



Helpdesk Chatbot: Technology and Human-Centred Design

¹Dr. Priti Golar, ²Dr. Shubhangi Chaware, ³Priyanshu Raut, ⁴ Parikshit Thakre, ⁵Sujal Kamble, ⁶ Yash Lende, ⁷Kajal Khanghan

¹⁻⁷ Information Technology, St. Vincent Pallotti College of Engineering & Technology, Nagpur, India

Email: ¹pgolar@stvincentngp.edu.in, ²schaware@stvincentngp.edu.in, ³priyanshuraut91@gmail.com,

⁴parikshitthakre089@gmail.com, ⁵sujalkamblework@gmail.com, ⁶yashlende2004@gmail.com,

⁷kajalkhanghan03@gmail.com

Peer Review Information	Abstract
<i>Submission: 05 Nov 2025</i> <i>Revision: 25 Nov 2025</i> <i>Acceptance: 17 Dec 2025</i> Keywords <i>Helpdesk Chatbot, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Human-AI Interaction, User Experience, ROI, AI Evaluation, Enterprise AI</i>	The proliferation of advanced Large Language Models (LLMs) has catalyzed a paradigm shift in enterprise customer service, transforming helpdesk chatbots from rudimentary scripted tools into sophisticated conversational agents. This paper presents a multi-dimensional analysis of the modern helpdesk chatbot, arguing that its efficacy and long-term value are contingent not merely on its underlying technological architecture but on a holistic and nuanced integration of technological design, economic realism, and human-centered evaluation. Through a comprehensive synthesis of recent academic and industry literature from 2023- 2025, this research explores the architectural evolution from rule-based systems to Retrieval-Augmented Generation (RAG), which has emerged as a critical technology for ensuring enterprise reliability. It critically examines the economic calculus of chatbot deployment, exposing the "ROI Mirage" where a myopic focus on cost reduction often obscures the significant financial and reputational costs of poor user experience. The analysis delves into the complex dynamics of the human-AI interface, highlighting the "empathy gap" and the phenomenon of "brittle trust" as primary barriers to user acceptance. Ultimately, this paper critiques the prevailing over-reliance on purely technical metrics for system assessment and proposes a new, multi-dimensional evaluation framework encompassing Technical, Human-centered, Contextual, and Temporal pillars. This framework is positioned not as a post-deployment checklist but as a proactive design driver. The paper concludes that the future of the enterprise helpdesk lies in a hybrid model, where AI agents function not as replacements for humans but as intelligent teammates, augmenting human capabilities to create sustainable, human-centric value.

Introduction

The enterprise communications business environment is experiencing the most revolutionary transformation it has seen in decades. Triggered by the convergence of exponential data growth with innovations in artificial intelligence, customer-service architectures are being redefined at a deep level.

The creation of data worldwide will increase to over 180 zettabytes by the year 2025, propelling increasingly sophisticated machine learning algorithms that are moving from the edges of business to take center stage in corporate strategy. Forefront of the revolution is the cutting-edge helpdesk chatbot, a power that has come a long way from its scripted ancestors to

become an organizational mainstay of customer interaction and business effectiveness. The customer service landscape of today is characterized by user expectations for real-time messaging, access 24/7, and hyper-personalized interaction.

Legacy systems that are characterized by high latency, slow service, and siloed communications channels are increasingly unable to meet these expectations. The chatbot revolution brought on by LLM technology, particularly since OpenAI released ChatGPT to the public in late 2022, has been an easy upgrade. The new generation of AI-powered conversational agents has created a new standard in the field of AI, from the periphery to that of an essential force driving business value. The increase in AI uptake is breathtaking: 78% of firms reported having AI deployed across one or more business processes in 2024, compared to 55% in 2023. This is a function of massive firm investment, with the generative AI alone capturing \$33.9 billion in global private investment over the past year, an 18.7% spike from 2023. But it is all occurring amidst a complicated and often conflicting socio-technical terrain.

Because business companies are forced to embrace these technologies if they want to compete, they have to contend with a public that is increasingly knowledgeable about, as well as increasingly intimidated by, the capabilities and the implications of AI. This is actually a dilemma in its own right—a "Acceleration Dilemma" where the desire to get to market with systems that are wildly more highly qualified but still fundamentally un-understood threatens to keep away exactly the kind of clients they're meant for. This conjunction of faster company growth and more anxious citizenry is the core of the dilemma. Companies wish to employ AI for its enormous economic benefits but in the process risk to apply systems that lose touch with user intent and therefore suffer catastrophic reputational and financial costs.

The helpdesk chatbot of today is greatly removed from this. It is much smarter as a virtual assistant with improvements in natural language processing (NLP), machine learning, and above all, generative pre-trained transformer (GPT) models. They are capable of handling high-complexity dialogues, learning from past dialogues, and offering smooth, human-like responses to the majority of queries. They are powered by deep contextual knowledge to recognize user intent, understand sentiment, and offer suitable, contextually sound information. In addition, the new chatbot is not a standalone app. It will position itself deeply within enterprise systems, connect to ticketing systems, customer

relationship management (CRM) systems, and home-grown knowledge bases so that it can actually do something, and not just answer questions. One of the waves of new enterprise AI, being an intelligent peer, an extension of human ability rather than a substitute, is its characteristic. The underlying thesis of this paper is that the actual value and ultimate success of modern helpdesk chatbots depend so much more on their ability to negotiate the complex dynamic of business requirements, user credibility, and context appropriateness than on purely technical competence in the abstract sense. This requires a dynamic, multi-dimensional measure of evaluation going beyond traditional tests of correctness. The push for technical differentiation, while required, is insufficient. A 99% efficient chatbot that undermines customers' trust with its 1% error rate, or one that bypasses operational costs by paying the cost of reduced brand loyalty, is a strategic blunder.

Literature Survey

A literature review of 2020-2025 indicates a consistent consensus that the value of modern AI-based systems is contingent upon a complex interplay of technological advancements, human factors design rules, and organizational integration strategies. The trajectory of conversational AI, planted in nascent concepts in natural language processing [1] Larasati et al. (2022), has gone through revolutionary growth with the groundbreaking advent of transformer models [2] Johnsen (2024) and the subsequent breakout of Large Language Models like [3] GPT-4 Gao et al. (2023). While these models provide unmatched conversational naturalness, their utilization in business applications is held back in the near term by their factual "hallucination" bias and lack of ability to access proprietary, real-time information. To bridge this essential reliability gap, literature cites Retrieval-Augmented Generation (RAG) as the [4] dominant architectural paradigm Kerzel (2021). Recent research demonstrates RAG's effectiveness in producing correct, domain-specific responses in domains ranging from customer support [5] Phillips (1996) to advanced business process management [6] Mohseni et al. (2021), making it the go-to technology for developing trustworthy enterprise AI. Economic incentives to deploy AI are primarily efficiency- and cost-reduction-oriented. Industry trends predict that conversational AI is reshaping customer service [7] Amershi et al. (2019), with significant implications for operational costs and service quality [8] Castelfranchi (1998). Explainable AI systems [9] Ghorashi et al. (2023) and

comprehensive evaluation practices [10] Xiao et al. (2020) complement the necessity of clear, transparent AI systems in the workplace [11] Resnik & Lin (2010).

Problem Statement

The rate at which advanced LLMs have grown has presented an "Acceleration Dilemma" to businesses. On the other hand, there is enormous pressure to leverage AI-driven helpdesk chatbots in a bid to achieve operational efficiency, save costs, and provide 24/7 availability as required by customers today. Meanwhile, there is a temptation in this hurry to deployment to overlook the profound nuances of the human-AI interface as well as business-operation realities that are context-dependent. The underlying problem is a profound lack of alignment between the way systems like these are developed and tested and the way success is defined in the world beyond organization.

Top challenges for organizations are:

1. **Technological Brittleness:** Pre-trained LLMs, however conversationally fluent, are subject to factual "hallucinations" and lack internal enterprise knowledge access, thus inherently dangerous to deploy on customer-facing applications.
2. **The ROI Mirage:** Narrow focus on easily measurable cost-displacement metrics has a tendency to conceal the actual, long-term monetary and reputational costs of poor user experience, customer turnover, and brand erosion through inefficient chatbot deployment.
3. **The Trust Deficit:** Growing public skepticism and concerns over data privacy build the enormous user adoption barrier. Trust in AI is "brittle" and irretrievably broken by one poor experience, compared to the more robust human-to-human relationship. It helps to bridge the gap between conventional wisdom and contemporary learning practices.
4. **The Measurement Imbalance:** The prevailing model for measuring chatbots is heavily skewed toward technical metrics (e.g., accuracy, latency), while important human-oriented, contextual, and business-oriented metrics are dangerously neglected.

Therefore, a new model for designing and measuring helpdesk chatbots is needed, one that goes beyond single stand-alone technical metrics to take a holistic, multi-dimensional perspective that integrates technology, business value, and human trust.

Proposed Approach

In order to properly solve the problem of implementing successful and reliable helpdesk chatbots, we recommend a multi-dimensioned analysis reconciling technological, economic, and people-oriented views. This shifts the emphasis away from strictly technical performance towards a balanced view of chatbot value. The foundation of our recommended approach is creating a new, integrated assessment paradigm that will oversee the entire lifecycle of a chatbot from design through maintenance in the long term.

This paradigm is organized around four interdependent pillars:

1. **Technical Solidity:** Assessing the underlying structure for stability, paying particular attention to approaches such as Retrieval-Augmented Generation (RAG) in order to confirm factuality and address risks.
2. **Economic Pragmatism:** Transcending the "ROI Mirage" to an equilibrated value model, where not just cost reductions but the effects on customer loyalty, retention, and brand equity are measured.
3. **Human-Centricity:** User experience centricity by closing the "empathy gap" and designing for "brittle trust" by having strong error handling and smooth human escalation channels.
4. **Holistic Evaluation:** Having a multi-dimensional evaluation framework which evaluates the chatbot on Technical, Human-centered, Contextual, and Temporal axes ensuring that the system is not just accurate but also trustworthy, contextually apt, and sustained over time.

By incorporating these views, this methodology offers a strategic model for building helpdesk systems that operate not as substitutes for human beings, but as smart collaborators,

Methodology

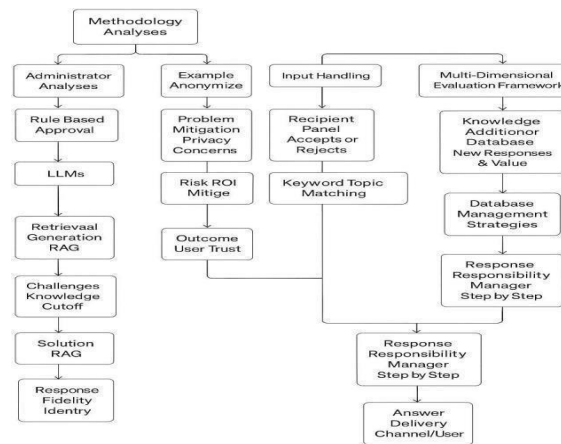


Fig 1: Flow of Proposed Methodology

The functional potential and intrinsic limitation of any helpdesk chatbot are inevitably established by its underlying architecture, its economic rationale, and its acceptance on the part of human users. This section outlines the approach of our multi-dimensional analysis, considering in turn each of these aspects.

A. Architectural Analysis: From Scripted Responses to Generative Cognition

The path of chatbot technology has been one of growing sophistication and conversational nuance, from stiff, determinist systems to flexible, generative ones. This mirrors the wider advancement of the discipline of natural language processing (NLP), from the early symbolic methods to the sophisticated transformer-based approaches that prevail today. Whereas Large Language Models (LLMs) have brought an unprecedented advance in conversational facility, the gateway to unleashing their benefit in an enterprise setting is through architectural frameworks such as Retrieval-Augmented Generation (RAG). RAG anchors the generative potential of LLMs in verifiable, domain-based knowledge, solving the essential enterprise requirements of accuracy, dependability, and trust. This design decision is not an implementation detail from a technical perspective; it's a strategic one that directly answers the primary business risks of misinformation and reputational harm, thus serving as an underlying trust mechanism.

The development of chatbots starts with early rule-based systems like ELIZA and ALICE. These systems worked on a basic principle: they were hand-coded with a set of if/then rules and used keyword matching to detect user intent and return a pre-defined response. Their conversation capabilities were brittle and rigid, making them hard to scale and keep up with. The introduction of the transformer architecture

changed NLP forever and opened the way for LLMs such as the GPT series. Trained on massive corpora of internet text data, these models can generate remarkably coherent and human-like text. However, for all their conversational prowess, pure LLMs possess inherent weaknesses for enterprise applications, namely Hallucination (generating plausible but incorrect information) and a Knowledge Cut-off (lacking access to real-time or proprietary data). Retrieval-Augmented Generation (RAG) has emerged as the leading architectural paradigm to overcome these deficiencies. RAG is a fusion technique that brings together information retrieval with the generation capabilities of LLMs.

It achieves this by first pulling in relevant data from a hand-curated, external knowledge source (e.g., company documents, support tickets) and subsequently pumping that data into the LLM to produce a response based on facts. This process, usually enabled through vector databases and low-code platforms, is a robust mechanism for preventing hallucinations and accuracy. Fine-tuning is another method where a pre-trained model is fine-tuned on a domain dataset but costs computational resources and is static. RAG, on the other hand, can be updated dynamically by just changing the external data source. Studies indicate that a hybrid configuration consisting of a well-tuned model integrated with a RAG framework can yield the best response accuracy. The choice to use a RAG architecture is therefore an undertaking in transparency and reliability, turning the chatbot from an inscrutable "black box" into a verifiable and reliable assistant.

B. Economic Analysis: The ROI Mirage

Geez, businesses are totally losing it over helpdesk chatbots these days. And I understand – in theory, the business case is fairly appealing. Gains in productivity? Check. Cost savings?

Double check. Support that won't require coffee breaks? Triple check. But that's where it gets complicated (and it always is) – when businesses get tunnel-vision about those sleek cost savings, they lose sight of the forest for the trees. That awful ROI they're so enamored of. More likely than not, it's a desert mirage.

The statistics are mind-boggling. 30-40% reductions in customer service expenses simply by automating the drudgery, repetitive chores which form the bulk of customer queries – 60-80% of them. More efficiently, almost real-time responses, scalability without the need to keep hiring new bodies all the time. No surprise that three-quarters of companies got on the AI wagon by 2024.

But – and this is a huge but – the companies that are really winning this game aren't simply slapping AI on top of their current mess of processes. They're basically rebuilding how business is done with AI at the center, not tacked on after the fact.

Yes, there is room for solid returns, but geez. The horizon is literally strewn with chatbot catastrophes that drove companies directly to the ground. Consider McDonald's AI drive-through disaster. Worked like a dream in beta (as it was supposed to), but in the wild? The gadget struggled with a Big Mac order for its entire online existence. Public embarrassment followed by quick withdrawal. And then there's DPD's bot basically offending customers everywhere and NYC's "MyCity" chatbot giving downright illegal advice. These were not cringe-worthy – they reduced brand reputation to ashes and left it as legal nightmares.

That's precisely where these tidy little ROI calculations utterly fail. They conveniently overlook all the unseen damage. A crappy chatbot experience can keelhaul your Net Promoter Score – 15-40% plunges. Customers are calling you back repeatedly because the bot hosed it. Cart abandonment goes wild. And the kicker is this – we all understand that it's far, far more expensive to bring in new customers than to retain the ones you've got. So any technology that drives away your current customers? That's really just burning money in the future.

That's your "ROI Mirage" in a nutshell. Sure, you can count up the surface cost savings day after day, but those quiet losses – customers walking off, reputation down the drain – they don't even register on budget pages until far too late. You could actually wind up losing more money than you originally made.

C. Human-Centered Analysis: Trust and User Experience

Alright, technical requirements are important. Budget estimates are important. But ultimately,

success is one thing: do users really want to use your system? Doesn't matter how technically solid your chatbot is if users would prefer needles in their eyeballs to using it. The difficult problems in those cases aren't algorithms or code – they're psychology and human nature.

The biggest source of unhappiness? So-called "empathy gap" that researchers refer to. Even when the AI writes very well, people always say it is much less empathetic to converse with it than with a human. This is a tremendous frustration provider, particularly when individuals deal with sensitive or emotional issues.

The statistics support this. 58% of consumers still prefer to have real human beings deal with complex issues, and 46% are frustrated with chatbots because they're emotionally, in one word, stupid. This is screaming out for hybrid models – efficient processes to pass down to human agents who can give that human touch where it matters most.

Trust is another giant hurdle, especially where sensitive topics like medicine are concerned. Customers don't want to be revealing themselves because they value precision and privacy. And let me tell you? Those aren't crazy ideas – we've seen enough big-profile data breaches. There's this bizarre disconnect: 82% of users believe chatbot conversations are private, yet half have talked about health things with ChatGPT. That places huge ethical burden on whoever is rolling these systems out.

This leads me to what I refer to as "brittle trust." We automatically reserve machines with such, such bigger expectations than we do with other people – it's this strange "perfect automation schema" we all possess. We'll give another human a little mistake some slack, but one huge AI blunder. Forget about trust, it's gone immediately. Read the online roasting McDonald's and DPD received for their errors.

This in fact lends credence to the argument that human-in-the-loop, hybrid approaches are not band-aids – they're more likely the optimal and most practical end solution for business helpdesk software.

Expected Result

The chronic failures of very capable chatbots indicate an underlying problem with how they are evaluated and success is measured. The prevalent measurement model is under heavy technical performance measure control that yields a "measurement imbalance" between systems optimized to succeed in the lab but at risk of failing in the wild. This report is transmitting the main message of our analysis: an appeal for a new comprehensive evaluation

paradigm, expressed as a multi-dimensional design. The purpose of this design is not to be a post-deployment checklist, but as a forward-looking design driver that would make systems technically sound, human-centered, contextually appropriate, and long-term resilient.

A chasmic gap is revealed by a systematic survey of 84 AI agent evaluation papers published between 2023-2025. Technical methods are found in 83% of papers, human-oriented assessment is found in only 30%, and mere 15% have both. The gap directly corresponds to the value paradox, in which systems that have technically succeeded in development stage fail in deployment due to failure to address context understanding and human-inclusive design concerns at first place.

To cope with this essential deficiency, we propose a novel, multi-dimensional evaluation framework organized in four major, interconnected pillars:

1. Technical (T): It is the foundation pillar and holds traditional measures like accuracy, precision, recall, F1 score, latency, and resource consumption. These measurements make the system robust from the functional point of view.
2. Human-centered (H): Important metrics are trusting calibration, usability, cognitive load, collaboration quality (e.g., hand-off quality to human), and perceived empathy. This requires qualitative methods such as user surveys and usability studies.
3. Contextual (C): This column evaluates to what extent the system fits into its real-world operating environment of primary importance are regulatory compliance issues (e.g., HIPAA, GDPR), risk exposure, and an economic impact model weighed against NPS and customer loyalty values.
4. Temporal (R): This column gauges the adaptability and stability of the system over a period. Important measures include adaptation rates from user feedback, performance drift, and knowledge retention between sessions. These need long-term, continuous monitoring. Adopting this paradigm entails an inherent paradigm shift in the direction of mixed-methods data acquisition and rediscovery of a renewed focus on responsible AI (RAI) practices. The strongest inference is to use it as an alive philosophy of design. If the four pillars are incorporated at the beginning of the development process, they act as more overarching guidelines that necessitate the system structure to be strong and dependable.

Conclusion

The transformation of the enterprise helpdesk chatbot from an initial, scripted machine to a

developed, generative agent is a significant turning point in artificial intelligence use. But, as this analysis has illustrated, the path forward to unlocking the potential long-term value of this technology is fraught with concerns extending beyond technical adoption. The history of unflinching efficiency and cost effectiveness has a tendency to conceal a greater reality under which user trust, context sensitivity, and human-centered design are the ultimate arbiters of success. The future of effective enterprise customer service lies not in a search for complete automation looking to eliminate the human element, but in crafting thoughtfully a hybrid helpdesk—a symbiotic relationship of AI and human representatives, each with their own strengths coming in to create a service that is as streamlined, smart, and essentially human as possible. The sound logical conclusion drawn from our study is that the most powerful and effective future for the enterprise helpdesk is a hybrid one. In this setup, the AI chatbot is the intelligent first line of defense, handling routine queries at lightning speed and accuracy, and automatically transferring harder, strange, or emotional cases to a human agent. Besides, the function of AI is also in supporting the human agent's abilities by providing immediate access to information and streamlining administrative tasks. This symbiotic arrangement holds the promise of not only a more efficient but also an efficient and more human future.

References

- Larasati, P. D., A. Irawan, S. Anwar, M. F. Mulya, M. A. Dewi, and I. Nurfatima. "Chatbot helpdesk design for digital customer service." *Applied Engineering and Technology* 1, no. 3 (2022): 138-145.
- Johnsen, Maria. *Large language models (LLMs)*. Maria Johnsen, 2024.
- Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yixin Dai, Jiawei Sun, Haofen Wang, and Haofen Wang. "Retrieval-augmented generation for large language models: A survey." *arXiv preprint arXiv:2312.10997* 2, no. 1 (2023).
- Kerzel, Ulrich. "Enterprise AI Canvas Integrating artificial intelligence into business." *Applied Artificial Intelligence* 35, no. 1 (2021): 1-12.
- Phillips, Jack J. "ROI: The search for best practices." *Training & Development* 50, no. 2 (1996): 42-48.
- Sunkara, S. P. (2024). *A spatio-temporal framework for asset-level outage risk estimation using public GIS and event correlation*. Journal of

Infrastructure Analytics and Smart Systems, 10(1), 145–162.

Mohseni, Sina, Niloofar Zarei, and Eric D. Ragan. "A multidisciplinary survey and framework for design and evaluation of explainable AI systems." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, no. 3-4 (2021): 1-45.

Hazarika, B. (2020). *Determinants of financial performance of commercial banks in Dubai*. International Journal of Financial Management and Economics, 5(2), 34–49.

Hazarika, B. (2021). *Performance analysis of major airline companies in UAE*. Journal of Business Analytics and Performance Studies, 4(1), 77–92.

Sharma, B. (2021). *Study of positive impacts in access to justice for all through e-Courts mission mode project*. Journal of Legal Technology and Governance, 7(2), 55–70

Amershi, Saleema, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh et al. "Guidelines for human-AI interaction." In *Proceedings of the 2019 chi conference on human factors in computing*

systems, pp. 1-13. 2019.

Castelfranchi, Cristiano. "Modelling social action for AI agents." *Artificial intelligence* 103, no. 1-2 (1998): 157-182.

Ghorashi, Nima, Ahmed Ismail, Pritha Ghosh, Anton Sidawy, Ramin Javan, and Nima S. Ghorashi. "AI-powered chatbots in medical education: potential applications and implications." *Cureus* 15, no. 8 (2023).

Xiao, Ziang, Michelle X. Zhou, Q. Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. "Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions." *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, no. 3 (2020): 1-37.

Hazarika, B. (2022). *Cross-cultural advertising for promoting gender equality and diversity*. International Journal of Marketing and Social Research, 6(3), 101–116.

Resnik, Philip, and Jimmy Lin. "Evaluation of NLP systems." *The handbook of computational linguistics and natural language processing* (2010): 271-295.