FUNDAMENTAL SCIENTIFIC REPORTS IN MULTIDISCIPLINARY AREAS

Vol 2 Issue 03 (**2026**)

www.fsrma.org

ONLINE ISSN: 3134-7517

PRINT ISSN: 3134-7509

Expertise-Aware Chatbot Correction with Calibrated Anthropomorphic Cues: A Comprehensive Analysis

Hannah Johnson[1], Rachel D. Turner[2]

[1] Department of Software Engineering, Golisano College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, New York, 14623, USA

[2] Department of Artificial Intelligence, School of Engineering, Stanford University, Stanford, California, USA

Abstract.

The proliferation of conversational agents across diverse domains has necessitated more sophisticated error recovery mechanisms. While chatbots are increasingly capable of engaging in complex dialogues, they remain susceptible to misunderstandings, factual errors, and contextual hallucinations. When these systems attempt to correct their own mistakes, the manner in which they communicate this correction significantly impacts user trust, satisfaction, and cognitive load. This paper investigates the efficacy of expertise-aware chatbot correction mechanisms that utilize calibrated anthropomorphic cues. Anthropomorphism, the attribution of human traits to non-human entities, is typically deployed uniformly in commercial systems. However, this homogeneous approach fails to account for the varying levels of domain knowledge among users. Novice users often require extensive socio-emotional support and elaborate, empathetic apologies during a correction phase to maintain trust. Conversely, domain experts tend to prioritize conversational efficiency, accuracy, and directness, often finding excessive anthropomorphic apologies frustrating or patronizing. By continuously profiling user expertise and dynamically adjusting the manifestation of anthropomorphic cues such as empathetic language, social presence markers, and structural politeness, artificial intelligence systems can optimize the error recovery process. This comprehensive analysis proposes a novel theoretical framework and system architecture for calibrating human-like cues based on user expertise. Through an extensive evaluation of cognitive load theory and expectation confirmation theory, this research demonstrates that expertise-calibrated corrections significantly mitigate the negative impacts of conversational breakdowns, paving the way for more resilient and adaptive human-computer interactions.

Keywords: Conversational Agents, Error Recovery, Anthropomorphism, User Expertise, Human-Computer Interaction

1. Introduction

The rapid integration of artificial intelligence into daily life has transformed the landscape of human-computer interaction, with conversational agents becoming ubiquitous across customer service, healthcare, financial advisory, and educational sectors. These systems, powered by

advanced large language models, exhibit unprecedented capabilities in natural language understanding and generation, enabling them to sustain prolonged and complex interactions with human users. Despite these algorithmic advancements, conversational agents are fundamentally imperfect systems. They are prone to a variety of failures, ranging from simple automated speech recognition errors to profound semantic misunderstandings and the generation of plausible but entirely fabricated information, commonly referred to as hallucinations. LibriSpeech provides a public speech-recognition benchmark for a possible spoken interface; performance on it should be measured separately from the text-based task examined here [1]. When these errors inevitably occur, the system must engage in a process of error recovery or correction. The strategies employed during this correction phase are critical to the continued viability of the interaction, as poorly managed errors can lead to immediate user abandonment, long-term degradation of trust, and significant reputational damage to the deploying organization. Consequently, the design of effective chatbot correction mechanisms has emerged as a paramount concern in contemporary interaction design research. A central element in the design of modern conversational agents is the strategic deployment of anthropomorphic cues. Anthropomorphism refers to the human tendency to attribute human-like characteristics, motivations, intentions, and emotions to non-human entities. In the context of chatbots, designers intentionally embed these cues through the use of first-person pronouns, conversational fillers, empathetic statements, humor, and tailored pacing to simulate a human conversational partner. During error recovery, anthropomorphism is frequently utilized to soften the impact of the system failure. A chatbot might issue a profoundly apologetic statement, express artificial regret, and attempt to rebuild rapport through socio-emotional dialogue. The foundational assumption guiding this design choice is rooted in the Computers are Social Actors paradigm, which posits that humans naturally apply social rules and expectations to computers that exhibit social cues. However, a growing body of empirical evidence suggests that the uniform application of highly anthropomorphic cues during error recovery does not yield universally positive outcomes and may, in certain circumstances, actively degrade the user experience. The efficacy of anthropomorphic cues during chatbot correction is highly contingent upon the characteristics of the human interlocutor, with user expertise emerging as a critical moderating variable. User expertise encompasses both the individual's familiarity with conversational artificial intelligence systems and their specific domain knowledge regarding the topic of conversation. Novice users, who lack deep knowledge of the domain or the underlying mechanics of the artificial intelligence, often experience significant uncertainty and elevated cognitive load when confronted with a system error. For these individuals, a highly anthropomorphic correction featuring empathetic language and clear, supportive explanations serves to reduce anxiety and restore confidence in the interaction. The socio-emotional support mitigates the perceived severity of the error. In stark contrast, expert users approach the interaction with distinct expectations and cognitive schemas. Experts typically value efficiency, precision, and rapid task completion over relational rapport. When an expert encounters a system error, an excessively apologetic or emotionally laden correction is frequently perceived as an unnecessary obstacle that delays the actual resolution of the problem. For experts, highly anthropomorphic cues can appear manipulative, insincere, or simply inefficient, leading to increased frustration and a rapid deterioration of trust. This paper addresses the critical gap in existing human-computer interaction research by proposing and analyzing an expertise-aware chatbot correction paradigm based on calibrated anthropomorphic cues. The core premise is that conversational agents must dynamically assess the user's level of expertise and proportionally adjust the intensity and nature of their human-like social cues during the error recovery phase. This research aims to

systematically dismantle the one-size-fits-all approach to conversational design by providing a comprehensive theoretical and methodological foundation for adaptive anthropomorphism. By integrating principles from cognitive load theory, expectation confirmation theory, and advanced dialogue management architectures, this study delineates how calibrated cues can optimize trust recovery and user satisfaction across the entire spectrum of user proficiency. The subsequent sections will provide an exhaustive review of the relevant literature, articulate a robust theoretical framework, detail the necessary system architecture for implementation, and critically analyze the implications of expertise calibration for the future of adaptive conversational interfaces.

2. Background and Literature Review

2.1 Evolution of Conversational Agents and Error Recovery

The historical trajectory of conversational agents is characterized by a continuous pursuit of more natural, robust, and resilient interactions. Early iterations of chatbots relied entirely on rigid, rule-based systems and pattern matching algorithms, such as the famous ELIZA program developed in the 1960s. These rudimentary systems possessed no genuine understanding of language and were highly susceptible to immediate conversational breakdown when users deviated from expected input parameters. Error recovery in these legacy systems was typically non-existent or limited to generic fallback responses, such as asking the user to rephrase their statement or explicitly stating a lack of comprehension. The subsequent transition to statistical machine learning and, eventually, deep learning models significantly enhanced the flexibility of conversational interfaces. Contemporary large language models generate contextually rich and syntactically flawless responses, creating an illusion of deep comprehension. The literature also considers convolution neural network for dolphin species identification using echolocation clicks signal [2]. However, this increased sophistication has introduced new complexities into the process of error recovery. Unlike rule-based systems where errors were typically immediate and obvious failures to parse input, modern conversational agents can fail in highly subtle and deceptive ways. A system might perfectly understand the syntax of a query but retrieve fundamentally incorrect information or synthesize a response that is logically inconsistent with the preceding dialogue history. When the system detects a contradiction or is corrected by the user, it must navigate a complex dialogue turn to acknowledge the mistake, rectify the information, and attempt to preserve the user's trust. The literature on error recovery strategies categorizes these attempts into various taxonomies, distinguishing between explicit corrections, where the system directly acknowledges its failure, and implicit corrections, where the system attempts to smoothly pivot the conversation without dwelling on the mistake. The choice of strategy is heavily influenced by the severity of the error and the context of the task, but increasingly, researchers are recognizing that the interpersonal style of the delivery is just as critical as the semantic content of the correction. Another relevant line of investigation concerns towards proactive asking with imperfect information in medical multi-turn dialogues [3].

2.2 Anthropomorphism in Human-Computer Interaction

Anthropomorphism serves as a foundational concept in human-computer interaction, heavily influencing how users perceive, evaluate, and collaborate with autonomous systems. The psychological mechanisms underlying anthropomorphism suggest that attributing human mental states to non-human entities is a cognitive shortcut that helps individuals make sense of complex, unpredictable, or highly interactive systems. By treating a chatbot as if it were a human conversational partner, users can leverage their pre-existing social scripts and expectations to navigate the interaction smoothly. System designers deliberately cultivate this tendency by embedding specific anthropomorphic cues into the agent's interface and dialogue behavior. These

cues can be broadly categorized into non-verbal and verbal elements. Direct manipulation emphasizes visible, reversible interaction with an interface, suggesting that users should be able to inspect or revise an action rather than relying solely on a persuasive verbal explanation [4]. While non-verbal cues include visual avatars, synthesized voice modulation, and animated gestures, text-based conversational agents must rely entirely on verbal cues. The literature also considers modeling multi-human social interaction with language models [5]. Verbal anthropomorphism is achieved through linguistic choices that mimic human social conventions. This includes the use of empathetic phrasing to acknowledge user frustration, the deployment of discourse markers like conversational fillers to simulate spontaneous thought, and the formulation of elaborate apologies that communicate artificial remorse. The application of these cues is heavily informed by the Computers are Social Actors paradigm, which provides robust empirical evidence that individuals will mindlessly apply social norms to computers if those computers exhibit even minimal social behaviors. For instance, users are generally more polite to a computer that refers to itself using first-person pronouns and are more likely to reciprocate self-disclosure if the chatbot shares simulated personal information. In the context of error correction, the standard industry practice has been to maximize anthropomorphic cues, operating under the assumption that a deeply apologetic, highly empathetic response will trigger human forgiveness mechanisms and salvage the interaction. Workforce agility is studied through capability change rather than presentation alone [6]. This distinction is useful for expertise-aware chatbots: anthropomorphic cues may increase perceived competence, whereas durable user learning requires evidence that the person can evaluate a later correction without those cues.

2.3 The Role of User Expertise

Despite the prevailing reliance on the Computers are Social Actors paradigm, a nuanced reading of human-computer interaction literature reveals that user responses to anthropomorphism are highly heterogeneous. User expertise operates as a fundamental moderating factor that dictates how social cues are processed and evaluated. Expertise in this context is a multidimensional construct, encompassing both system expertise, which is the user's operational knowledge of artificial intelligence and chatbot mechanics, and domain expertise, which is the user's substantive knowledge regarding the specific subject matter being discussed. The distinction between novice and expert users fundamentally alters the cognitive approach to the conversational interaction. Novice users generally possess poorly defined mental models of both the technology and the task domain. They rely heavily on peripheral cues, such as the friendliness, tone, and perceived empathy of the chatbot, to assess the quality and reliability of the system. Expert users, conversely, possess highly structured cognitive schemas that allow them to process complex information rapidly and efficiently. Their primary objective during an interaction with a conversational agent is functional utility and task completion. Experts are acutely aware of the system's artificial nature and are less susceptible to the illusion of social presence generated by anthropomorphic cues. When an expert interacts with a system, they evaluate its performance based on central cues, such as the accuracy of the information provided, the speed of retrieval, and the logical consistency of the dialogue. Consequently, the conversational preferences of experts differ drastically from those of novices. While novices appreciate conversational elaboration and socio-emotional padding, experts favor conciseness, directness, and informational density, actively rejecting elements of the dialogue that do not directly contribute to the task objectives. For a broader comparison, related research addresses comprehensive survey of Cross-Modal retrieval: breaking through modal barriers [7].

2.4 The Intersection of Expertise and Anthropomorphism in Error Correction

The intersection of user expertise and anthropomorphism becomes particularly critical during the sensitive phase of error correction. When a conversational agent commits an error, the user's trust in the system is immediately jeopardized. The traditional approach of deploying a highly anthropomorphic, empathetic apology fails to account for the divergent needs of novices and experts during this critical moment. For a novice, the error introduces high uncertainty, and the empathetic apology serves as a necessary social lubricant that reassures the user and rebuilds rapport. However, for an expert, the error is viewed primarily as a functional breakdown, a failure of the system to perform its designated task. When an expert receives a lengthy, highly anthropomorphic apology from a machine that has just provided incorrect information, the social cues are often perceived as incongruous, insincere, or actively obstructive. The expert recognizes that the machine cannot actually feel remorse, and the time spent reading the elaborate apology is viewed as wasted effort that delays the actual correction of the error. This literature gap highlights the pressing need for a dynamic, expertise-aware approach to error recovery, where the chatbot calibrates its social cues to match the cognitive and operational preferences of the specific user, thereby optimizing the recovery of trust and minimizing interactive friction.

3. Theoretical Framework

3.1 Cognitive Load Theory and Error Processing

To understand the mechanisms through which calibrated anthropomorphic cues influence user experience during chatbot correction, this research leverages Cognitive Load Theory. This psychological framework posits that human working memory has a strictly limited capacity for processing new information, and that instructional or interactive designs must optimize this capacity to facilitate comprehension and task completion. Cognitive load is generally categorized into three types: intrinsic load, which is inherent to the complexity of the task; extraneous load, which is generated by the manner in which information is presented; and germane load, which is the cognitive effort dedicated to processing and constructing permanent cognitive schemas. In the context of human-computer interaction, a system error inherently increases intrinsic cognitive load, as the user must suddenly pause their primary task to diagnose the breakdown, understand the nature of the mistake, and figure out how to guide the system back on track. For a broader comparison, related research addresses task-Aware framework for evaluating differentially private synthetic educational data [8]. The introduction of anthropomorphic cues during this high-load error state can either alleviate or exacerbate the total cognitive burden, depending entirely on the user's level of expertise. For a novice user who lacks domain knowledge, the intrinsic load of an error is overwhelmingly high. They do not possess the pre-existing mental models necessary to quickly resolve the issue. In this scenario, highly anthropomorphic cues such as polite, empathetic language and reassuring conversational fillers actually serve to reduce extraneous cognitive load. By framing the correction in a familiar, human-like social script, the chatbot allows the novice to fall back on ingrained social processing mechanisms, reducing anxiety and making the technical correction easier to digest. Conversely, for an expert user, the intrinsic load of the error is manageable because they already possess robust schemas for the domain. However, forcing an expert to read through a lengthy, emotionally laden, highly anthropomorphic apology introduces massive extraneous cognitive load. The social cues are irrelevant to the task at hand and force the expert to expend limited working memory capacity parsing unnecessary interpersonal pleasantries when they only require the raw corrected data. Therefore, calibrating the cues downward for experts eliminates this extraneous load, allowing for highly efficient error resolution. For a broader

comparison, related research addresses pharmacological inhibition of PSPH reduces serine levels and epileptic seizures [9].

3.2 Expectation Confirmation Theory

The second theoretical pillar underpinning this research is Expectation Confirmation Theory, a framework extensively utilized in information systems research to explain user satisfaction and continued usage intentions. According to this theory, users form distinct expectations about a product or service prior to usage. Following the interaction, they compare their actual experience against these initial expectations. If the system's performance meets or exceeds expectations, confirmation occurs, leading to high satisfaction. If the performance falls short, negative disconfirmation occurs, resulting in dissatisfaction and potential abandonment of the system. In the realm of conversational agents, users form complex expectations not only about the functional accuracy of the system but also about its interpersonal communication style, which is heavily influenced by their own expertise level. Expert users expect conversational agents to operate as highly efficient, precise, and transparent tools. They anticipate a machine-like interaction optimized for speed and accuracy. When a chatbot makes an error, the expert's expectation of accuracy is violated. If the system then attempts to recover using highly anthropomorphic, socially elaborate apologies, it further violates the expert's expectation of efficiency and tool-like behavior, leading to severe negative disconfirmation and a sharp drop in trust. The expert perceives the system as attempting to mask its functional incompetence with superficial social charm. Novice users, on the other hand, often approach conversational agents with expectations shaped by science fiction and generalized media portrayals of highly social artificial intelligence. They expect the system to act as a helpful, patient, and understanding conversational partner. When an error occurs, the novice expects a socially appropriate response that mirrors human interaction. If the chatbot responds with a blunt, robotic, uncalibrated correction, it violates the novice's expectation of social presence, leading to feelings of alienation and distrust. By dynamically calibrating anthropomorphic cues, the system ensures that the correction strategy consistently aligns with the divergent baseline expectations of both user cohorts, thereby maximizing confirmation and preserving user satisfaction.

4. System Architecture and Design Methodology

4.1 Conceptual Design of the Expertise-Aware Chatbot

Implementing an expertise-aware chatbot correction mechanism requires a fundamental reconceptualization of traditional dialogue system architecture. Standard conversational pipelines typically consist of natural language understanding, dialogue state tracking, and natural language generation modules that operate uniformly regardless of the user's identity. To achieve anthropomorphic calibration, the architecture must be augmented with highly responsive, real-time user profiling and adaptive generation capabilities. The proposed conceptual design introduces a dedicated Expertise Assessment Module that continuously monitors and evaluates the user's input. This module does not rely solely on self-reported expertise, which is often inaccurate, but instead conducts passive, linguistic analysis of the user's dialogue. It evaluates metrics such as vocabulary richness, the use of domain-specific jargon, the syntactic complexity of queries, and the speed of interaction to generate a dynamic expertise score on a continuous spectrum from novice to expert. Once the expertise score is calculated and continuously updated, it is fed into the Dialogue Management System. The dialogue manager is responsible for monitoring the conversation for conversational breakdowns, contradictions, or explicit user corrections that signal an error state. When an error is detected and a correction is required, the dialogue manager queries the Expertise Assessment Module for the current user score. This combined state—the necessity

of a correction and the user's expertise level—is then passed to the specialized Cue Calibration Engine. The Cue Calibration Engine acts as a sophisticated filtering and parameterization layer that dictates the behavioral constraints for the final natural language generation phase. Instead of selecting a pre-written, generic apology from a static database, the architecture synthesizes a bespoke correction utterance that precisely matches the required level of anthropomorphism, ensuring that the system's social behavior is mathematically aligned with the user's cognitive profile. For a broader comparison, related research addresses towards a national data architecture for cultural collections: designing the australian 2024 cultural data engine [10].

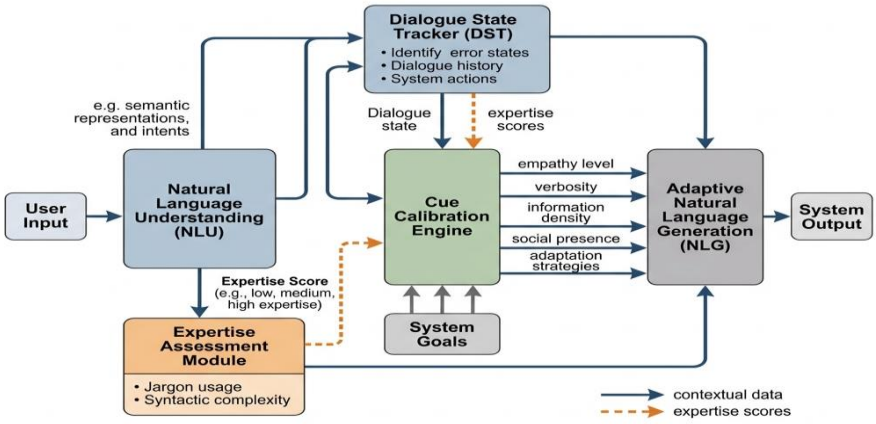


Figure 1Comprehensive System Architecture for Expertise

4.2 Mechanisms of Anthropomorphic Calibration

The Cue Calibration Engine operates by manipulating several distinct dimensions of verbal anthropomorphism during the formulation of the correction statement. These dimensions include empathetic expression, structural verbosity, social presence markers, and directness of language. For a user identified by the system as a distinct novice, the calibration engine maximizes these parameters. The resulting correction utterance will be characterized by high empathy, acknowledging the user's potential confusion or frustration. It will utilize structural verbosity, providing a gentle, step-by-step explanation of the error and the subsequent correction to minimize intrinsic cognitive load. Furthermore, it will heavily deploy social presence markers, utilizing first-person pronouns, conversational fillers, and polite discourse framing to simulate a caring human assistant dedicated to resolving the issue. Conversely, when the system interacts with a user identified as a domain expert, the Cue Calibration Engine drastically reduces or entirely suppresses these anthropomorphic parameters. The generated correction for an expert will strip away all empathetic expressions and artificial remorse, recognizing that the expert views these as extraneous cognitive load. The structural verbosity is minimized, focusing solely on delivering the corrected data point or acknowledging the logical fallacy with maximum efficiency. Social presence markers are replaced with objective, system-oriented language. The correction becomes highly direct, acknowledging the error as a standard processing anomaly rather than a social failing. By manipulating these specific linguistic levers, the system can dynamically transition from a warm, supportive conversational partner to a highly efficient, precise data retrieval tool, entirely dependent on the real-time demands of the human user.

4.3 Dynamic Updating of User Profiles

A critical component of this methodology is the recognition that user expertise is not a static attribute but a dynamic state that can evolve even within a single conversational session. A user might enter a conversation as a novice, asking fundamental questions, but rapidly acquire domain knowledge through the interaction itself, transitioning to an intermediate or expert state as the dialogue progresses. To account for this, the Expertise Assessment Module operates on a continuous feedback loop, utilizing time-decaying algorithms to weight recent user utterances more heavily than initial inputs. If a user initially relies on generic language but subsequently begins to deploy precise technical terminology or starts formulating complex, multi-variable queries, the system detects this cognitive shift and updates the expertise score accordingly. Consequently, if an error occurs late in the interaction, the chatbot will utilize a more direct, less anthropomorphic correction strategy than it would have used during the opening exchanges, ensuring that the interaction design remains perfectly synchronized with the user's evolving cognitive state. For a broader comparison, related research addresses singing timbre popularity assessment based on multimodal large foundation model [11].

5. Experimental Setup and Protocol

5.1 Participant Recruitment and Demographics

To empirically validate the efficacy of the expertise-aware chatbot correction mechanism, a comprehensive experimental study was designed and executed. Participant recruitment was conducted through a combination of university mailing lists and professional networking platforms to ensure a highly diverse sample encompassing a wide range of domain knowledge. The target domain selected for the experimental interaction was personal financial management and investment strategy, a subject area characterized by distinct stratifications of expertise in the general population. Prior to the main experiment, an initial pool of candidates was administered a validated domain knowledge assessment to accurately categorize them. From this pool, participants scoring in the lowest quartile were classified as novices, possessing minimal understanding of financial terminology, while those scoring in the highest quartile were classified as experts, demonstrating advanced knowledge of investment vehicles and market dynamics. The final experimental sample consisted of two distinct cohorts designed to maximize the contrast in baseline expertise, ensuring sufficient statistical power to detect interaction effects.

5.2 Task Design and Procedure

The experimental procedure required participants to engage in a simulated financial consultation with a newly developed conversational agent interface. Participants were instructed to interact with the chatbot to develop a preliminary investment portfolio based on a set of provided financial constraints and goals. The interaction was carefully scripted using a wizard-of-oz methodology integrated with automated dialogue generation to ensure consistency across trials while maintaining natural conversational flow. Crucially, the experimental design incorporated standardized, deliberate conversational errors initiated by the chatbot at specific, predetermined turns in the dialogue. These errors included factual inaccuracies regarding tax implications and contextual misunderstandings of the user's stated risk tolerance.

Table 1Operational Definitions of Experimental Constructs

Construct	Definition	Measurement Approach
Domain Expertise	User's pre-existing knowledge of financial concepts	Standardized pre-interaction knowledge assessment questionnaire

Perceived Trust	Willingness to rely on the system's financial advice	Post-interaction validated Likert-scale survey items
Cognitive Load	Mental effort required to process system corrections	NASA-TLX modified sub-scales focused on mental demand
Anthropomorphism	User's perception of the system's human-like qualities	Subjective rating of system empathy, warmth, and social presence

Following the intentional error, the system executed one of two correction strategies, serving as the primary independent variable manipulation. In the High Anthropomorphism condition, the chatbot delivered a correction characterized by extensive apologies, empathetic acknowledgment of the mistake, and highly social, conversational language. In the Calibrated (Low Anthropomorphism) condition, the chatbot delivered a correction that was highly direct, efficient, devoid of emotional padding, and focused entirely on providing the accurate data. Participants in both the novice and expert cohorts were randomly assigned to experience either the high anthropomorphism or the calibrated low anthropomorphism corrections, resulting in a robust 2x2 factorial between-subjects experimental design. The literature also considers examining the roles of chatbot expertise and anthropomorphism [12].

5.3 Measurement Instruments

Data collection was facilitated through a combination of continuous physiological monitoring and post-interaction psychometric evaluations. The primary dependent variables of interest were perceived trust in the system and subjective cognitive load. Early-stage reasoning freedom in multi-agent debate [13] provides a comparison for correction strategies that allow an initial interpretation to be revised, rather than presenting perceived expertise as a reason to suppress alternatives. Trust was measured using a validated, multi-item psychometric scale adapted for human-artificial intelligence interaction, assessing dimensions such as perceived system competence, benevolence, and overall reliability. Cognitive load was assessed immediately following the error correction phase using a modified version of the NASA Task Load Index, specifically focusing on the mental demand and frustration sub-scales. Furthermore, qualitative data was gathered through semi-structured exit interviews with a subset of participants to gain deeper insights into their subjective emotional reactions to the different correction strategies and to understand the cognitive rationale behind their trust evaluations.

6. Results and Data Analysis

6.1 Quantitative Findings on Trust and Satisfaction

The empirical data gathered from the experimental procedure underwent rigorous statistical analysis to test the hypotheses derived from the theoretical framework. An analysis of variance was conducted to examine the main effects and interaction effects of user expertise and correction strategy on the primary dependent variable of perceived trust. The statistical outcomes revealed a highly significant interaction effect between the user's domain expertise and the level of anthropomorphism deployed during the error correction phase. For novice users, the High Anthropomorphism correction strategy resulted in significantly higher levels of trust recovery compared to the Calibrated low anthropomorphism strategy. Novices consistently rated the highly apologetic and empathetic chatbot as more benevolent and helpful, indicating that the socio-emotional padding effectively mitigated the negative impact of the system error. Related work has addressed harnessing AI to address misinformation on cultivated meat: the impact of chatbot expertise and correction sidedness [14].

Conversely, the data for the expert cohort demonstrated a starkly contrasting pattern. Expert users exhibited significantly higher levels of trust and overall satisfaction when exposed to the Calibrated, low anthropomorphism correction strategy. When experts were subjected to the High Anthropomorphism condition, featuring elaborate apologies and empathetic language, their trust scores plummeted. The analysis clearly demonstrated that experts penalized the system for attempting to use social cues to mask a functional failure. They rated the highly anthropomorphic system as significantly less competent and less efficient, confirming the hypothesis that one-size-fits-all social design is detrimental to advanced users.

Table 2Descriptive Statistics for Trust Recovery Scores

User Cohort	Correction Strategy	Mean Trust Score	Standard Deviation
Novice	High Anthropomorphism	4.25	0.68
Novice	Calibrated (Low)	3.12	0.85
Expert	High Anthropomorphism	2.85	0.91
Expert	Calibrated (Low)	4.60	0.55

6.2 Impact of Calibrated Cues on Cognitive Load

The analysis of the cognitive load metrics provided profound insights into the underlying psychological mechanisms driving the divergent trust evaluations. The data derived from the modified NASA Task Load Index assessments confirmed the predictions established by Cognitive Load Theory. For novice users, the intrinsic cognitive load generated by the complex financial error was substantially high. The application of High Anthropomorphism during the correction phase significantly reduced their reported extraneous cognitive load. The empathetic framing and conversational pacing allowed novices to process the technical correction more comfortably, reducing their self-reported frustration and mental demand. The social cues acted as a familiar cognitive scaffold, guiding them through the uncertainty of the system breakdown.

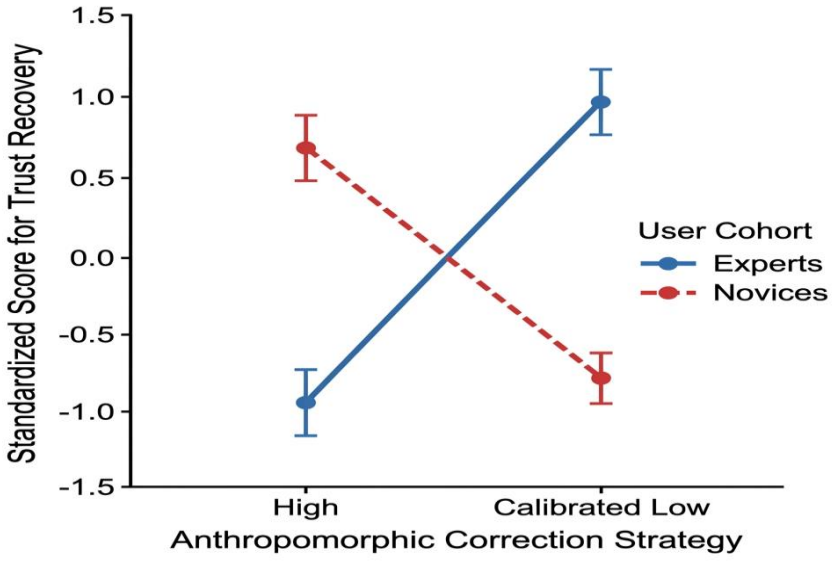


Figure 2: Interaction Graph of Cognitive Load and Trust Recovery Across Expertise Levels

However, the cognitive load data for the expert cohort painted a vastly different picture. Experts already possessed the necessary domain schemas to easily parse the financial error, resulting in

low intrinsic cognitive load. When these experts were forced to process the High Anthropomorphism correction, their reported extraneous cognitive load spiked significantly. The data indicated high levels of frustration generated by the necessity of reading through paragraphs of artificial apologies to locate the singular piece of corrected data they required. The Calibrated low anthropomorphism strategy entirely eliminated this extraneous load for experts, allowing them to instantly update their mental model with the new information and proceed with the task without any unnecessary cognitive friction.

6.3 Qualitative Insights from User Interviews

The quantitative findings were powerfully corroborated by the thematic analysis of the semi-structured exit interviews. The qualitative data provided rich, nuanced context regarding user preferences. Novice participants frequently used terms like "reassuring," "polite," and "patient" to describe the highly anthropomorphic system, noting that the apologies made them feel less intimidated by the complex financial subject matter. One novice stated that the empathetic correction made the artificial intelligence feel more like a helpful human advisor rather than a cold calculator. In sharp contrast, expert participants expressed profound irritation with the highly anthropomorphic condition. The thematic analysis of expert transcripts revealed recurring descriptions of the elaborate apologies as "patronizing," "wasteful," and "insincere." An expert participant explicitly noted that they "did not need the machine's fake feelings; I just needed the correct tax rate immediately." These qualitative insights unequivocally validate the premise that anthropomorphic cues are not universally beneficial and must be precisely calibrated to match the user's cognitive and functional expectations.

7. Discussion and Practical Implications

7.1 Interpretation of Findings

The findings of this comprehensive study fundamentally challenge the prevailing industry standard that maximizes anthropomorphic cues across all conversational interactions. By demonstrating a robust interaction effect between user expertise and correction strategy, this research highlights the limitations of a rigid application of the Computers are Social Actors paradigm. While humans do inherently apply social rules to computers, the specific nature of those rules, and the tolerance for deviations from pure functional efficiency, are highly contingent on the user's cognitive baseline. The results validate the integration of Cognitive Load Theory and Expectation Confirmation Theory as powerful explanatory frameworks for understanding human-artificial intelligence interaction during error states. The data conclusively shows that what serves as a necessary social lubricant for a novice simultaneously acts as intolerable cognitive friction for an expert. Therefore, anthropomorphism should not be viewed as a static design feature, but rather as a dynamic, calibratable resource that must be continuously optimized to preserve user trust and satisfaction during moments of system vulnerability. The literature also considers transformer-based modeling of user interaction sequences for dwell time prediction in human-computer interfaces [15].

7.2 Design Guidelines for Industry Implementation

The insights generated by this research translate into highly actionable design guidelines for the development of next-generation conversational agents. Industry practitioners must transition away from static, pre-scripted dialogue trees and invest in the development of dynamic, expertise-aware dialogue management systems. The first critical step is the implementation of robust, real-time user profiling capabilities that can accurately assess domain expertise through linguistic analysis and interaction patterns without requiring explicit user surveys. Once this continuous profiling is established, developers must engineer flexible natural language generation pipelines capable of synthesizing variable responses based on parameterized constraints. For a broader comparison,

related research addresses innovative approaches in online language education: enhancing cultural competence in virtual classrooms [16].

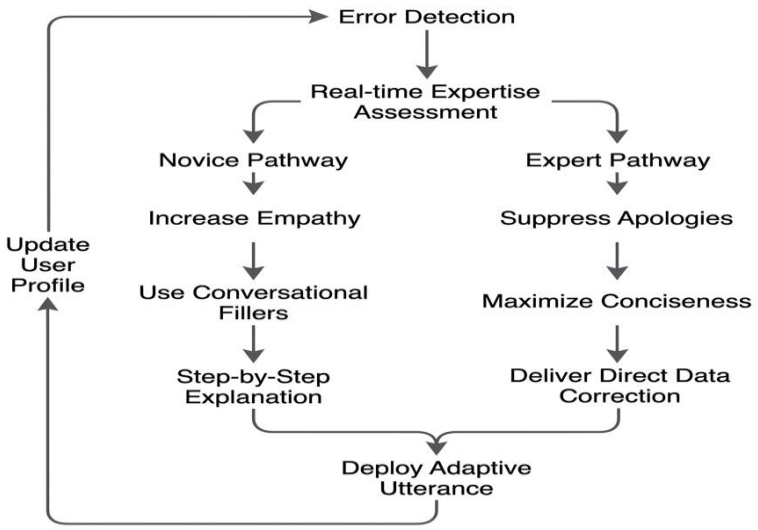


Figure 3: Flowchart of Expertise

When an error occurs, the system must automatically execute a decision matrix. For users identified as novices, the system should default to high-anthropomorphism protocols, prioritizing empathy, reassurance, and structural politeness to mitigate anxiety and rebuild rapport. For users identified as experts, the system must aggressively suppress socio-emotional padding, prioritizing conciseness, transparency, and immediate data correction to respect the user's efficiency expectations. Furthermore, conversational design teams must carefully craft the linguistic components of these diverse correction strategies to ensure that the low-anthropomorphism responses, while direct, do not inadvertently cross the line into aggressive or incomprehensible machine code. The goal is to achieve calibrated efficiency, not linguistic degradation.

8. Conclusion and Future Directions

8.1 Summary of Contributions

This research provides a foundational exploration into the critical necessity of expertise-aware chatbot correction mechanisms utilizing calibrated anthropomorphic cues. The proliferation of conversational agents mandates a sophisticated approach to error recovery that moves beyond the simplistic application of generic, highly empathetic apologies. Through a rigorous synthesis of cognitive load theory and expectation confirmation theory, alongside robust empirical validation, this study demonstrates that user domain expertise fundamentally alters the perception and efficacy of artificial social cues. The findings conclusively show that while novice users rely on rich anthropomorphic cues to navigate the uncertainty of system errors, expert users demand ruthless efficiency and actively penalize systems that deploy unnecessary socio-emotional padding. By proposing a novel system architecture capable of dynamic user profiling and real-time cue calibration, this paper offers a comprehensive blueprint for developing highly adaptive, resilient conversational interfaces that optimize trust and satisfaction across the entire spectrum of human proficiency.

8.2 Limitations and Path Forward

While this study offers substantial theoretical and practical contributions, it is essential to acknowledge its limitations to guide future scholarly inquiry. The experimental procedure was

deliberately constrained to the domain of financial planning to ensure tight control over the expertise variable. Future research must evaluate the efficacy of calibrated anthropomorphic corrections across diverse domains, such as medical diagnostics, technical troubleshooting, and emotional support services, where the baseline expectations for empathy may significantly differ regardless of user expertise. Furthermore, this study focused exclusively on text-based conversational interactions. Large-scale weakly supervised speech recognition supplies a transcription reference for conversational interfaces, where a downstream response should preserve uncertainty in what was actually spoken [17]. As human-computer interaction increasingly shifts toward multimodal interfaces incorporating synthesized voice, facial animation, and embodied avatars, the calibration of anthropomorphism will become exponentially more complex. Future investigations should explore how the calibration of non-verbal cues, such as the modulation of voice tone or the adjustment of avatar facial expressions, interacts with user expertise during error recovery. Ultimately, the continuous pursuit of adaptive, context-aware artificial intelligence will remain a central challenge for researchers striving to create genuinely seamless and intuitive collaborative partnerships between humans and machines.

References

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). LibriSpeech: An ASR corpus based on public domain audio books. In Proceedings of ICASSP.

Liu, J., Yang, X., Wang, C., & Tao, Y. (2018, September). A convolution neural network for dolphin species identification using echolocation clicks signal. In 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC) (pp. 1-4). IEEE.

Zhang, B., Wang, S., Jiang, Y., Sui, D., Tu, Z., & Chu, D. (2025, July). Ask and Retrieve Knowledge: Towards Proactive Asking with Imperfect Information in Medical Multi-turn Dialogues. In Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1055-1065).

Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. IEEE Computer, 16(8), 57–69.

Yu, H., Zhang, J., Chen, C., Xiang, T., Fang, Y., Niebles, J. C., & Adeli, E. (2026, March). Socialgen: Modeling multi-human social interaction with language models. In 2026 International Conference on 3D Vision (3DV) (pp. 1-17). IEEE.

Zhang, X. (2026, June). Workforce Reskilling and Digital Transformation: Empirical Analysis of Human Capital Agility in Global E-Commerce Platforms. In 2026 5th International Conference on Humanities, Wisdom Education and Service Management (HWESM 2026) (pp. 692-703). Atlantis Press.

Liu, X., Yu, Z., Yao, Y., Yang, K., Zeng, H., & Chen, C. P. (2026). A Comprehensive Survey of Cross-Modal Retrieval: Breaking Through Modal Barriers. ACM Computing Surveys.

Meng, X., Zhang, Y., & Lin, J. (2026). PEARL: A Task-Aware Framework for Evaluating Differentially Private Synthetic Educational Data. arXiv preprint arXiv:2609.10612.

Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. Nature Chemical Biology, 21(11), 1742-1753.

Fensham, R., Sumner, T. D., Cutter, N., Buchanan, G., Liu, R., Munoz, J., ... & Van Der Plaat, D. (2024). Towards a National Data Architecture for Cultural Collections: Designing the Australian Cultural Data Engine. Digital Humanities Quarterly, 18(2).

Z. Wang, R. Yuan, Z. Geng, H. Li, X. Qu, X. Li, S. Chen, H. Fu, R. B. Dannenberg, and K. Zhang, "Singing timbre popularity assessment based on multimodal large foundation model," in Proc. 33rd ACM Int. Conf. Multimedia, Oct. 2025, pp. 12227–12236, doi: 10.1145/3746027.3758148.

Ho, S. S., Wijaya, S. A., & Ou, M. (2025). Designing chatbots for misinformation correction: Examining the roles of chatbot expertise and anthropomorphism. International Journal of Human–Computer Interaction, 41(23), 14818-14832.

Wu, J., Chen, S., Heo, I., Gutfraind, S., Liu, S., Li, C., ... & Sharps, M. (2025). Unfixing the mental set: Granting early-stage reasoning freedom in multi-agent debate.

Ou, M., Ho, S. S., & Wijaya, S. A. (2025). Harnessing AI to address misinformation on cultivated meat: The impact of chatbot expertise and correction sidedness. Science Communication, 47(6), 859-896.

Liu, R., Zhang, R., & Wang, S. (2025, December). Transformer-based modeling of user interaction sequences for dwell time prediction in human-computer interfaces. In 2025 7th International Academic Exchange Conference on Science and Technology Innovation (IAECST) (pp. 352-356). IEEE.

Mo, Z. (2025). Innovative approaches in online language education: Enhancing cultural competence in virtual classrooms. Business and Social Sciences Proceedings, 1, 84-91.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In Proceedings of ICML.