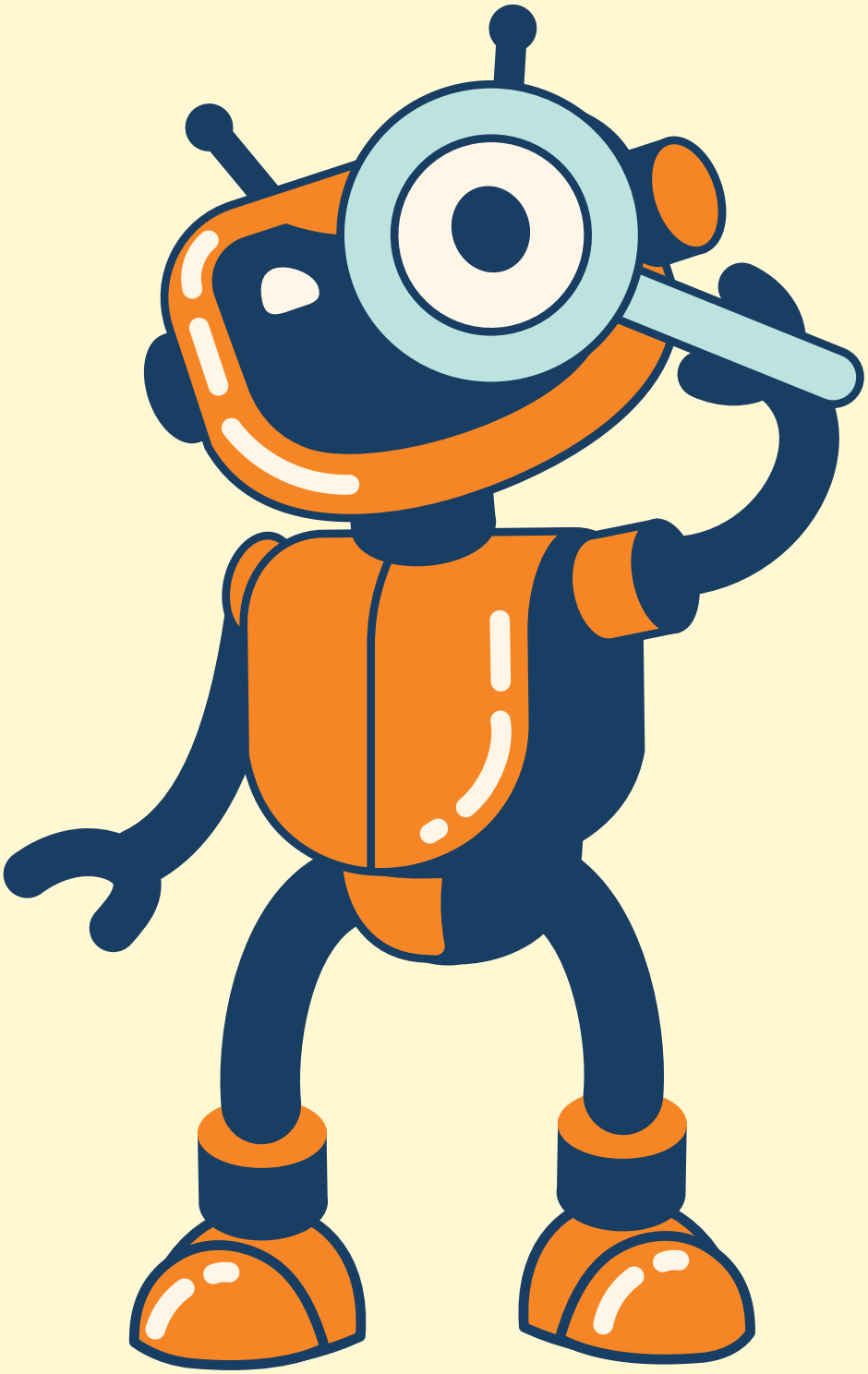

Skeptical Chatbot: A Theory-Based Design for Judgment and Decision-Making in Auditing

Chameera De Silva, James Cook University
Thilina Halloluwa, University of Queensland
Abhijit Das, University of Western Australia
Abhijeet Singh, Curtin University
Mohammad Azim, James Cook University

Insights

- Through structured prompt design and skepticism-focused training, large language models can replicate and reinforce skeptical behaviors among auditors.
- AI interactions grounded in the self-determination theory support deeper engagement and promote autonomy, competence, and relatedness.
- Techniques such as chain-of-thought prompting and SHAP/LIME visualizations make AI reasoning transparent and critical in regulated, high-stakes professional environments.

More than 60 years ago, Douglas Engelbart articulated a visionary perspective in his seminal 1962 report, "Augmenting Human Intellect: A Conceptual Framework." At its core, Engelbart's vision emphasized the use of technology to amplify—not replace—human capabilities. He argued that by enhancing how individuals interact with information, technology could accelerate problem-solving, innovation, and societal progress [1]. Central to this vision was the idea of humans' and machines' coevolution, wherein interactive systems are designed to empower users, not dictate behavior. In this context, motivational design becomes not a luxury but a necessity—crucial for ensuring that AI systems function as cognitive partners rather than passive instruments or autonomous authorities.

This philosophy aligns closely with Ping Zhang's [2] concept of *motivational affordances*, which asserts that effective information system design must address core psychological needs—autonomy, competence, and relatedness—as outlined in self-determination theory (SDT) by Richard M. Ryan and Edward L. Deci [3]. Zhang reframes the challenge of ICT design by arguing

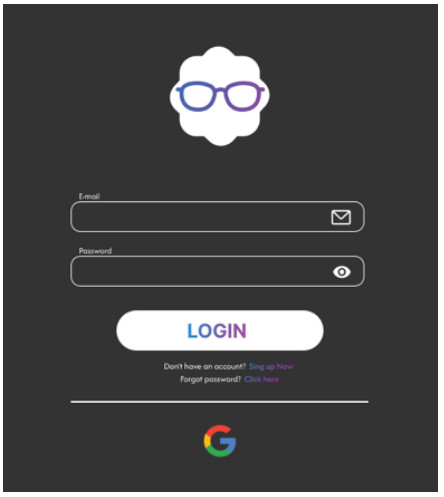

Figure 1. The Accountly chatbot onboarding screen introduces users to a skeptical assistant, setting expectations for critical dialogue rather than quick answers.

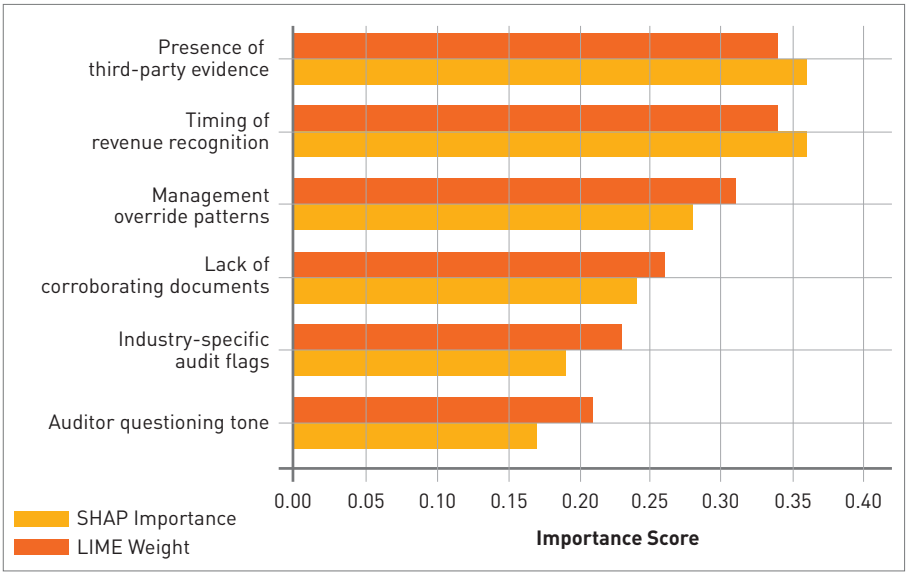

Figure 2. SHAP vs. LIME feature importance comparison for key audit risk indicators.

that system failures often don't stem from a lack of functionality, but from a failure to meet users' intrinsic motivational needs [2]. Crucially, motivational affordances are not inherent in technology itself—they emerge through the dynamic interaction between users and systems. When systems fail to offer users opportunities to express their identity, receive meaningful feedback, or engage in social connection, they risk becoming inert tools rather than agents of personal and professional development. Zhang's framework is particularly salient in the context of AI, where design often prioritizes performance and output efficiency while neglecting the deeper ways in which users engage, adapt, and learn.

In contrast to motivational richness, most current LLM applications are designed for task automation, overlooking the social and cognitive processes that underpin expert reasoning and human development. Chameera De Silva and Thilina Halloluwa, the coauthors of this article, confronted this oversight [4]. While they acknowledge LLMs' capacity to streamline creativity and knowledge work, they raise serious concerns about social disconnection and trust erosion. When AI replaces knowledge platforms, users lose access to collaborative sensemaking and peer validation. In professional domains, this kind of substitution may undercut both critical thinking and emotional resilience—traits that are deeply tied to social interaction and motivational grounding. De Silva and Halloluwa warn that the path toward convenience must not erode interpersonal communication or ethical reflection. Their work builds directly on Zhang's [2] emphasis on relatedness and competence, reminding us that people must feel their skills matter and that they remain embedded in a human network.

It is 4:45 p.m. on a Friday. A junior auditor receives a last-minute justification for a revenue spike. The CFO says it was a strategic sales push. No documents, just words. The auditor hesitates. What should she ask next?

In a near-future audit office, a skeptical chatbot trained on international audit standards and ethical reasoning nudges her: *"Request supporting documentation. Evaluate the timing of revenue recognition. Has delivery occurred?"* The moment is small, but the implications are profound. In auditing—as in healthcare, law, and education—AI is entering not to automate judgment, but to scaffold it.

This article describes the design of an LLM-powered chatbot that models professional skepticism (PS)—a critical behavior in auditing defined by inquiry, evidence corroboration, and ethical vigilance. It also argues for a critical human-technology perspective: Trust and integrity in AI begin with how we train it to question things.

AUDITING AS HUMAN-CENTERED RESPONSIBLE DUTY

External auditors face conflicting demands: They must trust and verify, assess and report, detect error and fraud, yet maintain rapport with client management. While statutory checklists and regulations abound, the intangible skill of skepticism often falters under pressure or ambiguity. Studies have shown that even trained professionals are sometimes inclined to accept client explanations or over-rely on past experiences, creating blind spots. With generative AI's rise, there is both risk and opportunity: Systems may mimic bias or reinforce compliance-only behavior. But if designed thoughtfully, AI can also embody the reflective, probing stance that auditors need to uphold public trust.

BUILDING A SKEPTICAL CHATBOT: WHAT AND WHY

This chatbot does not replace auditor judgment; it scaffolds it. Our goal is to fine-tune GPT-4 to exhibit domain-aligned PS. We trained it using a multisource corpus: regulatory texts, published peer-reviewed articles on behavioral bias, detection of error and

Trust and integrity in AI begin with how we train it to question things.

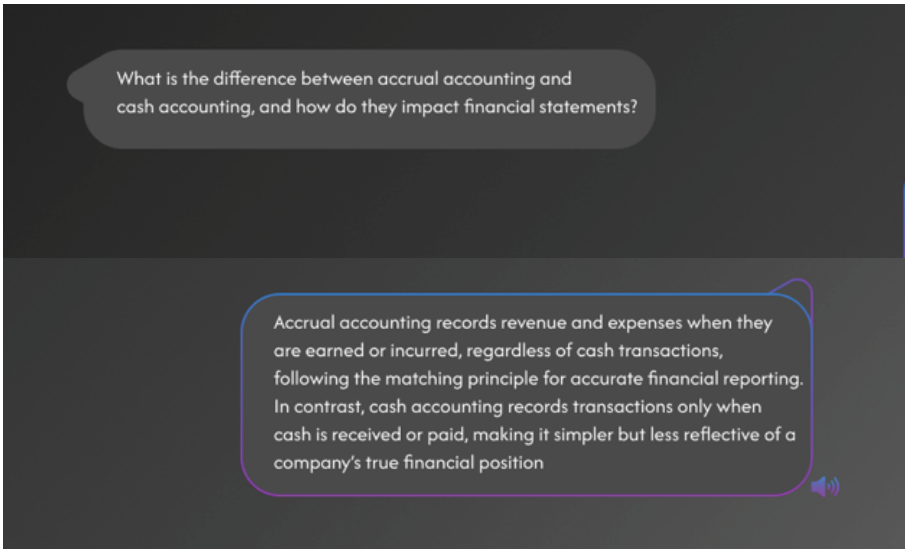

Figure 3. A sample interaction where the chatbot nudges the user to verify a CFO's verbal claim. The dialogue reflects skepticism and autonomy support.

fraud, audit failures, case transcripts, and expert-authored manuals. But it was not just about ingesting contents—it was about *embedding questioning behavior*.

It simulates conflicting evidence, challenges assumptions, and issues counterfactual prompts. By doing so, it cultivates the capacity to doubt constructively, meeting regulatory calls for heightened skepticism. Here, the chatbot functions not as a substitute for expertise, but as a tool to enhance cognitive reflection and ethical and moral awareness.

We employ prompt engineering, multi-turn dialogues, professional skepticism-labeled examples, and model tuning for prioritized segments rich in analytical reasoning. Most importantly, we grounded this artifact (i.e., chatbot) in Ryan and Deci's SDT, a cognitive motivational framework that supports the fundamental psychological needs. Those include autonomy, competence, and relatedness that must be fulfilled to foster intrinsic motivation and psychological growth of an individual.

DESIGNING FOR AUTONOMY, COMPETENCE, AND RELATEDNESS

Prompt design for our chatbot was not arbitrary. Each scenario was mapped to one or more SDT needs:

- *Autonomy:* Does the chatbot encourage self-directed inquiry, such as suggesting alternative procedures when audit evidence is missing?
- *Competence:* Does the chatbot foster task mastery and critical appraisal, such as questioning valuation assumptions or regulatory and statutory alignment?
- *Relatedness:* Does the chatbot maintain ethical and moral awareness and social coherence within professional norms, such as recognizing the risk of management override?

This motivational scaffolding helps align machine interactions with human learning, not just model outputs with audit codes (Figure 1). Unlike traditional automation tools, this system is framed as a collaborative agent, not a decision maker.

EXPLAINABILITY FOR TRUST IN HIGH-STAKES DOMAINS

To ensure transparency, we simulate explainability with LIME and SHAP (Figure 2). Each response can be traceable to influential features: lack of documentation, contradictory timing, vague justifications. Chain-of-thought prompting further enabled the chatbot to reason aloud, displaying logic paths that human users could critique.

These visual and linguistic cues make it possible not just to *see what the model predicts*, but *why it takes a skeptical stance*—a vital component for acceptance in regulated fields.

INTERFACE AS COGNITIVE PARTNER

While the chatbot's reasoning engine is grounded in SDT, its value is fully realized only when those motivational principles are embodied in the user interface. In high-stakes domains such as auditing, design is not a cosmetic layer—it is an extension of cognitive function. The interface must not merely deliver outputs but scaffold user judgment, stimulate inquiry, and reinforce ethical awareness.

The *mobile version* of the skeptical chatbot was purposefully designed for *just-in-time* interactions, ideal for auditors on the move or those working in remote locations. As shown in Figure 3, the design integrates motivational cues through tone calibration, color-based feedback, and pacing of conversational exchanges. For instance, when a user receives a vague verbal explanation for a revenue spike, the chatbot doesn't just flag it—it suggests evidence-gathering alternatives in a tone that encourages action without dictation.

This interaction rhythm helps users feel in control (autonomy), receive structured cognitive reinforcement (competence), and stay aligned with professional norms (relatedness). Even subtle user interface choices, such as spacing in message bubbles, feedback animations, or role-conditioning headers, contribute to the chatbot's *pedagogical role*. It becomes a *cognitive partner*, not just a question-answering machine.

FROM BEHAVIORAL FIDELITY TO MOTIVATIONAL ALIGNMENT

The evaluation of the chatbot, in this context, is not measured by speed or accuracy, but by the system's ability to enhance judgment and reflection. This paradigm shift—from performance-driven to motivation-driven AI—echoes Zhang's original vision [2]. Moreover, our chatbot's training loop incorporates both auditing standards

Prompt Scenario	Skeptical Response	SDT Alignment
Missing confirmations	"Request alternative procedures"	Autonomy, competence
Inventory undervaluation	"Challenge assumptions"	Autonomy, competence
Verbal-only explanation	"Insist on documentation"	Competence, relatedness
No discrepancies	"Maintain skepticism"	Autonomy, relatedness

Figure 4. Mobile design emphasizes motivational cues through language tone, color feedback, and conversational pacing—ensuring the experience aligns with ethical and psychological needs.

and psychological theory, translating motivation into a design principle rather than a theoretical backdrop. In this way, the system exemplifies interdisciplinary rigor: merging technical precision with motivational science. While LLMs are often seen as general-purpose engines, this chatbot is sculpted into a specialized sociocognitive partner, aligned with Engelbart's recursive model of human-computer augmentation: "getting better at getting better" (Figure 4).

Our evaluations (using a custom benchmark SCEPT-X) showed that the skeptical chatbot outperformed base GPT-4 and other domain-specific models on the following:

- Skepticism index
- Red flag detection accuracy
- Follow-up quality score
- Judgment alignment score.

More importantly, the review of responses generated by customized GPT reveals notable improvements in Socratic dialogic reasoning, moral acuity, and the articulation of unbiased judgment. The chatbot moved beyond merely replicating audit scripts: It demonstrated the ability to pause and probe thoughtfully, capturing the essence of professional skepticism.

DESIGNING FOR THE NEXT GENERATION OF PROFESSIONALS

What emerges here is not just a tool for audits. It is a model for designing AI in *any* domain where ethical and moral reasoning, judgment, and decision- making in ambiguous situations converge with trust. From education to healthcare, the need is the same: AI that emulates and supports—not shortcuts—critical human thinking.

As noted in recent work by the coauthors of the article, LLMs are increasingly being framed as copilots in HCI [4]. Our audit chatbot follows this trajectory, offering not just task support but also ethically aligned, dialogue-based scaffolding of professional judgment. Together, these three contributions—Zhang's motivational framework, De Silva and Halloluwa's ethical lens, and Ryan and Deci's SDT-aligned chatbot prototype—present a compelling model for the future of AI integration. As Engelbart cautioned, automation that bypasses human values poses a critical risk. The path forward, as these works collectively argue, lies not in replicating human behavior, but in designing systems that actively foster mental well-being.

In conclusion, a skeptical chatbot does not replace the auditor, but it can pose the right question at the right moment, particularly when the human might not due to cognitive load or other environmental constraints.

ACKNOWLEDGMENTS

This article is one of the research outputs from a project funded by Certified Practising Accountant (CPA) Australia through its Global Research Perspectives Program. The authors acknowledge and appreciate the financial support and encouragement, which made this ongoing research possible.

References

1. Engelbart, D. C. Augmenting human intellect: A conceptual framework. Stanford Research Institute, 1962.
2. Zhang, P. Motivational affordances: Reasons for ICT design and use. *Communications of the ACM 51*, 11 (2008), 145–147; https://bit.ly/4lFFjrF
3. Ryan, R. M. and Deci, E. L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist 55*, 1 (2000), 68–78.
4. De Silva, C. and Halloluwa, T. 2025. Augmenting human potential: The role of LLMs in shaping the future of HCI. *Interactions 32*, 2 (2025), 42–45; https://bit.ly/47SxQQ9

Chameera De Silva is a clinical data scientist at Annalise.ai and an academic at James Cook University. With extensive experience in data engineering and clinical research, he specializes in designing, developing, testing, and deploying innovative applications. Proficient in writing codes and algorithms, he uses various programming languages to build complex neural networks.
→ chameera.desilva@my.jcu.edu.au

Thilina Halloluwa is a lecturer in the School of Electrical Engineering and Computer Science at the University of Queensland in Australia. He has more than 15 years of academic and industry experience.
→ t.halloluwa@uq.edu.au

Abhijit Das is a lecturer at James Cook University.
→ abhijit.das@jcu.edu.au

Abhijeet Singh is a senior lecturer at Curtin University in Perth, Australia. His research interests include earnings management, earnings and audit quality, and corporate governance.
→ abhijeet.s@cbs.curtin.edu.au

Mohammad Azim is a senior lecturer in financial accounting and auditing at James Cook University. During his 23-plus years in academia, he has published more than 50 studies and three books.
→ mohammad.azim@jcu.edu.au