoning as more helpful and less prone to generating unspecific or confusing answers. This qualitative evidence further reveals that structured reasoning helps users in navigating complex, multi-criteria decisions.

Synthesizing these findings, we propose the following formal

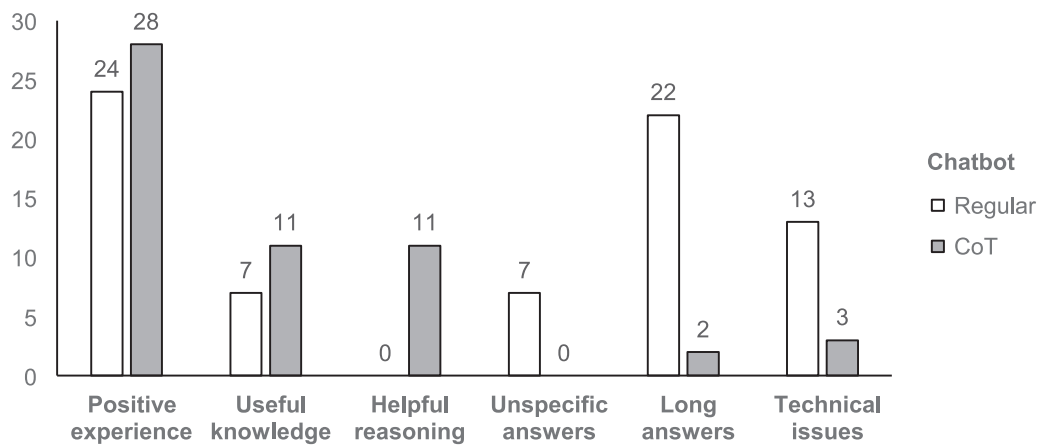


Fig. 8. Frequency of different evaluative criteria depending on type to chatbot.

propositions regarding the role of CoT in sustainable consumption behavior. These propositions conceptualize CoT not merely as a technical feature, but as a behavioral intervention mechanism that shapes how users interpret and act on AI-generated information.

Proposition 1. The technical enhancement of LLMs through GRPO-based fine-tuning translates directly into superior behavioral outcomes by bridging the gap between model capability and user comprehension. Human-interpretable explanations help generate downstream behavioral effects.

Proposition 2. Transparency in AI reasoning functions as a primary driver of sustainable behavior adoption, where the cognitive cost of processing detailed explanations is offset by the gains in trust and perceived utility. Users accept increased informational effort when it yields higher epistemic value.

Proposition 3. The efficacy of generative AI in behavioral change is contingent upon a dual mechanism where technical reasoning quality (backend) and perceived user-friendliness (frontend) must be optimized simultaneously.

7.1. Theoretical contributions and implications

This study makes several significant contributions to the literature on explainable AI, information systems, and sustainable consumer behavior.

First, this study advances the theoretical boundaries of human-computer interaction by resolving the intellectual puzzle introduced by CoT reasoning. Traditionally, foundational frameworks such as the TAM and the IS Success Model have operated on a linear assumption: increased information quality and transparency directly equate to higher perceived usefulness and ease of use (Davis, 1989; DeLone & McLean, 2003). Our findings challenge this deeply rooted assumption by proving that in generative AI, transparency is paradoxical. Building on prior research on information overload, explainability–usability trade-offs and the personalization privacy calculus (Awad et al., 2006; Dinev & Hart, 2006; Hoffmann et al., 2022b), we introduce a critical nuance to XAI literature by demonstrating that CoT reasoning activates competing psychological mechanisms—acting simultaneously as a promoter (enhancing trust and perceived knowledge) and a potential inhibitor (elevating cognitive load and privacy concerns). Crucially, our findings indicate that this paradox is not inherent to transparency per se, but depends on how transparency is structured and presented.

Second, by positioning our findings within this scholarly conversation, we argue that future theoretical models of AI adoption must evolve beyond descriptive, universally positive views of transparency. Instead, adoption models must account for this duality. More importantly, our

study points to boundary conditions under which transparency yields positive versus negative outcomes. Critically, our findings reveal that in high-involvement, normatively complex domains like sustainability, this paradox is resolved: when decisions involve uncertain trade-offs, CoT reasoning serves as a cognitive scaffold rather than a burden. It simultaneously adds information volume yet reduces interpretive complexity. This clarifies an ongoing debate in XAI literature (Chromik et al., 2021; Schneider & Handali, 2019), proving that the net effect of AI transparency is positive only when the reasoning is explicitly structured to guide users through domain-relevant trade-offs.

Third, our study bridges a critical gap between system explainability and user understanding (Doshi-Velez & Kim, 2017; Hoffman et al., 2018; Meske et al., 2022). Much XAI research evaluates explanations using computational metrics or expert assessments, yet few studies examine whether explanations improve end-user knowledge and behavioral outcomes (Poursabzi-Sangdeh et al., 2021). By integrating behavioral experimentation with model development, this study extends prior work by demonstrating that explainability can serve as a mechanism for learning, persuasion, and decision support. We demonstrate empirically that CoT reasoning enhances not only perceived transparency but also users' knowledge acquisition and behavioral intentions—showing that XAI benefits extend beyond trust calibration to include substantive cognitive gains. This positions explainable AI as an active driver of behavioral outcomes rather than merely a tool for model inspection.

Fourth, we contribute to consumer behavior research by establishing a causal link between AI-mediated reasoning and sustainable behavioral intentions. Our findings suggest that conversational agents employing structured reasoning can bridge the attitude-behavior gap (Carrington et al., 2010) by enhancing perceived knowledge through explicit trade-off articulation, which reduces decision uncertainty and strengthens behavioral intentions. In this sense, CoT reasoning functions as a form of decision justification, enabling consumers to rationalize and commit to sustainability-oriented choices (Kaiser & Brüggemann, 2025). This extends the SHIFT framework (White et al., 2019) by identifying AI-enabled reasoning transparency as a tool for supporting behavior change in complex decision contexts.

In summary, this study's core theoretical contribution lies in explicating the psychological mechanisms through which reasoning transparency operates in complex decision contexts and demonstrating that these mechanisms differ systematically based on domain characteristics—advancing both explainable AI research and information systems scholarship. By demonstrating how structured AI reasoning simultaneously influences cognitive processing, trust formation, and behavioral intentions, this research advances a new perspective on AI systems as behavioral intervention technologies in complex decision environments.

7.2. Implications for practice

The findings offer actionable insights for developers, policymakers, and sustainability advocates. These implications highlight that the value of CoT lies not only in improving model performance but in shaping user cognition and decision-making processes.

The novelty and originality of these findings translate directly into practical benefits for AI system developers and UI/UX designers' community. Based on our data analysis, which confirms that CoT significantly increases perceived knowledge and behavioral intentions, organizations deploying AI for complex domains should transition away from opaque 'black-box' recommendations. The primary benefit of our study is providing empirical validation that designing conversational agents with explicit, step-by-step reasoning structures (the 'glass-box' approach) actively mitigates user uncertainty and drives sustainable behavioral adoption.

For AI Developers and Engineers: The study validates the use of parameter-efficient fine-tuning techniques (GRPO with LoRA) to enhance reasoning capabilities in smaller models. Developers should prioritize the generation of high-quality, domain-specific synthetic data (as done in Study 1) to train models that can "show their work." This is particularly relevant for deploying efficient AI solutions on edge devices where massive models are impractical.

For Policymakers and NGOs: The strong positive effect of CoT on behavioral intentions suggests that AI chatbots can be powerful tools for public education. Organizations aiming to promote sustainable lifestyles should deploy conversational agents that are transparent about their sources and reasoning logic. Rather than providing static tips, these tools should be designed to explain the environmental impact of specific choices, thereby fostering greater user trust and knowledge retention. At the same time, the persuasive power of structured reasoning raises important ethical considerations. Policymakers should therefore ensure that such systems adhere to principles of responsible AI, including transparency, accountability, and non-manipulative design, as transparency does not inherently guarantee neutrality (Papagiannidis et al., 2025).

For UX/UI Designers: The dual presence of promoters and inhibitors highlights a critical design challenge. While users value transparency, the potential for cognitive load and privacy concerns must be managed. Designers should implement "progressive disclosure" interfaces where the CoT reasoning is available but not overwhelming—perhaps offered as an expandable option. Interfaces could be developed in such a way that they allow users to personalize the level of explanation detail, enabling adaptive explainability based on user preferences and expertise. Furthermore, addressing privacy concerns through clear data usage policies and "local-only" processing options could further enhance adoption. More broadly, explainability should be treated as an interactive design feature rather than a static system output.

8. Conclusions, limitations, and future research

8.1. Conclusions

This research successfully demonstrates that fine-tuning language models for reasoning is a viable and effective strategy for promoting sustainable consumption behavior. By combining rigorous technical optimization (Study 1) with robust behavioral testing (Study 2) and qualitative validation (Study 3), we show that CoT-enabled chatbots do more than just answer questions—they educate and persuade. The transition from a "black box" to a "glass box" model fosters trust, enhances perceived usefulness, and ultimately drives users toward more sustainable choices. The study provides a holistic blueprint for leveraging advanced AI to address critical societal challenges.

In summary, this study's core theoretical contribution lies in explicating the psychological mechanisms through which reasoning transparency operates in complex decision contexts and demonstrating that

these mechanisms differ systematically based on domain characteristics—advancing both explainable AI research and information systems scholarship.

8.2. Limitations and future research direction

Despite its contributions, this study has limitations that open avenues for future research. First, the sample for the behavioral studies (Study 2 and 3) consisted primarily of IT specialists. While this group is knowledgeable about technology, they may possess higher need for cognition and lower technology anxiety than the general population. Future research should replicate the experiment with a broader, more representative sample of consumers to validate the generalizability of the findings.

Second, the study focused on a specific model architecture (Qwen-2.5-3B) and a specific domain (sustainable consumption). The findings are therefore likely to generalize most directly to comparable decision-support contexts characterized by uncertainty, value-laden trade-offs, and multi-criteria decision-making, where users may benefit from transparent explanations and structured reasoning. Future studies could investigate whether the benefits of GRPO fine-tuning and CoT reasoning extend to other critical domains, such as healthcare (Bocking et al., 2022; Goel et al., 2024) or financial planning (Koskelainen et al., 2023). Given that sustainability decisions are inherently moral and high-involvement, future research should also examine whether the observed benefits of CoT are generalizable to low-involvement, routine, time-sensitive, or hedonic consumption contexts, where detailed reasoning may be less relevant or even perceived as inefficient (Longoni & Cian, 2022). In addition, future research should examine key moderating variables, such as AI literacy, domain involvement, decision complexity, and cultural context, to better understand for whom and under which conditions CoT is most effective (Husairi & Rossi, 2024; Mogaji & Jain, 2024; Schweidel et al., 2022). From a more methodological point of view, future studies could also test whether they hold across different model sizes, architectures, chatbot interfaces, and deployment settings.

Third, the behavioral intentions measured in Study 2 are self-reported. While intention is a strong predictor of behavior, it is not a perfect proxy. Future research should employ longitudinal designs or field experiments to track actual behavioral changes over time, such as verifying if users actually purchase eco-friendly products or reduce their energy consumption following the chatbot interaction. Such designs have been proven as fruitful to test influences of digital tools on sustainable consumption behavior (e.g., longitudinal studies: Lasarov et al., 2024, field experiments: Hoffmann et al., 2022b) and they should be applied with regard to XAI as well.

Fourth, future research should also employ longitudinal study designs to test how the cognitive friction and privacy concerns associated with 'glass-box' transparency evolve over time. In particular, future studies should investigate how repeated exposure to CoT reasoning affects user behavior over time, including potential habituation effects, automation bias, or diminishing marginal returns. It is plausible that as users build habituation and familiarity with explicit CoT reasoning structures, the initial cognitive load may decrease, further altering the delicate balance between the promoters and inhibitors of AI adoption.

Finally, future research should explore the potential "dark side" of CoT reasoning. This includes psychological vulnerabilities such as user overreliance on AI-generated rationales, the illusion of understanding, and the critical risk of persuasive but fundamentally flawed explanations. To mitigate these adverse effects and better align model outputs with human cognitive processing, we propose the exploration of Group Relative Sequence Optimization (GSPO) as a conceptual and technical advancement over current GRPO approaches. While GRPO traditionally optimizes reasoning based on relative outcome rewards within a group of generated responses, GSPO would explicitly optimize the structural logic, coherence, and safety of the reasoning *sequence* itself. By shifting

the algorithmic focus to sequential integrity, GSPO could significantly reduce the generation of misleading rationales, ensuring that "glass-box" AI models remain both computationally robust and psychologically safe for users navigating complex decision environments.

CRedit authorship contribution statement

Pejman Ebrahimi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stefan Hoffmann:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Johannes Schneider:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Ethics approval and consent to participate

This study involved human participants and was conducted in accordance with applicable ethical standards. Participants were provided with an online information and consent form prior to participation and gave informed consent by actively indicating agreement (e.g., selecting "I have read and understood the information above, and I voluntarily consent to participate"). The consent form described the study purpose, procedures, risks (no greater than those encountered in daily life), confidentiality guidance (participants were instructed not to provide identifying information in open-text fields), data collected (survey responses and evaluation choices), intended use of data

Appendix A

(statistical analyses and academic publication), and compensation terms (including potential disqualification without compensation if attention checks were failed).

Consent for publication

The authors consent to the publication of this work in the "International Journal of Information Management".

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT 5 in order to improve the language and readability of the manuscript, as the authors are non-native English speakers. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

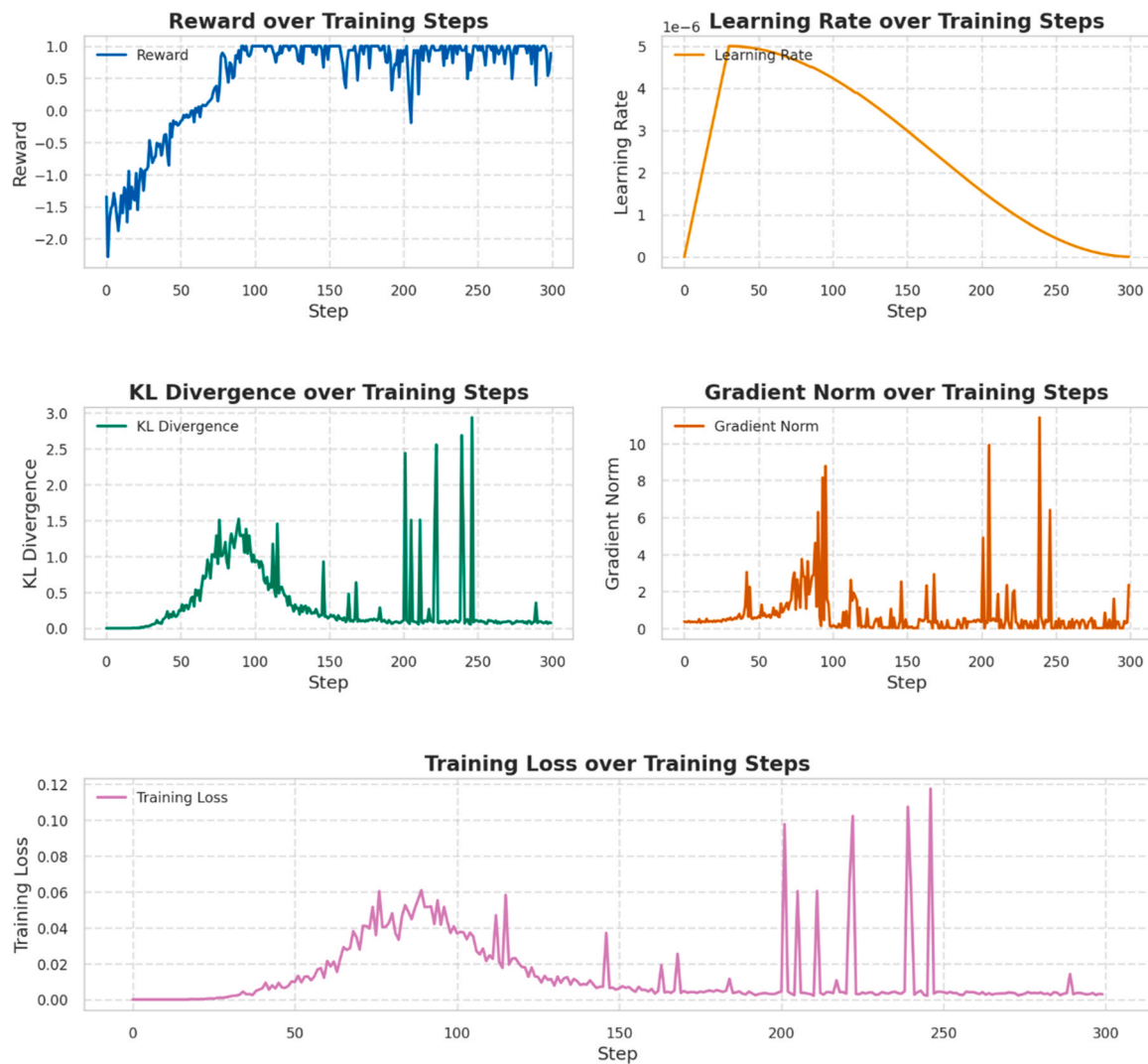


Figure A1. Training dynamics of GRPO fine-tuning for sustainable consumption reasoning model (Source: Authors' own work)

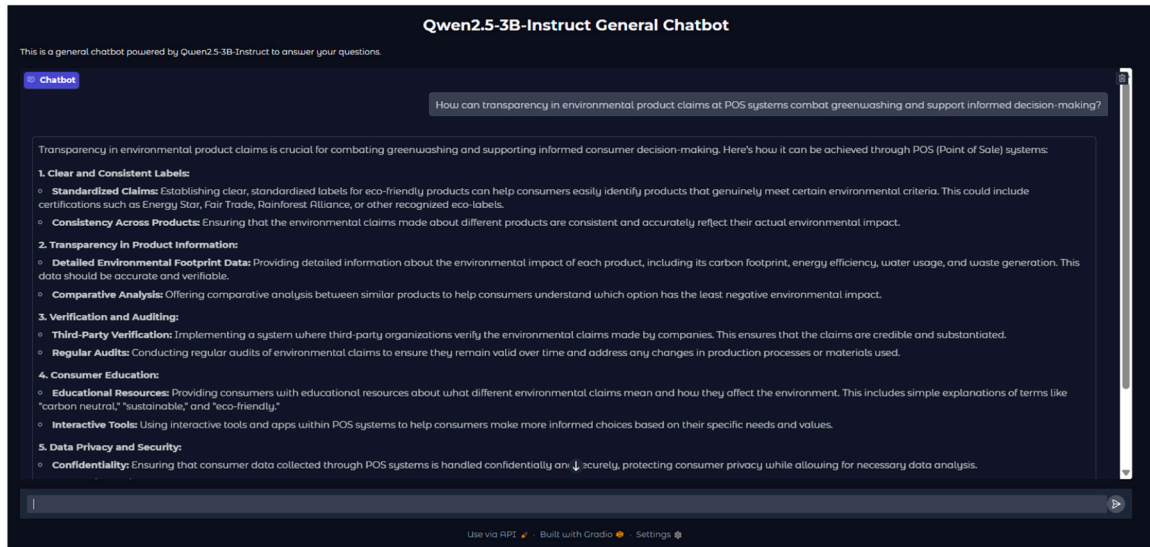
Table A1

Progressive performance metrics across GRPO fine-tuning phases for sustainable consumption reasoning enhancement

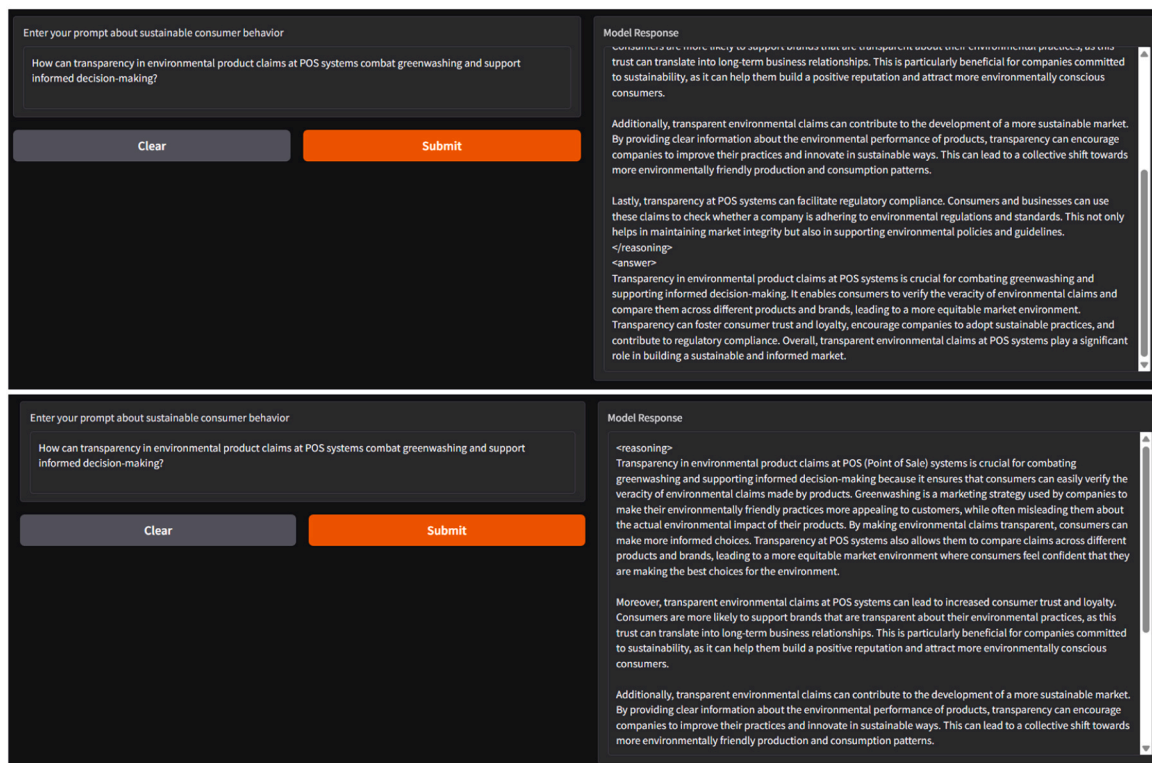
| Performance Metric | Early Phase(Steps 1–50) | Mid Phase(Steps 51–150) | Final Phase(Steps 151–300) |
|------------------------------------|-------------------------|-------------------------|----------------------------|
| Reward Score (Mean ± SD) | -1.35 ± 0.26 | -0.11 ± 0.08 | 0.94 ± 0.18 |
| Training Loss | 0.0001 ± 0.0002 | 0.012 ± 0.004 | 0.0036 ± 0.0008 |
| KL Divergence | 0.00002 ± 0.00001 | 0.32 ± 0.15 | 0.088 ± 0.022 |
| Completion Length (tokens) | 290.1 ± 65.3 | 107.4 ± 18.5 | 207.8 ± 28.4 |
| Strict Format Reward | 0.000 | 0.125 ± 0.177 | 0.456 ± 0.058 |
| Reward Standard Deviation | 0.264 ± 0.089 | 0.105 ± 0.048 | 0.312 ± 0.285 |
| Gradient Norm | ~0.0001 | 0.011 ± 0.004 | 0.0038 ± 0.0009 |
| Learning Rate (×10 ⁻⁶) | 5.0 | 3.2 ± 1.1 | 0.8 ± 0.6 |

Note: Training employed GRPO with LoRA (rank=64, alpha=64) over 300 steps, modifying 119.7 M of 3.09B parameters (3.99%); Configuration: AdamW-8bit optimizer (lr=5 × 10⁻⁶, β₁=0.9, β₂=0.99), cosine scheduling, batch size = 8, 8 generations per prompt.

Appendix B. Comparative response analysis of regular and Fine-tuned Models on sustainable consumption query

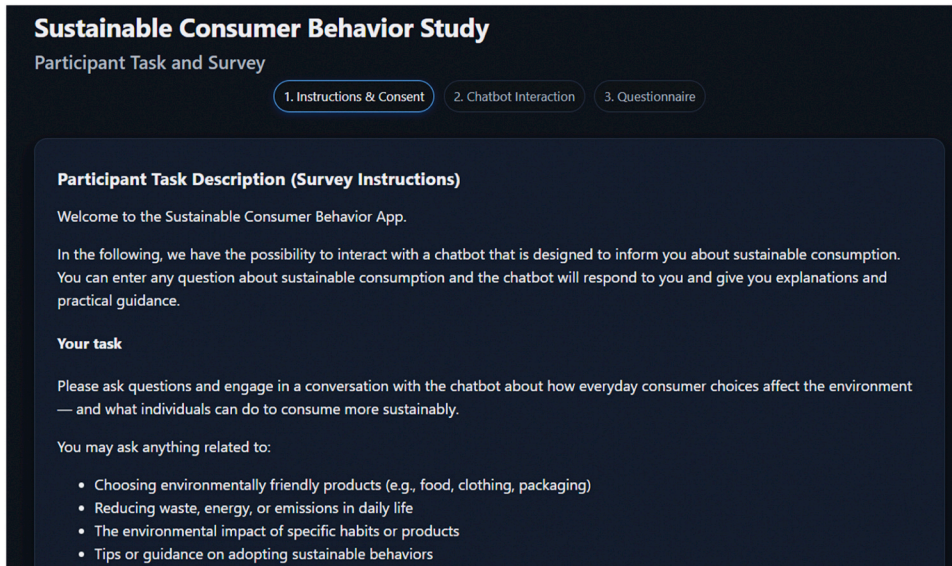


(a). Regular (Base) Model (Qwen2.5-3B-Instruct) response with structured format (Source: Authors' own work)

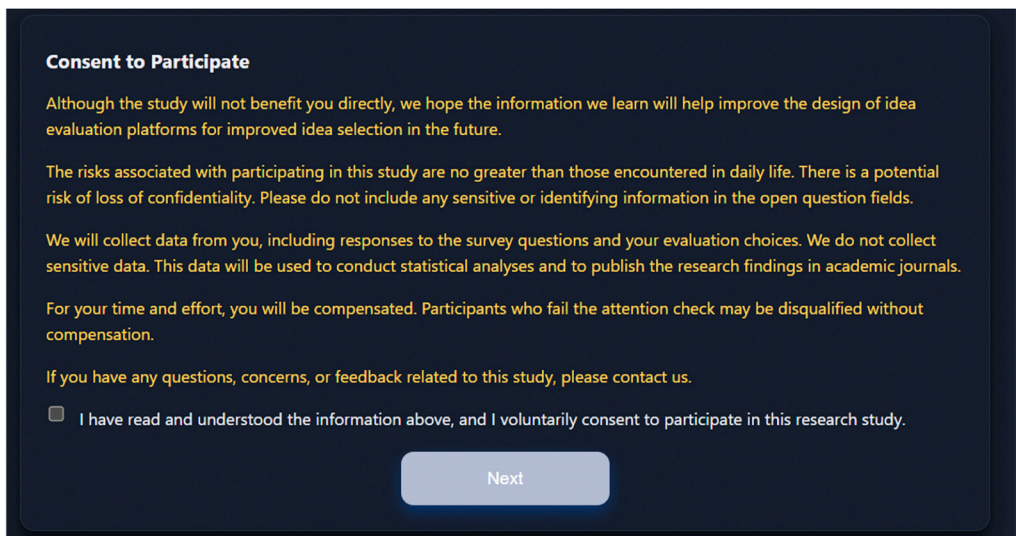


(b). Fine-tuned model response with reasoning chain and summarized answer (Source: Authors' own work)

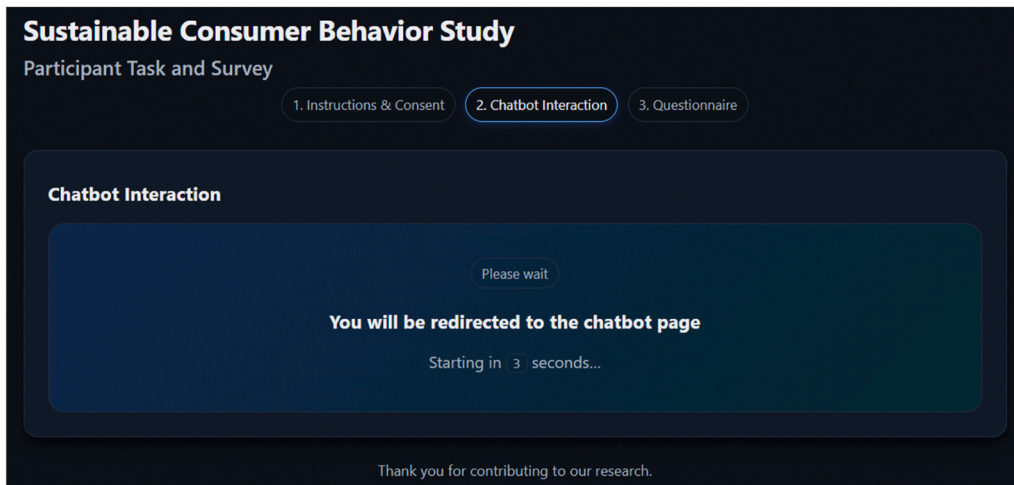
Appendix C. Web experiment (Screenshots) (Source: Authors' own work)



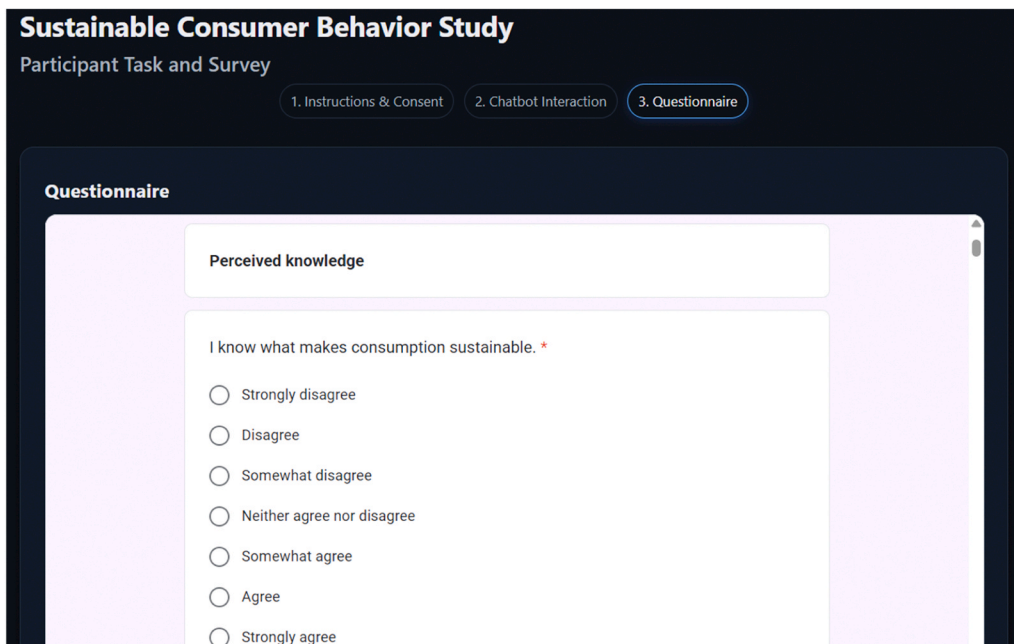
(a). Home page (based on GitHub.io)



(b). Consent form



(c). Interaction page



(d). Survey page

Appendix D. Sample characteristics (N = 417)

| | Frequency | Percent |
|--------------------------------------|-----------|---------|
| Gender | | |
| Female | 159 | 38.1 |
| Male | 258 | 61.9 |
| Age | | |
| 18–26 years | 170 | 40.8 |
| 27–36 years | 122 | 29.3 |
| 37–46 years | 74 | 17.7 |
| 47–56 years | 47 | 11.2 |
| > 56 years | 4 | 1.0 |
| Education | | |
| TAFE (certificate, diploma or below) | 122 | 29.3 |
| Bachelor | 137 | 32.9 |

(continued on next page)

(continued)

| | Frequency | Percent |
|-----------------|-----------|---------|
| Master | 117 | 28.0 |
| PhD / Doctorate | 41 | 9.8 |

Appendix E. Measurement

| Items | Loading |
|--|---------|
| Perceived knowledge | |
| I know what makes consumption sustainable. | 0.755 |
| I am aware of the impact my consumption choices have on the environment and society. | 0.687 |
| I know of practical ways to make my consumption more sustainable. | 0.726 |
| Behavioral intention | |
| I intend to make sustainable consumption decisions. | 0.673 |
| I plan to consume in a resource-conscious way. | 0.693 |
| I am willing to reduce my environmental footprint through my consumption behavior. | 0.694 |
| When I purchase something, I will consider how sustainable it is. | 0.721 |
| Perceived user friendliness | |
| The chatbot's interface is intuitive. | 0.747 |
| I find the chatbot to be easy to use. | 0.567 |
| The chatbot is easy to control. | 0.672 |
| Perceived usefulness | |
| Using the chatbot improves my understanding of sustainable consumption. | 0.743 |
| The chatbot provides me with useful information that I wouldn't be able to find on my own. | 0.624 |
| The chatbot offers practical guidance that is novel for me. | 0.758 |
| Personalized information | |
| The chatbot's personalized recommendations match my preferences and needs. | 0.734 |
| I feel that chatbots understands my preferences and needs better than other sources of information. | 0.763 |
| I value the personalized service that the chatbots provides and find it helpful when making consumption decisions. | 0.618 |
| Trust in chatbot | |
| I trust that the chatbot will provide me with accurate and reliable information. | 0.679 |
| I have trust in the chatbot's responses. | 0.811 |
| I believe that the chatbot is a trustworthy source of information and advice. | 0.712 |
| Perceived transparency | |
| The chatbot clearly explained how it arrived at its answers. | 0.706 |
| I understood the reasoning behind the chatbot's responses. | 0.731 |
| The chatbot's way of thinking was transparent to me. | 0.699 |
| Perceived cognitive overload | |
| The chatbot's explanation was so long that it made it hard for me to find the key information. | 0.674 |
| I felt mentally overloaded by the chatbot's lengthy response. | 0.682 |
| I lost track of the important points because the chatbot provided too much detail. | 0.713 |
| Perceived inefficiency | |
| The chatbot wasted my time by giving such a long explanation. | 0.642 |
| The response felt inefficient because it included too much information. | 0.661 |
| I could have found the answer faster without reading the entire explanation. | 0.687 |
| Privacy concerns | |
| If I were to use the app, I would be concerned that personal information ... | |
| ... submitted could be misused. | 0.778 |
| ... could be made available to unknown people or companies without my knowledge. | 0.634 |
| ... could be inappropriately used. | 0.671 |

Appendix F. Coding Scheme

| Category | Description | Example 1 | Example 2 | Example 3 |
|----------------------------|--|--|---|---|
| Positive experience | User expresses satisfaction, enjoyment, and generally positive evaluation. | Great experience and i like to use it in future. | Using the chatbot created a new experience for me that was enjoyable. | Really good chatbot. |
| Useful knowledge | User perceived the information as novel, helpful or usable. | The answers were reliable and useful to find better solutions. | The chatbot was very useful to find novel ways to improve my practical actions to make better decisions about my environmental effects. | Using this Chatbot help me to improve my knowledge about sustainability in my life. |
| Helpful reasoning | User comments positively on the reasoning process and structure of the answers. | I like that the <reasoning> tag feels like a teacher explaining things step-by-step. | It's fun to see how the chatbot reasons things out before giving the result | Even when the reasoning is long, the <answer> is short and straight to the point. |
| Unspecific answers | User perceived the answers as too generic, not specific and missing examples. | Kind of generic answers. | Didn't give examples like I expected. | Didn't really get what I was asking. |
| Long answers | User feels that responses are too long, repetitive, or confusing. | The chatbot's responses were very long. | I like that but i prefer to see answers shorter and highlight the main points. | A little confusing. |
| Technical issues | User expresses issues about slow response speed, incomplete answers, or design of the interface. | Design of user interface can be more better. | The chatbot is so slow, I don't know why but it took lot of time to answer my questions. | Sometimes the responses were incomplete. |

Appendix G. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ijinfomgt.2026.103075](https://doi.org/10.1016/j.ijinfomgt.2026.103075).

Data Availability

Data will be made available on request.

[COT-Sustainable-Consumption-Behavior](#) (Hugging Face Data repository)

References

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. <https://doi.org/10.1109/TKDE.2005.99>
- Alghamdi, A. D. (2025). MultiAgent-CoT: A multi-agent chain-of-thought reasoning model for robust multimodal dialogue understanding. *Computers, Materials & Continua*, 86(2), 1–35. <https://doi.org/10.32604/cmc.2025.071210>
- Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review*, 36(2), 247–271. <https://doi.org/10.5465/amr.2009.0188>
- Amrollahi, A., Yang, J., Fazal-e-Hasan, S. M., & Badreddine, B. (2026). Knowledge workers' trust and reception of generative AI's advice in complex tasks (Article) *International Journal of Information Management*, 88, Article 103031. <https://doi.org/10.1016/j.ijinfomgt.2026.103031>.
- Attari, S. Z., DeKay, M. L., Davidson, C. L., & de Bruin, W. B. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 16054–16059. <https://doi.org/10.1073/pnas.1001509107>
- Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, 30(1), 13–28. <https://doi.org/10.2307/25148715>
- Bansal, G., Nushi, B., Kamar, E., Horvitz, E., & Weld, D. S. (2021). Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(13), 11405–11414. <https://doi.org/10.1609/aaai.v35i13.17359>
- 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), 1431–1454. <https://doi.org/10.1287/isre.2023.1199>
- Bi, X., Yang, M., & Adomavicius, G. (2024). Consumer acquisition for recommender systems: A theoretical framework and empirical evaluations. *Information Systems Research*, 35(1), 339–362. <https://doi.org/10.1287/isre.2023.1229>
- Bocking, H., Russell-Bennett, R., & Letheren, K. (2022). I'm no expert, but...? Consumer use of supportive digital tools in health services. *Journal of Service Theory and Practice*, 32(2), 105–131. <https://doi.org/10.1108/JSTP-09-2020-0225>
- van Bussel, L. M., Kuijsten, A., Mars, M., & Van't Veer, P. (2022). Consumers' perceptions on food-related sustainability: A systematic review (Article) *Journal of Cleaner Production*, 341, Article 130904. <https://doi.org/10.1016/j.jclepro.2022.130904>.
- Carrington, M. J., Neville, B. A., & Whitwell, G. J. (2010). Why ethical consumers don't walk their talk: Towards a framework for understanding the gap between the ethical purchase intentions and actual buying behaviour of ethically minded consumers. *Journal of Business Ethics*, 97(1), 139–158. <https://doi.org/10.1007/s10551-010-0501-6>
- Chen, Y., Deng, H., Han, K., & Zhao, Q. (2025). Policy frameworks for transparent chain-of-thought reasoning in large language models. arXiv arXiv:2503.14521.
- Chen, Q. Q., Lin, L. M., & Yi, Y. (2025). Tailoring explanations in conversational recommendations: The impact of decision contexts and user interfaces. *Journal of Retailing and Consumer Services*, 85, Article 104281. <https://doi.org/10.1016/j.jretconser.2025.104281>
- Chen, C., Tian, A. D., & Jiang, R. (2024). When post hoc explanation knocks: Consumer responses to explainable AI recommendations. *Journal of Interactive Marketing*, 59(3), 234–250. <https://doi.org/10.1177/10949968231200221>
- Chromik, M., Eiband, M., Buchner, F., Krüger, A., & Butz, A. (2021). I think I get your point, AI! The illusion of explanatory depth in explainable AI. *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)* (pp. 307–317). Association for Computing Machinery. <https://doi.org/10.1145/3397481.3450644>
- Culnan, M. J., & Bies, R. J. (2003). Consumer privacy: Balancing economic and justice considerations. *Journal of Social Issues*, 59(2), 323–342. <https://doi.org/10.1111/1540-4560.00067>
- Chen, L., & Pu, P. (2008, October 23). A cross-cultural user evaluation of product recommender interfaces. In Proceedings of the 2008 ACM conference on Recommender Systems (RecSys '08) (pp. 75–82). Association for Computing Machinery. <https://doi.org/10.1145/1454008.1454022>.
- Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., Wang, L. C., & Chan, K. Y. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pac. J. Manag.*, 41, 745–783. <https://doi.org/10.1007/s10490-023-09871-y>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, 27(1), 113–118. <https://doi.org/10.1017/S1351324920000601>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>
- DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1), 60–95. <https://doi.org/10.1287/isre.3.1.60>
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30. <https://doi.org/10.1080/07421222.2003.11045748>
- Diekmann, A., & Preisendörfer, P. (2003). Green and greenback: The behavioral effects of environmental attitudes in low-cost and high-cost situations. *Rationality and Society*, 15(4), 441–472. <https://doi.org/10.1177/1043463103154002>
- Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), 61–80. <https://doi.org/10.1287/isre.1060.0080>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv arXiv:1702.08608. <https://doi.org/10.48550/arXiv.1702.08608>.
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- Ebrahimi, P., & Schneider, J. (2025). Fine-tuning image-to-text models on Liechtenstein tourist attractions. *Electronic Markets*, 35. <https://doi.org/10.1007/s12525-025-00806-7>
- ElHaffar, G., Durif, F., & Dubé, L. (2020). Towards closing the attitude-intention-behavior gap in green consumption: A narrative review of the literature and an overview of future research directions (Article) *Journal of Cleaner Production*, 275, Article 122556. <https://doi.org/10.1016/j.jclepro.2020.122556>.
- Evermann, J., & Rönkkö, M. (2023). Recent developments in PLS. *Communications of the Association for Information Systems*, 52, 663–667. <https://doi.org/10.17705/1CAIS.05229>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.2307/3151312>
- Foroughi, B., Iranmanesh, M., Yadegaridehkordi, E., Wen, J., Ghobakhloo, M., Senali, M. G., & Annamalain, N. (2025). Factors affecting the use of ChatGPT for obtaining shopping information. *International Journal of Consumer Studies*, 49, Article e70008. <https://doi.org/10.1111/ijcs.70008>
- Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4), 367–382. <https://doi.org/10.1016/j.ijhcs.2013.12.007>
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), 51–90. <https://doi.org/10.2307/30036519>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Goel, A., Singh, A., Taneja, U., & Jain, S. (2024). Consumer adoption of digital health services: A systematic literature review and research agenda. *International Journal of Consumer Studies*, 48(4), Article e13070. <https://doi.org/10.1111/ijcs.13070>
- Gorton, M., Tocco, B., Yeh, C. H., & Hartmann, M. (2021). What determines consumers' use of eco-labels? Taking a close look at label trust. *Ecological Economics*, 189, Article 107173. <https://doi.org/10.1016/j.ecolecon.2021.107173>
- Grinstein, A., Kodra, E., Chen, S., Sheldon, S., & Zik, O. (2018). Carbon inclusivity. *PLoS ONE*, 13(5). <https://doi.org/10.1371/journal.pone.0196282>
- Guo, D., Yang, D., Zhang, H., et al. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv arXiv:2501.12948. <https://doi.org/10.48550/arXiv.2501.12948>.
- Hair, J., Hult, G. T. M., Ringle, C., Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3rd ed., Sage.
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>

- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *New challenges to international marketing* (pp. 277–319). Emerald Group Publishing Limited.
- Henseler, J., & Schubert, F. (2025). Should PLS become factor-based or should CB-SEM become composite-based? Both! *European Journal of Information Systems*, 34(3), 551–563. <https://doi.org/10.1080/0960085X.2024.2357123>
- Hibbeln, M., Jenkins, J. L., Schneider, C., Valacich, J. S., & Weinmann, M. (2017). How is your user feeling? Inferring emotion through human–computer interaction devices. *MIS Quarterly*, 41(1), 1–21. <https://doi.org/10.25300/MISQ/2017/41.1.01>
- Hoffman, R.R., Mueller, S.T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. arXiv arXiv:1812.04608. (<https://arxiv.org/abs/1812.04608>).
- Hoffmann, S., Joerß, T., Mai, R., & Akbar, P. (2022b). Augmented reality-delivered product information at the point of sale: when information controllability backfires. *Journal of the Academy of Marketing Science*, 50(4), 743–776. <https://doi.org/10.1007/s11747-022-00855-w>
- Hoffmann, S., Lasarov, W., & Dwivedi, Y. K. (2024b). AI-empowered scale development: Testing the potential of ChatGPT (Article) *Technological Forecasting and Social Change*, 205, Article 123488. <https://doi.org/10.1016/j.techfore.2024.123488>.
- Hoffmann, S., Lasarov, W., & Reimers, H. (2022a). Carbon footprint tracking apps (Article) *What drives consumers' adoption intention? Technology in Society*, 69, Article 101956. <https://doi.org/10.1016/j.techsoc.2022.101956>.
- Hoffmann, S., Lasarov, W., Reimers, H., & Trabant, M. (2024a). Carbon footprint tracking apps. *Does feedback help reduce carbon emissions?* (Article) *Journal of Cleaner Production*, 434, Article 139981. <https://doi.org/10.1016/j.jclepro.2023.139981>.
- Hubert, M., Kazemi, S., Hubert, M., Carugati, A., & Mariani, M. M. (2025). The exploration of users' perceived value from personalization and virtual conversational agents to enable a smart home assemblage—a mixed method approach (Article) *International Journal of Information Management*, 80, Article 102850. <https://doi.org/10.1016/j.ijinfomgt.2024.102850>.
- Husairi, M. A., & Rossi, P. (2024). Delegation of purchasing tasks to AI: The role of perceived choice and decision autonomy. *Decision Support Systems*, 179, Article 114166. <https://doi.org/10.1016/j.dss.2023.114166>
- IPCC. (2023). *Climate change 2023: Synthesis report*. Intergovernmental Panel on Climate Change.
- Jannach, D. (2023). Evaluating conversational recommender systems: A landscape of research. *Artificial Intelligence Review*, 56(3), 2365–2400. <https://doi.org/10.48550/arXiv.2208.12061>
- Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2021). A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)*, 54(5), 1–36. <https://doi.org/10.1145/3453154>
- Jiang, Z., Chen, G., & Xu, Z. (2026). Building regulation question-answering system using retrieval-augmented generation with dual-stage fine-tuned large language model. *Advanced Engineering Informatics*, 69, Article 104089. <https://doi.org/10.1016/j.aei.2025.104089>
- Joerß, T., Hoffmann, S., Mai, R., & Akbar, P. (2021). Digitalization as solution to environmental problems? When users rely on augmented reality-recommendation agents. *Journal of Business Research*, 128, 510–523. <https://doi.org/10.1016/j.jbusres.2021.02.019>
- Joshi, R., Graefe, J., Kraus, M., & Bengler, K. (2024). Exploring the impact of explainability on trust and acceptance of conversational agents: A Wizard of Oz study. *Artificial intelligence in HCI: 5th International Conference, AI-HCI 2024, held as part of the 26th HCI International Conference, HCII 2024, Washington, DC, USA, June 29–July 4, 2024, proceedings, Part I* (pp. 199–218). Springer. https://doi.org/10.1007/978-3-031-60606-9_12
- Kaiser, F. G., & Brüggemann, M. (2025). Uncovering the relevance of reasons for behavior: The attitude–behavior gap revisited. *Journal of Environmental Psychology*, 107, Article 102762. <https://doi.org/10.1016/j.jenvp.2025.102762>
- Kang, S. E., Kim, M. J., Kim, J. S., & Olya, H. (2025). Can I Trust GenAI to Plan My Next Trip? A Multi-Method Approach to Optimizing Media Mix. *Journal of Travel Research*, 65(2). <https://doi.org/10.1177/00472875241305630>
- Kapoor, A., & Kumar, M. (2025). Frontiers: Generative AI and personalized video advertisements. *Marketing Science*, 44(4). <https://doi.org/10.1287/mksc.2023.0494>
- Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Vaughan, J. W. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376219>
- Kim, J. H., Kim, J., Baek, T. H., & Kim, C. (2025). ChatGPT personalized and humorous recommendations (Article) *Annals of Tourism Research*, 110, Article 103857. <https://doi.org/10.1016/j.annals.2024.103857>.
- Kim, H., & Lee, S. W. (2026). Sorry, it's my fault: Politeness, attribution, and anthropomorphism in managing generative AI hallucinations (Article) *International Journal of Information Management*, 86, Article 102996. <https://doi.org/10.1016/j.ijinfomgt.2025.102996>.
- Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390–2395). Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858402>
- Knijnenburg, B. P., Bostandjiev, S., O'Donovan, J., & Kobsa, A. (2013). Inspectability and control in social recommenders. *Proceedings of the sixth ACM conference on Recommender systems (RecSys '12)* (pp. 43–50). Association for Computing Machinery. <https://doi.org/10.1145/2365952.2365966>
- Knijnenburg, B. P., Willemsen, M. C., & Hirtbach, S. (2010). Receiving recommendations and providing feedback: The user experience of a recommender system. In In F. Buccafurri, & G. Semeraro (Eds.), *E-commerce and web technologies: EC-Web 2010* (Lecture Notes in Business Information Processing, 61 pp. 173–186). Springer. https://doi.org/10.1007/978-3-642-15208-5_19.
- Kocaballi, A. B., Berkovsky, S., Quiroz, J. C., Laranjo, L., Tong, H. L., Rezazadegan, D., ... Coiera, E. (2019). The personalization of conversational agents in health care: systematic review. *Journal of Medical Internet Research*, 21(11). (<https://preprints.jmir.org/preprint/15360>).
- Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *Advances in Neural Information Processing Systems (NeurIPS)*, 35, 22199–22213. <https://doi.org/10.5555/3600270.3601883>
- Koohang, A., Sargent, C. S., Nord, J. H., & Paliszkievicz, J. (2022). Internet of Things (IoT): From awareness to continued use (Article) *International Journal of Information Management*, 62, Article 102442. <https://doi.org/10.1016/j.ijinfomgt.2021.102442>
- Koskelainen, T., Kalmi, P., Scornavacca, E., & Vartiainen, T. (2023). Financial literacy in the digital age—A research agenda. *Journal of Consumer Affairs*, 57(1), 507–528. <https://doi.org/10.1111/joca.12510>
- Kovács, E.-Z., & Horváth, I. (2025). Trust, transparency, and AI adoption in business. *Acta Polytechnica Hungarica*, 22(6), 81–100. <https://doi.org/10.12700/APH.22.6.2025.6.6>
- Kovari, A. (2025). Explainable AI chatbots towards XAI ChatGPT: A review (Article) *Heliyon*, 11(2), Article e42077. <https://doi.org/10.1016/j.heliyon.2025.e42077>.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Sage.
- Lasarov, W., Hoffmann, S., & Dwivedi, Y. K. (2026). With a little help from my artificial friend: Functional and parasocial value in ChatGPT use (Article) *Technology in Society*, 85, Article 103220. <https://doi.org/10.1016/j.techsoc.2026.103220>.
- Lasarov, W., Hoffmann, S., Mai, R., & Schleich, J. (2024). Carbon footprint tracking apps: The spillover effects of feedback and goal-activating appeals (Article) *Ecological Economics*, 226, Article 108354. <https://doi.org/10.1016/j.ecolecon.2024.108354>.
- Laukkanen, T., Xi, N., Hallikainen, H., Ruusunen, N., & Hamari, J. (2022). Virtual technologies in supporting sustainable consumption: From a single-sensory stimulus to a multi-sensory experience. *International Journal of Information Management*, 63, Article 102455. <https://doi.org/10.1016/j.ijinfomgt.2021.102455>
- Li, Y., Mou, J., Ye, L., Long, J., & Huang, W. W. (2021). An empirical investigation of the utilitarian, social benefits in LBS information disclosure—The moderating effect of the gender based social role theory (Article) *International Journal of Information Management*, 56, Article 102243. <https://doi.org/10.1016/j.ijinfomgt.2020.102243>.
- Liao, Q. V., & Sundar, S. S. (2022). Designing for responsible trust in AI systems: A communication perspective. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)* (pp. 1257–1268). Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533182>
- Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., ... & Leike, J. (2023). Let's verify step by step. arXiv arXiv:2305.20050. (<https://arxiv.org/abs/2305.20050>).
- Liu, S., Chan, F. T., Yang, J., & Niu, B. (2018). Understanding the effect of cloud computing on organizational agility: An empirical examination. *International Journal of Information Management*, 43, 98–111. <https://doi.org/10.1016/j.ijinfomgt.2018.07.010>
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35. <https://doi.org/10.1145/3560815>
- Longo, L., Breci, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., ... Stumpf, S. (2024). Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions (Article) *Information Fusion*, 106, Article 102301. <https://doi.org/10.1016/j.inffus.2024.102301>.
- Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical AI. *Journal of Consumer Research*, 46(4), 629–650. <https://doi.org/10.1093/jcr/ucz013>
- Longoni, C., & Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The “word-of-Machine” Effect *Journal of Marketing*, 86(1), 91–108. <https://doi.org/10.1177/0022242920957347>
- Ma, J., Wang, P., Li, B., Wang, T., Pang, X. S., & Wang, D. (2025). Exploring user adoption of ChatGPT: A technology acceptance model perspective. *International Journal of Human-Computer Interaction*, 41(2), 1431–1445. <https://doi.org/10.1080/10447318.2024.2314358>
- Majid, G. M., Tussyadiah, I., Kim, Y. R., & Chen, J. L. (2025). Promoting pro-environmental behaviour spillover through chatbots. *Journal of Sustainable Tourism*, 33(11), 2440–2458. <https://doi.org/10.1080/09669582.2024.2393256>
- Mao, Y., Rafner, J., Wang, Y., & Sherson, J. (2024). HI-TAM, a hybrid intelligence framework for training and adoption of generative design assistants. *Frontiers in Computer Science*, 6. <https://doi.org/10.3389/fcomp.2024.1460381>
- Matheus, R., Faber, R., Ismagilova, E., & Janssen, M. (2023). Digital transparency and the usefulness for open government (Article) *International Journal of Information Management*, 73, Article 102690. <https://doi.org/10.1016/j.ijinfomgt.2023.102690>.
- Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: objectives, stakeholders, and future research opportunities. *Information Systems Management*, 39(1), 53–63.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Mogaji, E., & Jain, V. (2024). How generative AI is (will) change consumer behaviour: Postulating the potential impact and implications for research, practice, and policy. *Journal of Consumer Behaviour*, 23(5), 2379–2389. <https://doi.org/10.1002/cb.2345>
- Munro, P., Kapitan, S., & Wooliscroft, B. (2023). The sustainable attitude-behavior gap dynamic when shopping at the supermarket: A systematic literature review and framework for future research (Article) *Journal of Cleaner Production*, 426, Article 138740. <https://doi.org/10.1016/j.jclepro.2023.138740>.
- Najafian, S., Delic, A., Tkalcic, M., & Tintarev, N. (2021). Factors influencing privacy concern for explanations of group recommendation (pp. 134–139). In Adjunct

- Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization. <https://doi.org/10.1145/3450613.3456845>
- Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). Sage.
- Norman, D. A. (2013). *The design of everyday things (Revised and expanded ed.)*. MIT Press.
- Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, 34(2), Article 101885. <https://doi.org/10.1016/j.jsis.2024.101885>
- Pizzi, G., Scarpi, D., & Pantano, E. (2021). Artificial intelligence and the new forms of interaction: Who has the control when interacting with a chatbot? *Journal of Business Research*, 129, 878–890. <https://doi.org/10.1016/j.jbusres.2020.11.006>
- Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W., & Wallach, H. (2021). Manipulating and measuring model interpretability. *Proceedings of the 2021 CHI conference on human factors in computing systems (Article 237, pp. 1–52)*. Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445315>
- Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experimental perspective. *Journal of Marketing*, 85(1), 131–151. <https://doi.org/10.1177/0022242920953847>
- PraisonAI. (n.d.). PrasionAI [Computer software]. GitHub. <https://github.com/MervinPraison/PraisonAI>
- Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141. <https://doi.org/10.1007/s11747-019-00710-5>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). ACM. <https://doi.org/10.1145/2939672.2939778>
- Ries, T., & Rese, A. (2025). Can information help bridge the attitude–behavior gap in sustainable clothing consumption? (Article) *Cleaner and Responsible Consumption*, 17, Article 100278. <https://doi.org/10.1016/j.clrc.2025.100278>
- Ringle, C.M., Wende, S., & Becker, J.-M. (2024). SmartPLS 4. Bönningstedt: SmartPLS. Retrieved from (<https://www.smartpls.com>).
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Satinet, C., Ducarroz, C., & Fouss, F. (2025). Understanding the impact of sustainability-oriented recommender systems on consumers’ choices. *Electronic Commerce Research and Applications*, 74, Article 101540. <https://doi.org/10.1016/j.elerap.2025.101540>
- Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, Article 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>
- Schneider, J. (2024). Explainable generative AI (GenXAI): A survey, conceptualization, and research agenda. *Artificial Intelligence Review*, 57. <https://doi.org/10.1007/s10462-024-10916-x>
- Schneider, J. (2025). Generative to agentic ai: Survey, conceptualization, and challenges. arXiv preprint arXiv:2504.18875.
- Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2022). *Recommender systems handbook* (3rd ed.). Springer.
- Schneider, J., & Handali, J. (2019). Personalized explanation in machine learning: A conceptualization. arXiv arXiv:1901.00770. (<https://arxiv.org/abs/1901.00770>).
- Schweidel, D. A., Bart, Y., Inman, J. J., Stephen, A. T., Libai, B., Andrews, M., ... Thomaz, F. (2022). How consumer digital signals are reshaping the customer journey. *Journal of the Academy of Marketing Science*, 50(6), 1257–1276. <https://doi.org/10.1007/s11747-022-00839-w>
- Shao, Z., Wang, P., Zhu, Q., et al. (2024). Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv arXiv:2402.03300. <https://doi.org/10.48550/arXiv.2402.03300>
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, Article 102551. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277–284. <https://doi.org/10.1016/j.chb.2019.04.019>
- Silalahi, A. D. K. (2025). Can generative artificial intelligence drive sustainable behavior? A consumer-adoption model for AI-driven sustainability recommendations. *Technology in Society*, 83, Article 102995. <https://doi.org/10.1016/j.techsoc.2025.102995>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Tam, K. Y., & Ho, S. Y. (2006). Understanding the impact of web personalization on user information processing and decision outcomes. *MIS Quarterly*, 30(4), 865–890. <https://doi.org/10.2307/25148757>
- Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (2nd ed., pp. 353–382). Springer. https://doi.org/10.1007/978-1-4899-7637-6_10
- Trudel, R. (2019). Sustainable consumer behavior. *Consumer Psychology Review*, 2(1), 85–96. <https://doi.org/10.1002/arc.1045>
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. United Nations General Assembly.
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837. <https://doi.org/10.5555/3600270.3602070>
- White, K., Habib, R., & Hardisty, D. J. (2019). How to SHIFT consumer behaviors to be more sustainable: A literature review and guiding framework. *Journal of Marketing*, 83(3), 22–49. <https://doi.org/10.1177/0022242919825649>
- Wu, T., Li, R., Deng, Z., Zhou, L., & Lu, L. (2026). Seeking help from AI: Understanding patient use of intelligent guidance applications (Article) *International Journal of Information Management*, 88, Article 103032. <https://doi.org/10.1016/j.ijinfomgt.2026.103032>
- Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., ... & Qiu, Z. (2025). Qwen3 technical report. arXiv arXiv:2505.09388. <https://doi.org/10.48550/arXiv.2505.09388>
- Yang, B., Sun, Y., & Li, Q. (2026). To be credible or to be creative? Understanding the mechanisms driving user satisfaction with AI-generated content from a dual fit perspective (Article) *International Journal of Information Management*, 86, Article 102980. <https://doi.org/10.1016/j.ijinfomgt.2025.102980>
- Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2024). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36, 11809–11822. (<https://arxiv.org/abs/2305.10601>).
- von Zahn, M., Bauer, K., Mihale-Wilson, C., Jagow, J., Speicher, M., & Hinz, O. (2024). Smart green nudging: Reducing product returns through digital footprints and causal machine learning. *Marketing Science*, 44(4). <https://doi.org/10.1287/mksc.2022.0393>
- Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1), 1–101. <https://doi.org/10.1561/15000000066>
- Zhao, Y., Wang, S., Geng, Q., Yu, E., & Li, J. (2025). "Less is More": Reducing Cognitive Load and Task Drift in Real-Time Multimodal Assistive Agents for the Visually Impaired. arXiv arXiv:2511.00945.
- Zheng, X., Men, J., Yang, F., & Gong, X. (2019). Understanding impulse buying in mobile commerce: An investigation into hedonic and utilitarian browsing. *International Journal of Information Management*, 48, 151–160. <https://doi.org/10.1016/j.ijinfomgt.2019.02.010>



Pejman Ebrahimi is a postdoctoral researcher in Data Science and Artificial Intelligence at the University of Liechtenstein and a senior data scientist at GilTech Megoldások in Budapest. His research focuses on agentic AI, machine learning, secure enterprise AI, and game-based educational systems, and he actively supervises master's and PhD research projects. He holds a PhD in Economic and Regional Sciences from MATE University, Hungary, and regularly serves as a peer reviewer for international journals. In 2025, he was selected to participate in the 8th Lindau Nobel Laureate Meeting in Economic Sciences. He has been listed among Hungary's top scientists in the AD Scientific Index from 2020 to 2024. His applied work bridging academic innovation and industry-grade AI solutions.



Stefan Hoffmann is professor of marketing at Kiel University, Germany. He received his doctorate and completed his habilitation at the Technical University of Dresden, Germany. His research interests include digital marketing, responsible innovation, and sustainable consumption behavior. He has published in journals such as *Journal of the Academy of Marketing Science*, *International Journal of Research in Marketing*, *Journal of Consumer Psychology*, *Psychology and Marketing*, and *Tourism Management*.



Johannes Schneider is Full Professor of Data Science and Artificial Intelligence at the University of Liechtenstein. He previously conducted research at IBM and ABB industrial research laboratories. He holds a PhD and Master's degree in Computer Science, as well as a Master of Advanced Studies in Management, Technology, and Economics from ETH Zurich. His research spans theoretical foundations and practical applications of AI and data science and has received multiple best paper awards. His work has been published in leading conferences such as NeurIPS, AAAI, IJCAI, and ICIS, and in prestigious journals including the *Journal of the ACM*, *European Journal of Information Systems*, and *Business & Information Systems Engineering*.