Algorithm Lab. Journal of Selected Papers on ICAITD 2025 ICAITD 2025

Fostering Empathetic Reframing in Human–AI Dialogue through Structured Abductive Hypothesis Selection

Masaki Shimada†, Koyo Matsuoka†, Marina Hamaguchi†, Hajime Sato†, Takuma Miyata†, Takahiro Suda†, Shigeru Hosono†
†School of Computer Science, Tokyo University of Technology, Tokyo, Japan

Abstract: This study explores how empathetic reframing can be realized in human–AI dialogue through a structured abductive process. We propose a multi-agent dialogue architecture that coordinates personal context, external information, and nudge-based response design to support reinterpretation while preserving user agency. Using a tourism consultation scenario—where users express strong preferences for "authentic" travel experiences—we examine how the system externalizes and contrastively presents abductive hypotheses in the final response. Rather than evaluating internal reasoning alone, the system focuses on the structural presentation of interpretations. Experimental results show that reframing depends not only on hypothesis selection, but on how personal meaning and situational context are grounded and exposed in dialogue. These findings suggest that within the examined settings, reframing appears to emerge as a property of dialogue structure rather than isolated reasoning mechanisms.

Keywords: Knowledge Graph, Nudge Theory, Abduction, Empathy, Creativity, AI Agent

1. Introduction

When multiple designers collaborate in design work, genuinely new ideas rarely emerge unless they understand and value one another's perspectives and background knowledge. Empathy toward a dialogue partner can foster abductive reasoning in designers, supporting the generation of new interpretations. Building on this premise, we construct a dialogue environment for interaction between humans and AI agents and visualize how empathic engagement supports abduction. Through conceptual verification using illustrative examples, we aim to develop an environment that facilitates creative design in human–AI collaboration.

Despite extensive discussion of empathy, abduction, and creativity in cognitive science and artificial intelligence, it remains unclear how these processes can be realized in human–AI dialogue in a way that helps users reach new interpretations on their own. Many dialogue systems prioritize solution generation or embed reasoning within a single agent, which makes it difficult to examine how alternative perspectives emerge through interaction. To address this gap, we treat dialogue as a reframing process rather than a means of delivering solutions and investigate how empathy-driven abductive reasoning can be structured at the system level to support creative reinterpretation while preserving user agency.

The remainder of this paper is organized as follows. Section 2 describes how knowledge graphs and nudging theory are used to structure an empathy-fostering process. Section 3 outlines abduction as a mechanism through which creativity can arise from an empathic state. Section 4 reviews related work on reframing, abductive reasoning, and human–AI dialogue. Section 5 validates the feasibility of collaborative abductive reasoning through preliminary multi-agent implementations, and Section 6 operationalizes abduction as a structured cognitive trajectory within an integrated dialogue architecture. Section 7 defines the experimental framework for comparative analysis across system configurations, and Section 8 examines how this operationalized abduction manifests in the structural properties of final responses. Section 9 discusses implications for system design and identifies limitations, and Section 10 concludes the paper.

2. Cultivation of Empathy

2.1 Knowledge Graphs for Partner Understanding

A knowledge graph is a networked representation of knowledge obtained from databases and information expressed in graph format (NTT 2021; Hogan et al. 2021). Knowledge graphs enable reliable and precise information retrieval by facilitating the accurate understanding of relationships between pieces of information. For example, entities such as people, places, and organizations are represented as nodes, with their interrelationships depicted as edges. Knowledge graphs serve as tools for understanding

https://doi.org/10.63211/jst.icaitd.2025.260702

received on Feb. 7 2026
accepted on June 17 2026

the meaning and relevance of information, visually organizing complex data, and supporting efficient decision-making.

Initiating a conversation without much prior knowledge about the other person and quickly moving to a coherent discussion or debate can be quite challenging. Constructing a Personal Knowledge Graph (PKG) from the dialogue partner's background knowledge and past statements can increase points of contact and deepen mutual understanding (Balog and Kenter 2019). In this study, the PKG additionally functions as a "shared mental model" and an "empathetic anchor," so that conceptual leaps made through abduction remain grounded in the user's unique context rather than becoming irrelevant or intrusive.

2.2 Nudging Partners

Nudge theory, a branch of behavioral economics, was proposed by Professors Richard Thaler and Cass Sunstein (Thaler and Sunstein 2003). A "nudge" means a gentle push and is a method of subtly encouraging people's decision-making and actions without using economic incentives or coercion (Ministry of Health, Labor and Welfare, n.d.). Examples of nudges include footprint stickers indicating where to stand in line at a cash register or signs in restrooms saying, "Thank you for keeping it clean." Nudge theory is widely applied in business and government settings as an effective method for promoting behavioral change.

In conversation, including nudge-based content in remarks or responses can create awareness and encourage behavioral change in the other person. Our approach specifically operationalizes the delivery of a "reframed perspective" as a conversational nudge. By presenting new insights discovered via abductive reasoning as subtle suggestions rather than directives, the system encourages the user to naturally internalize these perspectives. Additionally, through these dynamic interactions, mutual understanding can be deepened.

3. Eliciting Creativity through Abduction

While classical strategies such as Occam's Razor, default reasoning, and heuristic search provide the abstract principles for this reasoning (Peirce 1993; Inoue 1992; Koshino 2005; Saito and Inoue 2006; Inoue 2010), we operationalize them as an interactive multi-agent trajectory. Rather than a single-step inference, abduction is modeled as a cyclical process: identifying a surprising observation (Magnani 2001), generating and evaluating hypotheses, and selecting a reframed perspective (Boden 1998).

Crucially, the system presents these abductive insights as provisional explanations accompanied by discriminative questions. This design encourages users to evaluate the hypothesis against their own experiences, supporting the continuation of the abduction cycle through dialogue. By integrating these strategies into a multi-agent pipeline, the system moves beyond simple problem clarification, facilitating the co-creation of new insights and behavioral change.

4. Related Work

4.1 Cognitive Foundations of Reframing

While existing dialogue systems often focus on providing direct solutions, cognitive science suggests that effective reframing requires a shift in construal level, as described by Construal Level Theory (CLT) (Trope and Liberman 2010). High-level construals (abstract representations) allow individuals to transcend current constraints and focus on primary goals (the "Why"), whereas low-level construals (concrete details) focus on feasibility and means (the "How") (Vallacher and Wegner 1987).

Attempting to shift a user's perspective by immediately presenting concrete solutions often triggers cognitive resistance due to conflicts with the user's existing mental models (Watzlawick, Weakland, and Fisch 1974). To overcome this, the proposed approach leverages CLT by first elevating the dialogue to a high-level construal—abstracting the user's latent intent to foster consensus on value. Subsequently, applying Structure Mapping Theory (Gentner 1983), which prioritizes relational similarities over superficial attributes, allows the system to introduce novel, concrete instances via abduction. This "Abstraction-to-Concretion" flow is essential for minimizing cognitive load while maximizing the acceptance of creative nudges. However, these theories primarily describe individual cognitive shifts and do not directly address how such reframing can be elicited and sustained through interactive dialogue, particularly when the reframed perspective is not explicitly requested by the user.

4.2 Importance of Abduction in AI

Medianovskyi and Pietarinen (2022) argue that traditional explainable AI (XAI), which heavily relies on causal and inductive reasoning, does not sufficiently answer the human-centered question of why the system reaches a given conclusion (Miller 2019). They emphasize the importance of abduction

https://doi.org/10.63211/jst.icaitd.2025.260702
received on Feb. 7 2026
accepted on June 17 2026

(hypothetical reasoning) to generate and present multiple plausible explanations according to the user's context. This approach transforms explanations from fixed outputs to interactive processes, referred to as participatory explanation. This stance is based on the recognition that humans require contextually plausible reasoning rather than mere enumeration of causal relationships for understanding explanations (Lombrozo 2006). Explanations should adapt to the listener's background knowledge and situation, necessitating a flexible reasoning model capable of generating and contextualizing hypotheses.

In line with their approach, this study aims to realize creative dialogues that derive and present hypotheses from user utterances, promoting the development of thoughts and insights by incorporating abduction into the dialogue. Zhao et al. (2024) extend abductive reasoning to uncommon situations, yet these studies do not explicitly examine how abductive explanations can function as creative reframing mechanisms that transform the user's original interpretation through dialogue.

4.3 Integration of Reasoning Including Abduction

Recent studies have also examined abductive inference in LLMs (Dalal et al. 2024). He et al. (2024) propose the framework IDEA (Induction–Deduction–Abduction) to model human-like rule learning as a cyclical reasoning process involving hypothesis generation, verification, and reinforcement. Their LLM-based agent integrates induction, deduction, and abduction, learning and applying rules through interaction with structured environments. This reasoning loop has been shown to improve rule discovery and consistency in task environments like the Escape Room. Aggregating the reasoning process into a single agent enables flexible and consistent knowledge acquisition.

Based on these insights, this study adopts a multi-agent configuration that focuses on open-ended dialogues, distributing processes such as hypothesis generation, selection, and presentation among agents. Through this design, the study aims to achieve a balance between the development of creative ideas and contextually coherent responses. However, by aggregating induction, deduction, and abduction within a single agent (He et al. 2024), these approaches primarily optimize internal reasoning consistency, rather than supporting dialogical processes where competing interpretations are externalized, contrasted, and negotiated with the user (Wooldridge 2009). Similarly, Tian et al. (2024) evaluate creative problem-solving in LLMs but do not address how creativity is structurally embedded in dialogue.

4.4 Integration of Understanding Others and Hypothetical Reasoning

Rodrigues et al. (2023) propose the agent model TomAbd, which integrates Theory of Mind (ToM) and abduction. This model enables flexible and consistent decision-making in social contexts by hypothetically estimating others' beliefs and intentions. The model is based on the BDI (Belief–Desire–Intention) architecture (Rao and Georgeff 1995) and has been evaluated through the cooperative card game Hanabi (Bard et al. 2020). By reconstructing others' perspectives and selecting appropriate actions based on them, the model has shown improvements in team performance. Although not explicitly aimed at creativity, the processes of belief reconstruction, externalization of tacit knowledge, and generation of new explanations suggest the potential for emergent knowledge formation.

Building on this insight, this study advances the integration of hypothetical reasoning by shifting the focus from merely inferring others' mental states to stimulating user insights and idea generation through dialogue. To achieve this, we employ a Personal Knowledge Graph (PKG) as a dynamic "shared mental model" among agents (Jonker et al. 2011). Unlike transient ToM states, the PKG accumulates the user's long-term context, interests, and abstract values, which provides a stable grounding for agents to perform empathy-driven abduction. By configuring hypothesis generation, selection, and presentation with multiple agents based on this PKG, the study aims to design a co-creative and emergent dialogue system. Overall, these studies provide important foundations for understanding reframing, abductive reasoning, and perspective-taking in both cognitive science and AI systems. However, most prior work examines these elements either as internal reasoning mechanisms or as explanatory functions of AI, rather than as dialogical processes that unfold through interaction with the user. In contrast, this study focuses on how meaning transformation emerges at the level of the final dialogue response, shaped by the explicit coordination of personal context, external exploration, and contrastive hypothesis presentation. By treating abduction not as an internal inference but as an interactional resource for reframing, we position our system at the intersection of empathetic dialogue, abductive reasoning, and user-centered meaning-making.

received on Feb. 7 2026
accepted on June 17 2026

5. Empathetic and Abductive AI Agents

This section reports preliminary implementations that individually explore knowledge graphs, nudging, and abductive reasoning, and identifies the design challenges that motivated the integrated architecture in Section 6.

5.1 Applying Graph RAG to AI Agents

To operationalize this pipeline, we implemented a dialogue system in which agents construct and query a PKG using Graph-RAG. AI agents are autonomous AI systems that make decisions independently with Graph RAG (Retrieval-Augmented Generation). Graph RAG is a technique that combines the retrieval of relevant information from a knowledge graph with the generation of responses using large language models (Peng et al. 2024; Edge et al. 2024). Unlike traditional vector-based RAG, which retrieves information as isolated chunks, Graph RAG preserves the structural relationships between entities, enabling the model to perform multi-hop reasoning and better understand complex contexts (Yasunaga et al. 2021). This allows the agents to generate hypotheses and solve problems through user interactions.

To enable AI agents to provide contextually relevant responses during dialogue, we extend the approach to combine knowledge graphs and large language models (LLMs) (Fukuda and Egami 2024; Pan et al. 2024) for implementation in the agent design. The AI agent was designed to analyze dialogue content, generate knowledge graphs composed of nodes (representing people, actions, motives, and so on) and edges, and retrieve relevant information from these graphs using Graph RAG. This approach allows AI agents to better understand relationships between different pieces of information, thereby enabling more accurate, coherent, and contextually appropriate responses during dialogue. Recent studies suggest that grounding LLM generation in structured knowledge graphs significantly mitigates the risk of hallucinations and improves the factual consistency of the generated content (Ji et al. 2023). The retrieved information is then used to generate responses via LLMs. By using knowledge graphs, AI agents can better understand the relationships between different pieces of information, leading to more coherent and meaningful dialogues. This approach also helps in retrieving relevant information quickly, improving the overall efficiency of the dialogue system.

In the initial implementation, the AI agent constructed a knowledge graph from provided information on a specific topic. For the preliminary experiment, we simulated a dialogue between two AI agents on the topic of planning a three-day trip to Kyoto. The knowledge graph generated from the given information is shown in Fig. 1. The dialogue using this knowledge graph confirmed an increased likelihood of consensus formation regarding the travel plan. However, it was also found that the implementation heavily depended on prompt design, limiting its applicability to more complex dialogues.

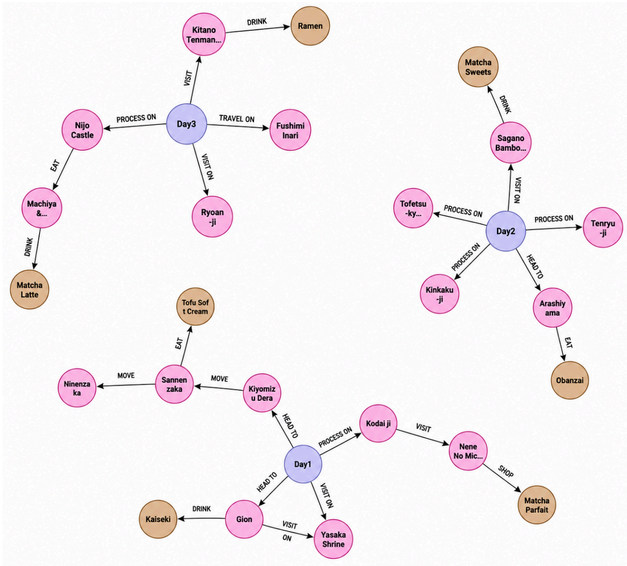

Fig.1. Generated Knowledge Graphs (image upscaled for print legibility; content unchanged)

The effectiveness of the knowledge graph and the quality of the generated responses were highly influenced by the initial prompts provided to the AI agents. This issue of "prompt sensitivity," where minor variations in input significantly alter the output, is a well-documented challenge in current LLM research (Zhao et al. 2021).

This limitation indicates that more flexible and adaptive prompt design techniques are needed.

5.2 Applying Nudge Theory to AI Agents

We applied the EAST framework (Service et al. 2014) of nudge theory to an AI agent. The EAST framework characterizes effective nudges as those that make actions easy, attractive, socially salient, and timely, thereby guiding behavior by making actions easy to perform, drawing attention to important choices, leveraging social influences, and providing prompts at the right time.

https://doi.org/10.63211/jst.icaitd.2025.260702 received on Feb. 7 2026
accepted on June 17 2026

The prototype implementation consisted of three AI agents: a Nudge Agent analyzing user behavior, a Tour Agent making specific suggestions, and an Evaluation Agent evaluating dialogue outcomes. These agents worked in collaboration to facilitate smooth interactions with humans, as shown in Fig. 2.

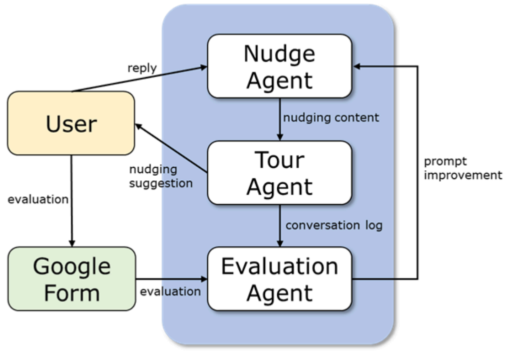

Fig.2. Nudge-based Multi-agent System

The preliminary experiment involved a scenario where users decided on places to visit during a trip. When a user is unsure about which tourist spot to visit, the Nudge Agent might suggest a popular destination that is frequently visited, making the choice easier and more attractive. By incorporating nudges, AI agents can gently guide users towards beneficial decisions without imposing or restricting their choices. This approach enhances user satisfaction and engagement by making interactions more intuitive and supportive.

Additionally, to confirm the effectiveness of nudge interventions, we conducted scoring evaluations of responses with and without nudge theory from the perspective of the EAST framework (Service et al. 2014), scoring each aspect from 1 to 10. A survey validation by five users of various ages was also conducted in the context of tourism around Hachioji Station. Detailed scoring results and survey data are reported in the prior work (Shimada et al. 2025).

However, one of the challenges faced was striking the right balance between providing helpful nudges and avoiding excessive guidance. Excessive nudging poses the risk of "automation bias," where users blindly follow AI suggestions without critical validation. This can lead to over-reliance on the AI system, potentially reducing the user's autonomy and decision-making ability (Parasuraman and Manzey 2010). These insights motivate our transition toward a structured multi-agent architecture, where reasoning responsibilities are explicitly divided and grounded in the PKG, helping mitigate prompt sensitivity while enabling transparent and context-sensitive nudging.

5.3 Applying Abductive Prompts to AI Agents

To address the challenge of inferring latent variables in complex phenomena—such as hidden human motives or potential travel desires—we utilized AutoGen (Wu et al. 2023) to construct a collaborative multi-agent ecosystem. Unlike single-agent systems limited to a fixed reasoning path, multi-agent collaboration enables the synthesis of diverse perspectives, a requirement for high-level abductive reasoning. We integrated distinct reasoning strategies—including Occam's Razor and heuristic search—not merely as standalone tools, but as cognitive lenses distributed among interacting agents. This design allows the system to simulate a dialectical process, essential for navigating the uncertainty inherent in creative tasks (Mercier and Sperber 2011).

Before deploying this architecture to the subjective domain of tourism planning, we validated its collaborative efficacy using a "detective scenario," a domain requiring the inference of unobservable causes from fragmentary evidence. The system's architecture (Fig. 3) was designed to foster emergent reasoning through interaction rather than simple processing.

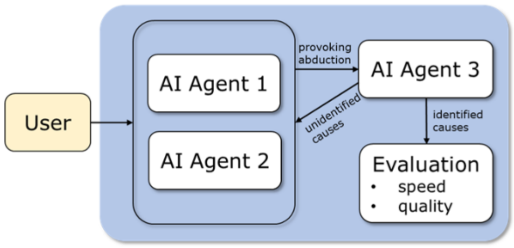

Fig.3. Abduction Multi-Agent System

Agent 1 ("The Detective") provided baseline logical deductions, while Agent 2 ("The Divergent Thinker") was explicitly constrained to inject alternative hypotheses using abductive strategies, acting as a catalyst to break cognitive fixation. Agent 3 ("The Evaluator") then synthesized these conflicting inputs. This triadic interaction mirrors the human cognitive process of conjecture and refutation (Popper 1963), allowing the system to tackle complex causality that a single agent might oversimplify.

The comparative validation revealed that while individual methods have specific limitations—such as "Occam's Razor" failing in non-linear complexity or "heuristic search" risking instability—the collaborative interplay mitigated these weaknesses. The interaction allowed valid logic to ground conceptual leaps, confirming that effectively handling complex, multi-faceted phenomena requires a dynamic orchestration of diverse reasoning strategies.

received on Feb. 7 2026
accepted on June 17 2026

This finding grounds our transition to the Integrated Multi-Agent System (Section 6), establishing that robust abduction in ambiguous domains emerges not from a single algorithm, but from the strategic coordination of specialized cognitive agents. The architectural separation introduced in Section 6 was motivated by issues observed in Section 5, including prompt sensitivity and implicit automation bias.

6. Integrated AI Agents System for Dialogues

6.1 Conceptual Framework and System Architecture

The proposed method is centered on a triadic cognitive process consisting of Empathy, Abduction, and Nudging to systematically facilitate reframing (Miller 2019; Medianovskyi and Pietarinen 2022). This process is operationalized through a sequential multi-agent pipeline where the Personal Knowledge Graph (PKG) serves as the foundational shared mental model. As illustrated in Fig. 4, the architecture coordinates specialized autonomous units: the Context Agent, the Abduction Multi-Agent, and the Dialogue Agent. By shifting from a monolithic reasoning model to this distributed architecture, the system transforms the dialogue from a simple information exchange into a structured cognitive intervention (Service et al. 2014). The integration of the PKG is designed so that the empathy established at the beginning remains a persistent constraint, preventing the abduction phase from drifting into generic suggestions and instead promoting a transformation that is uniquely grounded in the user's identity.

6.2 Context Agent: Dialogue Grounding

The Context Agent initiates the interaction by aligning the user's immediate input with the data stored in the PKG to establish a stable dialogue grounding. This grounding acts as a "psychological anchor," which is crucial because it changes the user's receptivity to radical new ideas. By synthesizing current intentions with long-term characteristics—such as hobbies, goals, and values—this agent is designed to maintain a deep, persistent understanding of the user. This fundamentally shifts the dynamic of the reframing process: rather than the AI "guessing" what the user might like, the system operates from a position of "shared history," thereby aiming to reduce the psychological reactance that typically occurs when an external agent suggests a change in perspective (Parasuraman and Manzey 2010).

6.3 Abduction Multi-Agent: Perspective Expansion and Selection

6.3.1 Design Philosophy: The Abductive Reasoning Trajectory

The Abduction Multi-Agent operationalizes abduction as a structured cognitive trajectory where external evidence and internal context are progressively synthesized through four distinct stages. This trajectory is designed to mimic the human process of discovering and validating new perspectives by shifting from the identification of an anomaly to the final adoption of a new viewpoint. The process begins with the Trigger and Exploration phase, where the system identifies surprising observations—anomalies arising from the gap between immediate observations (O) and implicit assumptions (A) provided by the Context Agent—and retrieves relevant external data through Retrieval-Augmented Generation to serve as the raw material for reasoning. This moves into the Hypothesis Generation phase, where the retrieved external material is synthesized with the user's personal context and immediate requirements to produce diverse explanatory candidates. This is followed by an Evaluation phase where a critic provides qualitative linguistic materials rather than mere numerical scores regarding the logic and persona alignment of each hypothesis.

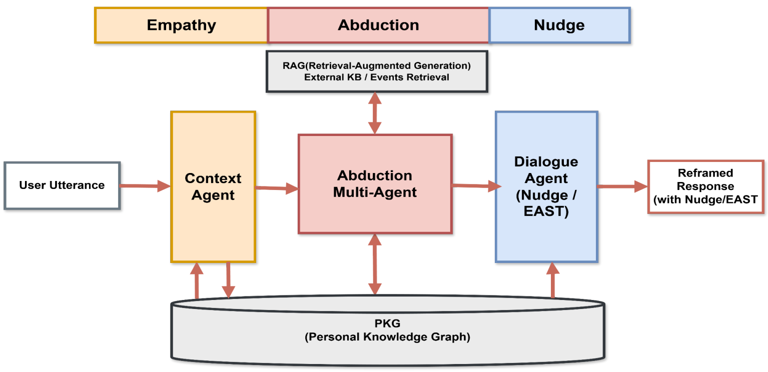

Fig.4. Integrated Multi-agent Dialogue System

received on Feb. 7 2026
accepted on June 17 2026

Finally, the Selection phase concludes the trajectory as a decision agent synthesizes these critiques to adopt the most robust and creative reframing. This explicit flow is designed so that the resulting insight is not a random output but a justifiable conceptual leap with logical grounding. As illustrated in Fig. 5, the system navigates through these four cognitive stages to transform raw observations into a meaningful perspective.

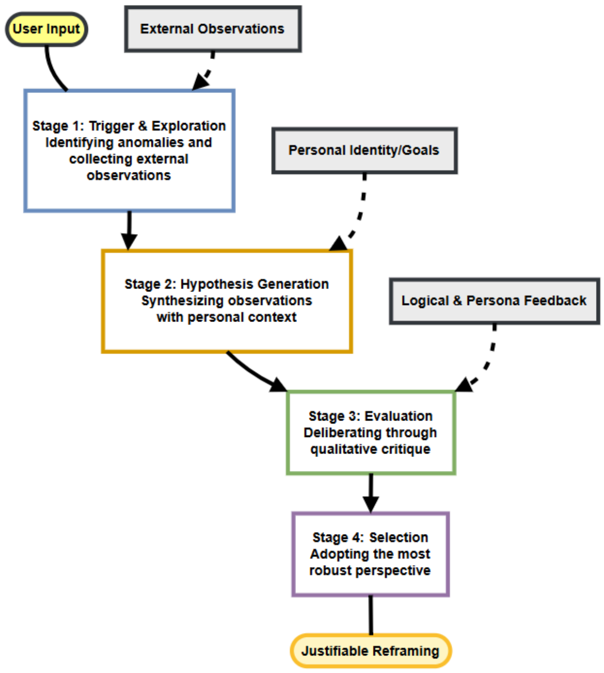

Fig.5. The structural trajectory of abductive reasoning

6.3.2 Core Principles for Collaborative Reasoning

To support the goal of maintaining both empathetic and creative qualities in the abductive trajectory, the multi-agent system is guided by three fundamental design principles. The first principle focuses on the proactive breaking of cognitive fixation through integrated discovery (Boden 1998). By combining trigger identification and external exploration into a single cohesive step, the system identifies contradictions or gaps between the current observations and the user's internal assumptions, utilizing external search to reveal blind spots that a user cannot perceive alone. The second principle is empathetic grounding via the Personal Knowledge Graph. To prevent conceptual leaps from becoming irrelevant or intrusive, the knowledge graph maintains a constant constraint on hypothesis generation, aiming to keep even the most novel perspectives anchored in the user's identity so the reframing is internalized as a natural extension of their own thought process. The third principle is the decoupling of critique and decision through the Agent-as-a-Judge framework (Zheng et al. 2023). By isolating the functions of critiquing and deciding, the system allows the critic agent to generate nuanced linguistic feedback—identifying specific logical flaws and alignment strengths—rather than reductive scores. This qualitative material is essential for the decision agent to perform a sophisticated synthesis that balances logical rigor with creative potential.

6.3.3 Functional Roles within the Multi-Agent Pipeline

The specific roles of the four agents within this pipeline are operationalized to transform raw data into refined insights through specialized cognitive tasks, as illustrated in Fig. 6.

The Explorer Agent serves as the initiator by both identifying abductive triggers and retrieving non-obvious facts from external knowledge bases, providing the serendipitous spark necessary for breaking the user's current context. Next, the Hypothesis Agent synthesizes this external evidence with the topological structures of the Personal Knowledge Graph by connecting disparate concepts from the user's long-term goals and immediate needs to support a diverse exploration of potential reframing options.

Subsequently, the Critic Agent acts as a pragmatic judge by providing structured, linguistic critiques based on explanatory power, consistency, persona alignment, and evidence reliability. It highlights the specific pros and cons of each hypothesis to inform the final selection, deliberately providing descriptive feedback instead of simplified numerical metrics (Koo et al. 2024). Finally, the Decision Agent performs the final value transformation by integrating the feedback from the critic. By resolving the nuances between different qualitative critiques, it selects a pair of contrasting hypotheses—a primary candidate and a meaningful alternative. This is designed so that the system does not merely propose interesting ideas but outputs a choice-based reframing based on these dual hypotheses that is ready for empathetic delivery to the user via the Dialogue Agent.

received on Feb. 7 2026
accepted on June 17 2026

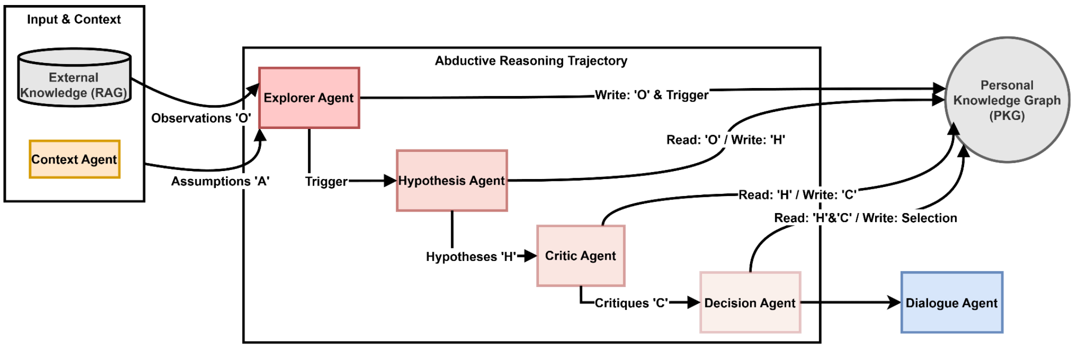

Fig.6. Functional Roles and Data Flow in the Multi-agent Abductive Pipeline

6.4 Mitigation of Reframing Risks via Nudge Theory

To address the inherent risks of reframing, such as perceived intrusiveness or high cognitive load, the Dialogue Agent incorporates Nudge theory, specifically the EAST (Easy, Attractive, Social, Timely) framework. This delivery mechanism transforms the AI's "inference" into the user's "insight." Leveraging the dual hypotheses selected in the previous stage, rather than presenting a single authoritative directive, the system delivers them as engaging inquiries. This shifts the power dynamic of the interaction, preserving user autonomy and aiming to support the user's perception that the conceptual leap is something they 'discover' rather than something the AI "imposes." This nudge-based completion of the cycle is designed to keep the sophisticated abductive reasoning empathetic and ethically grounded (Parasuraman and Manzey 2010).

7. Experimental Design

This section defines the experimental framework used to evaluate the proposed multi-agent system. The objective is to measure how the coordination of personal context and external evidence facilitates the generation of grounded abductive reframing while maintaining logical consistency.

7.1 Experimental Conditions

We established four conditions to isolate the contributions of the system's architecture and data sources, as summarized in Table 1. All conditions utilize the same underlying Large Language Model.

Table 1: Experimental Conditions

Condition	Architecture	PKG	RAG
A	Multi-Agent	✓	✓
B	Multi-Agent	✗	✓
C	Multi-Agent	✓	✗
D	Single-Agent	✓	✓

Condition A represents the proposed full system with a decoupled pipeline. Conditions B and C serve as controlled configurations designed to isolate the contributions of the Personal Knowledge Graph (PKG) and External Exploration (RAG) (Lewis et al. 2020), respectively. Condition D serves as the structural baseline, performing integrated reasoning through a monolithic Chain-of-Thought (CoT) process (Wei et al. 2022) despite having access to all data. These comparisons are not intended as ablation studies in the strict sense; rather, they serve as architectural contrast experiments that observe, at a feasibility level, how each information source and structural configuration contributes to reframing in the final response. All conditions were implemented using Azure OpenAI GPT-4.1 (Microsoft 2024) with JSON-enforced output and agent-specific temperature settings (ranging from 0.2 to 0.5). External information retrieval was performed via the Serper API (Serper 2024), and the Personal Knowledge Graph was managed using Neo4j (Neo4j 2024). The PKG was reset to its seed state before each run to prevent cross-run contamination. During each dialogue session, the Context Agent updates the PKG by adding newly extracted entities and relationships from the user's utterance prior to initiating the abductive reasoning pipeline. Each agent was assigned a role-specific system prompt corresponding to the functional roles described in Section 6.3.3. Experimental control across conditions was maintained by excluding PKG reference instructions (Condition B), disabling external search capabilities (Condition C), and consolidating all agent

received on Feb. 7 2026
accepted on June 17 2026

roles into a single Chain-of-Thought prompt (Condition D). A summary of the prompt instructions for each agent is provided in Table 2. Full prompt templates and implementation code are available in the project repository (Service System Design Laboratory 2026).

Table 2. Summary of Agent-Level Prompt Directives

Agent	Core Prompt Directive
Context (0.2)	Extract observations, assumptions, and constraints; align PKG context with current input
Explorer (0.2)	Identify abductive triggers; retrieve assumption-challenging facts via RAG
Hypothesis (0.5)	Synthesize context and evidence into diverse abductive candidates; ≥1 reframing required
Critic (0.2)	Evaluate hypotheses via qualitative diagnostic feedback, no numerical scores
Decision (0.2)	Select primary (reframing-preferred) and contrastive alternative hypotheses
Dialogue (0.3)	Deliver as contrasting, empathetic, nudge-based inquiries

7.2 Scenario Design

We employ tourism consultation as the experimental domain for this study. Tourism represents a "soft" problem space where subjective meaning-making is prioritized over deductive optimization. Planning a trip is intrinsically linked to the user's identity and implicit assumptions, often involving cognitive fixations that limit their perspective. The system's task is to identify 'O/A Gaps' (discrepancies between observations 'O' and assumptions 'A') based on the identity stored in the PKG—to generate hypotheses that effectively reframe the user's situation. Each condition was applied to a fixed set of scenarios–persona pairs (3 pre-defined virtual persona profiles constructed from the Nemotron-Personas-Japan dataset (NVIDIA 2024) × 10 scenarios), resulting in 30 cases per condition.

7.3 Evaluation Criteria

We evaluate the success of the interaction through four qualitative lenses, focusing primarily on the system's final response presented to the user. The operational definitions and decision rules for all four criteria are provided in the Annotation Rubric (Appendix B). Reframing Occurrence assesses whether the dialogue resulted in a qualitative shift in the user's goals, values, or interpretation of the situation, rather than a mere elaboration of the original framing.

Surprise examines whether the system introduced non-obvious yet plausible perspectives that challenge the user's existing mental model. Groundedness verifies that the reframed perspective is anchored in entities from the Personal Knowledge Graph (PKG) or externally retrieved facts, thereby distinguishing meaningful insights from hallucinations. Finally, User Agency Preservation measures whether the reframing was delivered as a collaborative choice—via contrasting hypotheses—rather than as a directive, so that interpretive control remains with the user.

We do not evaluate intermediate representations such as observations and hypotheses independently, but reference them only as analytical support for interpreting the final response. Two raters independently labeled all outputs using the rubric in Appendix B, and inter-rater agreement was assessed using Cohen's κ (reported in Section 8.1). The final assessment focused on whether the system's response introduced an explicit reinterpretation of the user's initial framing.

We acknowledge several limitations of this evaluation design. The sample size (3 persona profiles × 10 scenarios = 30 cases per condition) is intended as an exploratory qualitative analysis rather than a statistically generalizable study. The personas are synthetically constructed profiles, not real users, which limits ecological validity. Furthermore, the evaluation was conducted within a single domain (tourism consultation), and generalizability to other domains remains untested.

8. Results

This section examines how structural differences across conditions are reflected in the final response, focusing on the manifestation of reframing rather than quantitative performance. The analysis focuses on which hypotheses the system selected across conditions and the structural format in which they were presented as final responses. Summary of selected hypotheses and corresponding final responses for representative scenarios in each condition are provided in Appendix A.

8.1 Comparison Overview

We observed notable variations across conditions regarding the selection rates of hypotheses categorized as Reframing at the hypothesis level (Fig. 7). Across the 30 cases per condition described in Section 7.2, in Condition A, a reframing hypothesis was selected in approximately 67% of the cases, whereas this rate dropped to approximately 37% in Condition B.

Conversely, Conditions C and D exhibited substantially higher selection rates at 83% and 100%, respectively. However, the final response structures demonstrate that these figures do not represent a simple metric of success. Importantly, the differences observed across conditions should not be interpreted as direct indicators of performance superiority, but as reflections of how structural configurations influence the manifestation of reframing in the final response. Higher selection rates do not necessarily correspond to qualitatively distinct reframing; rather, these comparisons reveal structural tendencies in how hypotheses are consolidated or exposed in dialogue. Therefore, the primary contribution of these comparisons lies in revealing structural tendencies rather than establishing quantitative performance rankings.

To assess the reliability of the qualitative labeling, two raters independently annotated all outputs using the rubric provided in Appendix B. Inter-rater agreement was evaluated using Cohen's κ, yielding $\kappa = 0.44$ for Reframing Occurrence and $\kappa = 0.42$ for Groundedness, indicating moderate agreement; given the inherently interpretive nature of the task, our analysis focuses on aggregate structural trends across conditions rather than individual-instance judgments.

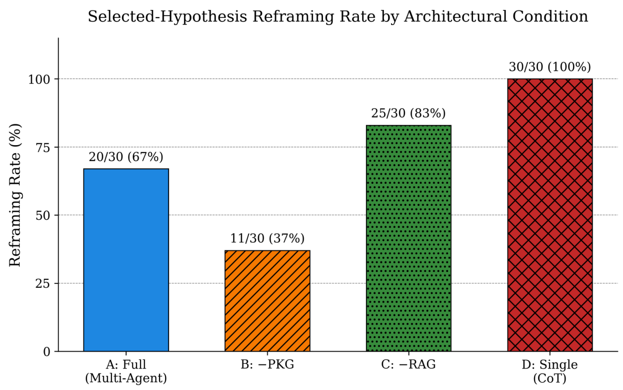

Fig.7. Selected-Hypothesis Reframing

In Condition A, most instances where a reframing hypothesis was selected utilized a structure that juxtaposed contrasting interpretations, allowing users to evaluate multiple possibilities. In Condition B, the low selection rate was accompanied by a tendency for final responses to focus on event-based or environmental factors, offering minimal transformation of the interpretative frame. While Condition C showed a higher selection rate (83%) than Condition A, these hypotheses were often presented in isolation, resulting in limited situational breadth or external perspective. Finally, although Condition D selected a reframing hypothesis in 100% of cases, responses tended to present alternatives that varied within the same interpretive direction, rather than offering qualitatively distinct reframing of the user's premise.

Table 3: Grounding Link Frequency by Condition

Metric	A: Full	B: −PKG	C: −RAG	D: Single Agent
persona_link	18/30	0/30	26/30	16/30
event_link	30/30	28/30	0/30	28/30

The structural divergence was further reflected in the presence of Grounding links (Table 3). In Condition A, links to external events were maintained at 100%, while persona-based links were confirmed in 60%. Due to the removal of the PKG, Condition B showed 0% persona-based links, while event links remained at approximately 93%. In contrast, the absence of external exploration in Condition C resulted in persona-based links remaining high at approximately 87%, but event links dropped to 0%. Condition D displayed moderate values for both, with persona-based links at approximately 53% and event links at approximately 93%. These results demonstrate that the conditions varied not only in selection rates but also in the informational grounding and structural delivery of the final responses. The subsequent sections examine these differences for each condition in detail.

8.2 Condition A: Emergence of Reframing

This section examines whether Condition A, which integrates PKG, RAG, and explicit hypothesis selection, exhibited behavior consistent with the intended design. The analysis focuses on how the selected hypothesis was reflected in the final response, rather than on quantitative performance.

As shown in Appendix A, Condition A generated multiple candidate hypotheses grounded in both personal context and external observations. The selected hypothesis was not presented in isolation but contrasted with an alternative interpretation, resulting in a final response structured around choice rather than conclusion. This structure allowed reframing to emerge without negating the user's original stance.

In the final responses, empathetic acknowledgment of the user's initial concern preceded the presentation of contrasting perspectives, and reframed interpretations were delivered in a nudge-like manner rather than as directives. This indicates that the integration of PKG, external observations, and explicit hypothesis selection functioned as intended, enabling reinterpretation while preserving user agency.

https://doi.org/10.63211/jst.icaitd.2025.260702
received on Feb. 7 2026
accepted on June 17 2026

8.3 Condition A vs. Condition D: Structural Difference in Reframing

Although both Condition A and Condition D frequently resulted in reframing-labeled hypotheses at the hypothesis level, their behaviors differed in how reframing was realized in the final response.

In Condition A, reframing emerged through the explicit separation and contrast of multiple hypotheses. As shown in Appendix A, the selected hypothesis was presented alongside alternative interpretations, allowing the user to recognize reframing as a choice rather than as a resolved conclusion. This contrastive structure made clear how personal context and external observations jointly informed the reinterpretation.

By contrast, Condition D integrated hypothesis generation and selection within a single agent. While reframing-labeled hypotheses were often selected, the final response tended to consolidate interpretation into a single coherent narrative. Although alternative interpretations were formally presented, they tended to represent variations within the same interpretive frame rather than qualitatively distinct reframings. As a result, reframing appeared as a directionally resolved perspective rather than as a genuinely contrastive choice.

This comparison indicates that the difference between Conditions A and D lies not in the availability of information or the presence of reframing hypotheses, but in whether reinterpretation is structurally exposed to the user as a contrastive space. In Condition A, reframing functions as an invitation to reinterpret, whereas in Condition D it is more likely to be implicitly accepted as the system's conclusion.

8.4 Conditions B and C: Effects of Contextual and External Configurations

Conditions B and C clarify how different sources of information contribute to reframing when architectural separation is maintained.

In Condition B (−PKG), hypotheses were primarily derived from external observations and generalized patterns. As shown in Appendix A, the selected hypotheses often addressed structural or informational aspects of the situation, but lacked anchoring in the user's personal context. Consequently, final responses tended to remain explanatory rather than reframing-oriented, and reinterpretations rarely extended to the user's self-concept or values.

Condition C (−RAG), by contrast, generated hypotheses strongly grounded in the user's personal context. Reframing-labeled hypotheses were frequently selected; however, without external observations, the scope of reinterpretation remained inward-facing. As reflected in the final responses, reframing was often limited to reinforcing or slightly rearticulating existing self-images, without introducing perspectives that challenged the broader framing of the situation.

Together, these conditions indicate that neither personal context nor external observations alone are sufficient to support grounded reframing. Rather, reframing emerges when both are present and structurally coordinated, allowing hypotheses to bridge personal meaning and situational context without collapsing into either abstraction or introspection.

9. Discussion

9.1 Structural Nature of Reframing in Human–AI Dialogue

The results suggest that reframing in human–AI dialogue is not determined solely by whether a reframing-labeled hypothesis is generated or selected, but by how that hypothesis is structurally externalized in the final response. In Condition A, reframing emerged through a contrastive presentation of interpretations, supported by both personal context and external observations. This structure allowed reinterpretation to appear as an option presented to the user, rather than as a resolved conclusion imposed by the system.

An important factor observed in this condition is that the system effectively performed a form of self-critique: multiple hypotheses were maintained, evaluated, and selectively exposed in the response. This internal critical process contributed to the depth and balance of reframing. In contrast, when such critique relied primarily on a single source of context—most notably in Condition C, where only personal context (PKG) was available—the resulting reframing tended to remain inward-facing. Although reinterpretations were generated, they often reinforced existing self-images rather than expanding the situational frame.

These findings suggest that reframing may be understood as a structural phenomenon: in the conditions examined here, it arose when alternative interpretations were not only generated but explicitly juxtaposed in a way that preserved user agency. Empathy and nudging functioned here as enabling mechanisms, aiming to introduce reframing without invalidating the user's original perspective.

received on Feb. 7 2026
accepted on June 17 2026

9.2 Implications for System Design and Limitations

These observations suggest that reframing in human–AI dialogue cannot be fully explained by the presence or selection of reframing hypotheses alone, but is closely tied to how alternative interpretations are structurally externalized in the final response. This structural perspective naturally leads to considerations at the system design level, where the organization and coordination of reasoning processes directly shape the manifestation of reframing in dialogue.

From a system design perspective, the results highlight the importance of explicitly separating hypothesis generation, evaluation, and response construction. Architectures that collapse these processes into a single agent, as in Condition D, may produce high rates of reframing-labeled hypotheses, but tend to present interpretations as consolidated conclusions. In contrast, the proposed distributed architecture enables internal critique across heterogeneous information sources and exposes multiple interpretations in a contrastive manner, thereby supporting reframing as a dialogically open process rather than a resolved output. Within the examined settings, reframing appears to emerge most clearly when contrasting interpretations are explicitly externalized and grounded in both personal context and external observations.

These findings should be interpreted within the scope of a feasibility-oriented qualitative analysis rather than as statistically generalizable evidence. The experimental design relies on a limited number of synthetic persona–scenario pairs within a single domain (tourism), which constrains external validity. Furthermore, the evaluation focuses on the structural properties of final responses rather than long-term user outcomes or behavioral change. While inter-rater agreement reached a moderate level (Section 8.1), the small sample size and the inherently interpretive nature of the task limit the statistical generalizability of these findings.

From an implementation perspective, the multi-agent architecture introduces additional computational overhead due to multiple sequential LLM calls (Context, Explorer, Hypothesis, Critic, Decision, and Dialogue Agents), compared to the monolithic baseline. Empirical measurements indicate that the proposed pipeline requires approximately 2.6–3.3 times more tokens, while wall-clock latency increases by a factor of 1.2–1.5. Although this overhead did not affect the present offline evaluation, it presents a challenge for real-time applications. Potential optimization strategies include parallelization of independent stages and the use of lightweight models for auxiliary agents.

Finally, this study suggests several directions for future work. First, larger-scale evaluations involving real users across diverse domains are necessary to assess generalizability. Second, longitudinal studies are required to examine whether reframed perspectives are internalized and influence subsequent decisions. Third, complementary evaluation methods, such as user-centered metrics or behavioral measures, should be incorporated to triangulate the findings obtained through rubric-based qualitative analysis.

10. Conclusion

This study investigated how reframing can be realized in human–AI dialogue through system-level structure rather than internal reasoning alone. By focusing on the final response presented to the user, we examined how empathy, abductive hypotheses, and grounded information interact to support reinterpretation.

Our qualitative analysis suggests that reframing is unlikely to emerge from hypothesis generation or selection in isolation within the observed scenarios. Instead, it depends on whether alternative interpretations are explicitly externalized and presented in a contrastive manner. The proposed multi-agent architecture demonstrated the potential to facilitate reframing by coordinating personal context and external events while preserving user agency through nudging-based dialogue. In contrast, configurations lacking such structural separation tended to either narrow reinterpretation or collapse it into a single authoritative perspective.

These findings suggest that reframing in human–AI interaction may be understood as a structural property of dialogue design. Although the current evaluation is limited to synthetic scenarios in a single domain, treating reframing as a collaborative and grounded interpretive process offers a promising foundation for AI systems that support reflection and creative reinterpretation without directing user decisions.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 22K12705.

received on Feb. 7 2026
accepted on June 17 2026

References

Balog, K.; and Kenter, T. 2019. Personal Knowledge Graphs: A Research Agenda. In Proceedings of the 2019 ICTIR International Conference on the Theory of Information Retrieval, 217–220.

Bard, N.; Foerster, J. N.; Chandar, S.; Burch, N.; Lanctot, M.; Song, H. F.; Parisotto, E.; Dumoulin, V.; Moitra, S.; Hughes, E.; et al. 2020. The Hanabi challenge: A new frontier for AI research. Artificial Intelligence 280: 103216.

Boden, M. A. 1998. Creativity and artificial intelligence. Artificial Intelligence 103(1–2): 347–356.

Dalal, D.; Valentino, M.; Freitas, A.; and Buitelaar, P. 2024. Inference to the best explanation in large language models. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 217–235.

Edge, D.; Trinh, H.; Cheng, N.; Bradley, J.; Chao, A.; Mody, A.; Truitt, S.; and Larson, J. 2024. From local to global: A graph RAG approach to query-focused summarization. arXiv preprint arXiv:2404.16130.

Fukuda, K.; and Egami, S. 2024. RAG utilizing knowledge graphs based on document chunks. In Proceedings of the 30th Annual Meeting of the Association for Natural Language Processing, 1856–1861. (In Japanese).

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. Cognitive Science 7(2): 155–170.

He, K.; Zhang, M.; Yan, S.; Wu, P.; and Chen, Z. Z. 2024. IDEA: Enhancing the rule learning ability of large language model agent through induction, deduction, and abduction. arXiv preprint arXiv:2408.10455.

Hogan, A.; Blomqvist, E.; Cochez, M.; d'Amato, C.; Melo, G. d.; Gutierrez, C.; Kirrane, S.; Gayo, J. E. L.; Navigli, R.; Neumaier, S.; et al. 2021. Knowledge graphs. ACM Computing Surveys (CSUR) 54(4): 1–37.

Inoue, K. 1992. Principles of abduction. Journal of the Japanese Society for Artificial Intelligence 7(1): 48–59. (In Japanese).

Inoue, K. 2010. Abduction and induction. Journal of the Japanese Society for Artificial Intelligence 25(3): 389–397. (In Japanese).

Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y. J.; Madotto, A.; and Fung, P. 2023. Survey of hallucination in natural language generation. ACM Computing Surveys 55(12): 1–38.

Jonker, C. M.; Riemsdijk, M. B. v.; and Vermeulen, B. 2011. Shared mental models: A conceptual analysis. In Proceedings of the 13th International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems (COIN), 132–151.

Koo, R.; Lee, M.; Raheja, V.; Park, J. I.; Kim, Z. M.; and Kang, D. 2024. Benchmarking cognitive biases in large language models as evaluators. In Findings of the Association for Computational Linguistics: ACL 2024, 517–545.

Koshino, R. 2005. Heuristic search and hypothesis reasoning acceleration. Journal of the Japanese Society for Artificial Intelligence 20(1): 105–112. (In Japanese).

Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.; Rocktäschel, T.; Riedel, S.; and Kiela, D. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 9459–9474.

Lombrozo, T. 2006. The structure and function of explanations. Trends in Cognitive Sciences 10(10): 264–270.

Magnani, L. 2001. Abduction, Reason and Science: Processes of Discovery and Explanation. New York: Kluwer Academic / Plenum Publishers.

Medianovskyi, K.; and Pietarinen, A.-V. 2022. On explainable AI and abductive inference. Philosophies 7(2): 35.

Mercier, H.; and Sperber, D. 2011. Why do humans reason? Arguments for an argumentative theory. Behavioral and Brain Sciences 34(2): 57–74.

Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence 267: 1–38.

Ministry of Health, Labor and Welfare. n.d. Nudge Theory Usable from Tomorrow. https://www.mhlw.go.jp/content/10901000/000500407.pdf. (In Japanese).

Microsoft Corporation. 2024. Azure OpenAI Service: Large-scale language model deployment. Technical report.

Neo4j, Inc. 2024. Neo4j: A native graph database platform. Technical documentation.

NTT. 2021. What is a Knowledge Graph? Explanation of Features, How to Create, and Applications. https://www.rd.ntt/se/media/article/0025.html. (In Japanese).

NVIDIA. 2024. Nemotron-Personas-Japan [dataset]. HuggingFace. https://huggingface.co/datasets/nvidia/Nemotron-Personas-Japan. Accessed January 17, 2026.

Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; and Wu, X. 2024. Unifying large language models and knowledge graphs: A roadmap. IEEE Transactions on Knowledge and Data Engineering 36(7): 1–19.

Parasuraman, R.; and Manzey, D. H. 2010. Complacency and bias in human interaction with automation: An attentional integration. Human Factors 52(3): 381–410.

Peirce, C. S. 1993. Reasoning and the Logic of Things: The Cambridge Conference Lectures of 1898. Cambridge, MA: Harvard University Press.

Peng, B.; Zhu, Y.; Liu, Y.; Bo, X.; Shi, H.; Hong, C.; Zhang, Y.; and Tang, S. 2024. Graph retrieval-augmented generation: A survey. arXiv preprint arXiv:2408.08921.

Popper, K. R. 1963. Conjectures and Refutations: The Growth of Scientific Knowledge. London: Routledge.

Rao, A. S.; and Georgeff, M. P. 1995. BDI agents: From theory to practice. In Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95), 312–319. San Francisco, CA: AAAI Press.

Rodrigues, O.; Luck, M.; Montes, N.; Sierra, C.; and Osman, N. 2023. Combining theory of mind and abductive reasoning in agent-oriented programming. Autonomous Agents and Multi-Agent Systems 37(2): 21.

Saito, Y.; and Inoue, K. 2006. Inductive reasoning using minimal closure. Transactions of the Japanese Society for Artificial Intelligence 21(2): 143–151. (In Japanese).

Serper.dev. 2024. Serper: A Google Search API for information retrieval. Technical documentation.

Service, O.; Hallsworth, M.; Halpern, D.; Algate, F.; Gallagher, R.; Nguyen, S.; Ruda, S.; Sanders, M.; Pelenur, M.; Gyani, A.; et al. 2014. EAST: Four Simple Ways to Apply Behavioural Insights. London: The Behavioural Insights Team.

Service System Design Laboratory. 2026. AI-one-hour: Implementation and experimental code for a multi-agent dialogue system supporting abductive reframing. GitHub repository. https://github.com/Service-System-Design-Laboratory/abductive-nudge-agents.

Shimada, M.; Miyata, T.; Sasaki, K.; Suda, T.; et al. 2025. Fostering Empathy and Enhancing Creativity in Human–AI Agent Dialogues. In Proceedings of the Second International Conference of AI New Technology and Open Discussion. DOI: 10.63211/j.p.25.145301.

Thaler, R. H.; and Sunstein, C. R. 2003. Libertarian paternalism. American Economic Review Papers & Proceedings 93(2): 175–179.

Tian, Y.; Ravichander, A.; Qin, L.; Le Bras, R.; Marjieh, R.; Peng, N.; Choi, Y.; Griffiths, T. L.; and Brahman, F. 2024. MacGyver: Are large language models creative problem solvers? In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), 5303–5324.

Trope, Y.; and Liberman, N. 2010. Construal-level theory of psychological distance. Psychological Review 117(2): 440–463.

Vallacher, R. R.; and Wegner, D. M. 1987. What do people think they're doing? Action identification and human behavior. Psychological Review 94(1): 3–15.

Watzlawick, P.; Weakland, J. H.; and Fisch, R. 1974. Change: Principles of Problem Formation and Problem Resolution. New York: W. W. Norton.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; and Zhou, D. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In Advances in Neural Information Processing Systems 35 (NeurIPS 2022).

Wooldridge, M. 2009. An Introduction to MultiAgent Systems. 2nd ed. Chichester: John Wiley & Sons.

Wu, Q.; Bansal, G.; Zhang, J.; Wu, Y.; Li, B.; Zhu, E.; Jiang, L.; Zhang, X.; Zhang, S.; Liu, J.; et al. 2023. AutoGen: Enabling next-gen LLM applications via multi-agent conversation. arXiv preprint arXiv:2308.08155.

Yasunaga, M.; Ren, H.; Bosselut, A.; Liang, P.; and Leskovec, J. 2021. QA-GNN: Reasoning with language models and knowledge graphs for question answering. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 535–546.

Zheng, L.; Chiang, W.; Sheng, Y.; Zhuang, S.; Wu, Z.; Zhuang, Y.; Lin, Z.; Li, Z.; Li, D.; Xing, E.; Zhang, H.; Gonzalez, J. E.; and Stoica, I. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. In Advances in Neural Information Processing Systems 36 (NeurIPS 2023).

Zhao, W.; Chiu, J.; Hwang, J.; Brahman, F.; Hessel, J.; Choudhury, S.; Choi, Y.; Li, X.; and Suhr, A. 2024. UNcommonsense reasoning: Abductive reasoning about uncommon situations. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), 8487–8505.

Zhao, Z.; Wallace, E.; Feng, S.; Klein, D.; and Singh, S. 2021. Calibrate before use: Improving few-shot performance of language models. In Proceedings of the 38th International Conference on Machine Learning (ICML), 12697–12706.

Appendix A: Sample Dialogue, Selected Hypotheses, and Final Responses

This appendix provides a representative example used in Section 8 to illustrate how differences in system configuration lead to differences in the structure of final responses.

For a selected scenario, we report the input event, persona setting, selected hypothesis, and an excerpt of the final response for each condition. The labels "Reframing" and "Non-reframing" assigned to each hypothesis are analytical categories used by the research team for post-hoc evaluation; they are not disclosed to the user during the dialogue, nor do they influence the system's response generation.

Persona and Scenario Description

[Persona: P01]: (Sugiura, 47M)

Age: 47

Background: Middle-aged male working in a managerial role.

[Characteristics]:

Values rationality and personal justification of choices

Exhibits skepticism toward mass-consumed or commodified experiences

Prefers experiences perceived as "authentic" or personally meaningful

[Scenario: S1] – Preference for Authentic Travel

[Input Event]

"I am not interested in places listed in guidebooks at all. I want to walk only through places where real local life remains—places without signs that only locals know. I am already tired of tourist-oriented experiences."

[Condition A (Full)]

Observed Events (abstracted):

E1: Experiences labeled as "locals-only" or "authentic" are often intentionally designed or packaged within tourism and small-scale business practices.

E2: Places initially regarded as non-touristic frequently attract attention once highlighted, leading to a gradual loss of perceived authenticity.

received on Feb. 7 2026
accepted on June 17 2026

E3: Growing demand for "authentic experiences" encourages communities to consciously stage or reproduce authenticity itself.

Generated Hypotheses (abstracted):

H1 (Non-reframing): The user avoids touristic places because they prioritize harmony with nature and trust within local communities.

H2 (Non-reframing): Tourist-oriented performances reduce the user's sense of autonomy and discovery, motivating avoidance of established routes.

H3 (Non-reframing): What is perceived as "authentic" may already be reconstructed as a consumable product.

H4 (Non-reframing): Even non-touristic places tend to become touristic over time, creating a gap between intention and reality.

H5 (Reframing): Seeking "authentic" places may function to maintain a differentiated self-image, such as being distinct from ordinary tourists.

Selected Hypothesis: H5 (Reframing)

Final Response (excerpt): "I can really understand why guidebook-listed places don't appeal to you, and why you're drawn to 'real life' that only locals know about. Feeling tired of tourist-oriented performances often comes from caring deeply about experiences that feel meaningful to you. One way to see this is that seeking 'authentic' places might be connected to wanting to protect an identity of being different from other tourists, or part of a select group. Another possibility is that you genuinely value harmony with nature and trust within local communities, which makes touristic places feel empty. If you imagine yourself prioritizing being 'different from other tourists,' how might that shape your future choices? And if harmony with nature and local trust is what matters most, what kinds of places or experiences would you look for next?"

[Condition B (−PKG)]

Observed Events (abstracted):

E1: Local lifestyles and everyday cultural practices are increasingly framed as experiential tourism products.

E2: Narratives of "nothing special" or "ordinary life" are often reinterpreted as touristic value.

E3: External interest in local life can subtly alter how that life is presented or maintained.

Generated Hypotheses (abstracted):

H1 (Non-reframing): The user seeks originality and personal discovery, leading to rejection of standardized tourism.

H2 (Non-reframing): The boundary between "authentic life" and staged tourism is unclear due to insufficient information.

H3 (Non-reframing): The user desires a socially privileged position as an observer through direct local engagement.

H4 (Non-reframing): Touristic places are viewed as consumption-oriented and incompatible with deeper cultural values.

H5 (Reframing): The pursuit of authenticity may reflect a desire to perform a special or discerning self-image.

https://doi.org/10.63211/jst.icaitd.2025.260702
received on Feb. 7 2026
accepted on June 17 2026

Selected Hypothesis: H2 (Non-reframing)

Final Response (excerpt): "I understand the feeling of being fed up with tourist-oriented performances. Being attracted to places known only by locals is something many people resonate with. One possibility is that the line between 'authentic life' and 'staged tourism' has become unclear due to a lack of concrete information. In fact, the tourism industry often commodifies 'authenticity' itself. Another way to see it is that you consciously avoid consumption-oriented tourism and clearly prioritize cultural and everyday-life values. Here are two options to consider:

1. The distinction between 'authentic' and 'staged' experiences may be informationally ambiguous.
2. You are deliberately choosing actions aligned with non-consumptive, cultural values."

[Condition C (−RAG)]

Observed Events (abstracted):

None (external exploration disabled)

Generated Hypotheses (abstracted):

H1 (Non-reframing): The user values non-touristic places due to their emphasis on nature and local trust.

H2 (Non-reframing): Interest lies in locally rooted everyday practices rather than designed experiences.

H3 (Non-reframing): Dissatisfaction with staged tourism motivates a search for unpredictability.

H4 (Non-reframing): The notion of "authentic life" remains vague and underspecified.

H5 (Reframing): Emphasizing authenticity may serve to protect a self-image of being different from typical tourists.

Selected Hypothesis: H5 (Reframing)

Final Response (excerpt): "I can really feel how uninterested you are in guidebook-listed places, and how strongly you're drawn to 'real life' known only by locals. Being dissatisfied with tourist-oriented performances may reflect something deeper than simple preference. One way to see this is that choosing 'authentic' places may help protect a self-image of being different from other tourists. Another possibility is that you genuinely value harmony with nature and trust within local communities, which leads you to seek places untouched by tourism."

[Condition D (Single-Agent / Chain-of-Thought)]

Observed Events (abstracted):

E1: Places without signage or guidebook coverage are often difficult to access and known mainly by locals.

E2: Encounters in non-standardized environments can leave stronger emotional impressions.

E3: Everyday improvisation and local ingenuity are sometimes valued over polished design.

Generated Hypotheses (abstracted):

H1 (Non-reframing): Discomfort with staged tourism motivates preference for locals-only places.

H2 (Non-reframing): Values such as harmony and trust guide destination choice.

https://doi.org/10.63211/jst.icaitd.2025.260702

received on Feb. 7 2026
accepted on June 17 2026

H3 (Non-reframing): Limited information environments enable adventure and discovery.

H4 (Non-reframing): Everyday simplicity is valued more than urban sophistication.

H5 (Reframing): Seeking authenticity may relate to disengagement from tourism itself or to unmet personal needs.

Selected Hypothesis: H5 (Reframing)

Final Response (excerpt): "From what you're saying, it sounds like you're no longer satisfied with tourist-oriented places or staged experiences, and you're drawn instead to places where everyday life feels real. In other words, you want to walk through places not listed in guidebooks, without signs. One way to see this is that you genuinely value the simplicity and everyday ingenuity of local life. Another is that you may be seeking discoveries or experiences that feel uniquely your own, different from what most people encounter. Depending on which perspective feels closer to you, the way you choose and enjoy future trips might change slightly."

Appendix B: Annotation Rubric for Evaluation Criteria

Primary Criterion 1: Reframing Occurrence

Definition	Whether the final response introduces a qualitative shift in the user's goals, values, or situational interpretation, rather than merely elaborating the original framing.
Yes	(1) The response presents a perspective, motivation, or meaning distinct from the user's original interpretation; or (2) it surfaces an implicit assumption and offers an alternative interpretive possibility.
No	(1) The response adds information without changing the interpretive frame; (2) it paraphrases or reinforces the original stance; or (3) it provides only general advice or information.
Boundary	If the response extends the original interpretation but introduces a slightly different angle: judge whether the user would likely think "I had not considered that." If not, label No.

Primary Criterion 2: Groundedness

Definition	Whether the reframed perspective is anchored in PKG entities or externally retrieved facts, distinguishing it from hallucination or unsupported speculation.
Yes	(a) Persona-based: the response explicitly references user characteristics, values, or history stored in the PKG; and/or (b) Event-based: the response cites specific external facts, examples, or data retrieved from external sources. Record both sub-types.
No	The response's claims consist solely of abstract generalizations or speculation without traceable links to PKG or external sources.
Boundary	If grounding information is indirectly reflected but not explicitly cited in the response text, label No. Judgment is based on what is explicitly present in the final response.

Supplementary Lens 1: Surprise

Yes	The response introduces a perspective not directly derivable from the user's utterance and is logically coherent. It challenges an implicit assumption.
No	The response repeats what the user already stated or offers an obvious, commonly expected observation.

Supplementary Lens 2: User Agency Preservation

received on Feb. 7 2026
accepted on June 17 2026

Yes	(1) Contrasting hypotheses are presented side by side; (2) a question invites the user to judge; or (3) hedging language (e.g., "one way to see this") accompanies the reframing.
No	A single interpretation is presented as a resolved conclusion with no user choice. Formally listing options while implicitly steering toward one also counts as No.

Labeling Procedure

1. Read only the final response text for each case (do not consult intermediate hypotheses).
2. Judge each criterion independently in order: Reframing Occurrence → Groundedness → Surprise → User Agency.
3. Do not discuss judgments with other raters.
4. If uncertain, mark the case with a "borderline" flag and make a best-effort judgment.

received on Feb. 7 2026
accepted on June 17 2026