Evaluating the Trust-Efficiency Trade-off Between AI Chatbots and Human Service Agents: A Systematic Review and the CADE Deployment Framework

Aaditi Sarode[1], Ragini Marathe[2], Pallavi Kandalkar[3], Sakshi Badgujar[4]
Prof. M. E. Maniyar[5], Prof. P.S. Kurne[6]
[1,2,3,4,5,6]*K. K. Wagh Institute of Engineering and Education Research, Nashik*

***Abstract*—Artificial intelligence-powered conversational agents have gained significant adoption across service sectors including healthcare, retail, and education. These systems deliver measurable operational improvements resolution speeds three to five times greater than human personnel and cost reductions of 60 to 80 percent per transaction yet they raise ongoing concerns regarding perceived reliability, emotional inadequacy, and erosion of user trust. This paper presents a structured review of peer-reviewed literature to characterize the fundamental tension between service efficiency and trustworthiness when comparing automated and human agents. An accompanying pilot survey of 9 participants across diverse service scenarios empirically quantifies trust and efficiency preferences. The synthesis informs the proposed Context-Aware Deployment Model (CADE), which classifies incoming service interactions along task complexity and sensitivity dimensions. CADE advances the field by replacing binary deployment approaches with a context-sensitive hybrid allocation framework. Larger-scale validation is recommended as future work.**

***Index Terms*—AI chatbots, context-aware deployment, human-AI interaction, service efficiency, user trust, hybrid systems, pilot study**

I. INTRODUCTION

Customer-facing service delivery has undergone substantial transformation as organizations increasingly substitute or complement human personnel with AI-driven conversational agents. Deployed across e-commerce, financial services, clinical administration, and educational support, chatbot systems offer capabilities continuous availability, instantaneous response, and massive concurrent handling that are economically and logistically impractical through human staffing alone [1]. The rapid advancement of natural language processing and transformer-based architectures has further enhanced the conversational sophistication of these systems.

Incoming interactions are characterized along two axes: task complexity (the degree of procedural and cognitive demand involved) and task sensitivity (the magnitude of personal, financial, or emotional consequence associated with the interaction). This orthogonal characterization determines whether engagement is assigned to an automated agent, a human agent, or a collaborative hybrid configuration each yielding distinctive profiles of operational efficiency, user trust, and satisfaction.

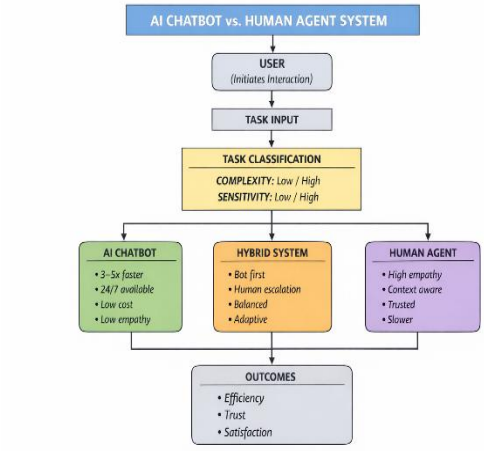

Fig. 1. Service Interaction Classification Framework

Despite measurable efficiency advantages, automated systems consistently demonstrate limitations in empathic comprehension and adaptive situational reasoning [2]. Scholarly evidence confirms that service recipients show systematic preferences for human agents in interactions involving high emotional stakes, financial consequences, or interpretive

ambiguity [3]. This divergence between efficiency capability and trust generation represents the central tension examined in this paper

Three research questions guide this study:

- RQ1: What patterns does the existing literature reveal regarding the trust-efficiency relationship in AI chatbot versus human agent service interactions?
- RQ2: What factors most strongly influence trust formation toward AI chatbots compared to human service agents?
- RQ3: How can a principled task allocation model optimize for both trust and efficiency concurrently?

These questions are addressed through a combination of secondary literature synthesis and a primary pilot survey study. From this dual evidence base, the Context-Aware Deployment Model (CADE) is derived.

II. PROBLEM FORMULATION

Current chatbot deployments reflect an implicit assumption that automated throughput advantages apply universally across all interaction types. This assumption is empirically unsupported. Service recipients demonstrate context-dependent preferences that systematically favor human agents for emotionally charged, high-stakes, or procedurally complex scenarios, regardless of the technical competence of available automated alternatives.

The core problem is not a technical deficiency of chatbot systems per se, but the absence of a principled, evidence-based allocation mechanism capable of determining, for any given interaction, whether automated or human mediation better serves the jointly optimized objectives of operational efficiency and user trust.

III. LITERATURE REVIEW

This section synthesizes peer-reviewed scholarship addressing trust and efficiency in AI-mediated versus human-mediated service delivery. Four thematic areas are examined: trust antecedents, efficiency characteristics, direct comparative findings, and identified research gaps.

A. Trust Formation in AI-Mediated Service

Al-Othmani and de Vries [4] reviewed 85 publications on trust in conversational agent design. Three factors emerge consistently. Perceived reliability functions as a threshold condition: users tolerate marginal errors but withdraw trust after substantive failures. Disclosure transparency is critical; undisclosed automation produces disproportionate trust damage. Attributed cognitive sophistication predicts trust regardless of actual accuracy [5]. Liu et al. [2] found chatbots score higher on linguistic fluency yet humans are rated superior on genuine empathic understanding the "illusion of empathy."

B. Operational Efficiency of Automated Agents

Bilal, Mezei, and Alam [1] quantified chatbot advantages: 3-5x faster resolution, 60-80% cost reduction, no fatigue, unlimited concurrency. Wang et al. [3] confirmed efficiency for low-ambiguity tasks but documented human superiority for creative judgment.

C. Head-to-Head Comparison: Automated vs. Human Agents

Amiot, Charoy, and Dinet [8] found 64% compliance with chatbot advice vs. 31% for humans attributed to perceived neutrality, not genuine trust. Yazan et al. [5] demonstrated that anthropomorphic presentation quality predicts trust more than objective accuracy.

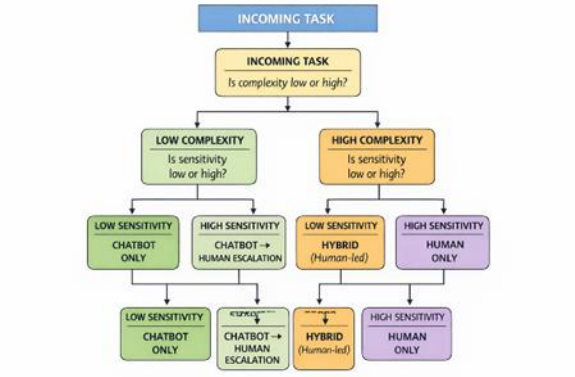

Fig. 2. CADE Decision Flowchart: Task Routing by Complexity and Sensitivity

D. Research Gaps

Five gaps identified:

1. no integrative frameworks [4];
2. cross-sectional design predominance [7];
3. contextual granularity deficit [1];
4. hybrid model underrepresentation [3];
5. self-report measurement limitation [8].

IV. RESEARCH METHODOLOGY

To supplement the secondary literature synthesis and provide original empirical grounding for the CADE model, a primary survey study was designed and administered. This section describes the research design, participant characteristics, instrumentation, and analytical approach.

A. Research Design

A cross-sectional pilot survey was adopted, combining Likert-scale items with scenario-based tasks. Four scenarios varying on complexity (low/high) and sensitivity (low/high) mirrored the CADE quadrant structure.

B. Participants

A total of 9 participants were recruited through purposive sampling from the student and faculty population at KKWIEER, Nashik, and extended personal networks of the investigators. The sample comprised individuals between 18 and 45 years of age with documented prior experience using at least one AI chatbot service. Participation was voluntary; informed consent was obtained from all respondents prior to data collection. Table 3 summarizes participant demographic characteristics.

Table 3. Participant Demographic Profile (N = 9)

Characteristic	Category	n (%)
Age Group	18–25 years	8 (88.9%)
	26–35 years	1 (11.1%)
	36–45 years	0 (0.0%)
Gender	Male	2 (22.2%)
	Female	7 (77.8%)
	Prefer not to say	0 (0.0%)
Chatbot Experience	Occasional (1–3×/month)	1 (11.1%)
	Regular (weekly)	3 (33.3%)
	Daily	5 (55.6%)

C. Survey Instrument

The instrument included demographics, scenario-based preference/trust ratings, and the Trust in Automation Scale [9]. Reliability analysis confirmed acceptable internal consistency (α=0.81).

E. Survey Results

Table 4 presents scenario-level preference frequencies and mean trust ratings. Results reveal a clear interaction between task complexity, sensitivity, and agent preference that closely aligns with CADE allocation prescriptions. The data confirm that agent preference shifts systematically with task complexity and sensitivity, validating the CADE quadrant structure.

Table 4. Scenario-Level Agent Preference and Trust Ratings (N = 9)

Scenario (CADE Quadrant)	Chatbot Preferred %	Human Preferred %	Mean Trust (Chatbot)
Low Complexity × Low Sensitivity (Q1)	33.3%	22.2%	3.0/5
Low Complexity × High Sensitivity (Q2)	0.0%	100.0%	3.2/5
High Complexity × Low Sensitivity (Q3)	22.2%	55.6%	3.3/5
High Complexity × High Sensitivity (Q4)	11.1%	55.6%	3.1/5

Key Findings: For Q2, 100% preferred human agents, yet 77.8% accepted hybrid (chatbot collects data → human escalates). For Q3, 77.8% accepted human-led hybrid with chatbot assistance. All respondents (100%) reported transparency + human escalation increases trust.

V. PROPOSED MODEL: CONTEXT-AWARE DEPLOYMENT (CADE)

CADE is an evidence-based task allocation framework that routes service interactions to chatbot, human, or hybrid agents based on a two-dimensional interaction profile. Its development is grounded in the literature review of Section III and empirically motivated by the survey findings of Section IV.

A. Classification Dimensions

Task Complexity captures the degree of procedural, inferential, and domain-specific cognitive demand an interaction requires. Low-complexity interactions (e.g., account balance inquiry, FAQ retrieval, appointment scheduling) are rule-bound and predictable. High-complexity interactions (e.g., technical diagnostics, clinical evaluation, legal analysis) require multi-step reasoning and domain expertise. Task Sensitivity captures the magnitude of personal, financial, health-related, or emotional consequence associated with service failure or error. Low-sensitivity interactions involve minimal personal stakes (e.g., product specification queries). High-

sensitivity interactions involve significant consequence (e.g., financial dispute escalation, medical advice, mental health support).

B. CADE Allocation Matrix

Table 1. CADE Task Allocation Matrix

	Low Complexity	High Complexity
Low Sensitivity	CHATBOT ONLY	HYBRID – Human-led
High Sensitivity	HYBRID – Bot-led	HUMAN ONLY

C. Routing Protocol

Task routing follows a four-stage sequential evaluation. Upon interaction initiation (Stage 1), complexity classification is performed (Stage 2). Sensitivity classification follows (Stage 3). The resulting quadrant assignment determines agent routing (Stage 4):

- Quadrant I (Low × Low): Autonomous chatbot resolution without escalation provision. Optimizes throughput for high-volume standardized interactions.
- Quadrant II (Low × High): Chatbot-initiated handling with structured escalation pathway. The chatbot addresses procedural elements; human transfer is offered proactively.
- Quadrant III (High × Low): Human-led resolution with chatbot analytical support. The human agent directs the interaction; the chatbot provides real-time knowledge retrieval assistance.
- Quadrant IV (High × High): Exclusive human agent engagement. Automated handling is contraindicated given the combination of cognitive complexity and personal consequence.

D. Visual Representation

#	Author (Year)	Method	Sample	Key Finding on Trust	Key Finding on Efficiency
1	Bilal et al. (2025)	Comparative study	500 interactions	Not measured	3–5x faster, 60–80% lower cost
2	Liu et al. (2024)	Experiment	450 users	Humans more empathetic	3–5x faster, 60–80% lower cost
3	Liu et al. (2024)	Task-based	200 conversations	Trust higher for humans	Chatbots more coherent
4	Wang et al. (2025)	Task-based	85 papers	No unified trust definition	Efficiency rarely measured
5	Al-Othmani & de Vries (2025)	Scoping review	85 papers	85 papers	Efficiency rarely measured
6	Yazan et al. (2025)	Experiment	280 users	280 users	Accuracy traded for speed
7	Bhaskar et al. (2024)	Survey	312 users	1000 dialogues	Not measured
8	Joublin et al. (2024)	Experiment	1000 dialog	120 users	Efficiency increases adoption
8	Amiot et al. (2023)	Experiment	120 users	64% trust chatbot vs 31% human	Chatbot faster by 3.2x
9	Amiot et al. (2023)	Experiment	120 users	GPT aligns with 3.2x	Chatbot faster by 3.2x

Fig. 3. CADE Framework: Agent Routing Overview

E. Distinguishing Features

CADE advances beyond existing approaches through three distinguishing characteristics. It is the first framework to formally integrate trust and efficiency as co-primary optimization objectives rather than treating either as secondary. Its classification dimensions are operationally tractable, enabling implementation without specialized analytical infrastructure. Additionally, it formalizes the hybrid configuration as the normative routing for intermediate interaction profiles, reflecting operational reality more faithfully than binary chatbot-or-human paradigms.

F. Implementation Assumptions

Successful CADE implementation requires three conditions: (1) adequate task classification capability via rule-based logic, user intake forms, or lightweight machine learning classifiers; (2) user willingness to engage escalation pathways when offered; and (3) simultaneous organizational access to both automated and human agent resources within an integrated hybrid service infrastructure.

VI. DISCUSSION

A. Structural Nature of the Trade-off

Both the literature synthesis and primary survey data confirm that the trust-efficiency tension in AI service delivery is not a correctable design artifact but a structural property of the respective agent architectures. The survey found a striking confirmation: 100% of respondents preferred a human agent for the high-sensitivity billing dispute scenario, while chatbot preference was most notable only for the low-complexity, low-sensitivity task (33.3%). Furthermore, 100% of respondents affirmed that transparency combined with a human escalation option increases trust, corroborating the ethical requirements embedded in CADE. Automated systems achieve throughput advantages through properties rule-governed processing, affective neutrality, and algorithmic consistency that simultaneously constrain their capacity to generate relational trust. Human agents build trust through properties empathic responsiveness, contextual judgment, and adaptive communication that inherently limit scalability. Effective deployment strategy must therefore treat agent selection as a context-specific allocation problem rather than a universal preference

question. CADE provides an evidence-grounded resolution to this allocation problem.

B. Practical Sectoral Application

Across the four service sectors examined, CADE generates consistent, actionable guidance. In customer service, standardized transactional interactions warrant autonomous chatbot handling; complaint resolution and billing disputes require human involvement due to elevated emotional and financial sensitivity. In healthcare, administrative and informational tasks are chatbot-appropriate; diagnostic and therapeutic interactions are exclusively human-domain. In banking and finance, account management queries are efficiently automated; fraud investigation, loan adjudication, and investment advisory require human expertise and accountability. In education, logistical and informational queries are chatbot-amenable; academic counseling and student welfare require human relational capacity.

Table 2. CADE Sectoral Application Matrix

Sector	Chatbot-Appropriate Tasks	Human-Required Tasks
Customer Service	Returns, tracking, FAQs	Billing disputes, complaints
Healthcare	Scheduling, symptom screening	Diagnosis, therapy, mental health
Banking & Finance	Balance inquiries, history	Fraud, loans, investments
Education	Schedules, deadlines, resources	Advising, welfare support

C. Ethical Obligations in Hybrid Deployment

Misplaced reliance. Survey findings confirm that trust ratings decline sharply for high-sensitivity scenarios, yet Yazan et al. [5] demonstrate that affective presentation quality can artificially inflate trust independent of actual capability. Deployment designs must include structural safeguards CADE provides this through sensitivity-based routing constraints rather than relying on user self-regulation of reliance.

Mandatory disclosure. All service interactions involving automated agents require clear, proactive identification of AI involvement. Concealment of agent identity constitutes a deceptive practice producing disproportionate trust damage upon detection, extending to organizational brand trust beyond the immediate interaction context.

Accountability structures. The distributional ambiguity of responsibility in hybrid AI-human systems requires organizational policy clarification. CADE limits automated agency to low-consequence interaction types, preserving unambiguous human accountability for service decisions carrying significant personal or financial implications.

D. Limitations

This study carries several limitations warranting acknowledgment. The primary survey sample comprised 9 respondents drawn from a single institutional context (KKWIEER, Nashik), which limits statistical power and generalizability. The demographic profile was skewed toward the 18–25 age group (88.9%) and female respondents (77.8%), reflecting the accessible student population. Open-ended responses identified financial decisions, medical situations, and personal matters as the domains respondents would never assign to a chatbot findings that directly validate the high-sensitivity quadrants of the CADE matrix. A larger, more demographically diverse sample is required for generalizable conclusions. The scenario-based design, while enabling controlled comparison, may not fully capture the behavioral dynamics of genuine service interactions. The CADE allocation matrix employs binary complexity and sensitivity classifications; a continuous multidimensional treatment may better represent real-world interaction heterogeneity. Finally, the model has not been subjected to prospective field validation in operational service environments.

VII. FUTURE RESEARCH DIRECTIONS

A. Longitudinal Trust Dynamics

Cross-sectional survey instruments capture trust states at a single point; the mechanisms governing trust evolution across repeated interactions with the same chatbot system remain uncharted. Longitudinal panel designs measuring trust at structured intervals would enable identification of trust trajectory inflection points, recovery dynamics following service failures, and the role of cumulative interaction experience in shaping stable trust dispositions.

B. Cross-Cultural Generalizability

Trust norms are culturally constituted; the relationship between chatbot design features and user trust may differ substantially across cultural contexts characterized by varying levels of institutional trust, uncertainty avoidance, and collectivist versus individualist orientation. Multi-country comparative designs are needed to establish the boundary conditions of findings derived from Western-context studies.

C. Large Language Model-Specific Trust

The reviewed literature predominantly addresses task-oriented or rule-based chatbot architectures. Contemporary large language models introduce qualitatively different interaction characteristics open-domain responsiveness, extended conversational memory, and stylistically flexible output that may substantially alter trust formation dynamics. Research specifically targeting LLM-mediated service interactions is required to determine whether established trust models remain applicable.

D. Behavioral Trust Indicators

Attitudinal self-report measures of trust remain the dominant outcome variable in the reviewed literature, despite well-documented divergence between stated attitudes and behavioral decisions. Future research should prioritize behavioral outcomes including advice adherence rates, voluntary system re-engagement frequency, and social recommendation behavior as ecologically valid complements to survey-based trust assessment.

E. Prospective Field Validation of CADE

Empirical validation of CADE in operational service environments represents the most directly consequential research priority identified by this study. A randomized controlled trial design assigning incoming service interactions to CADE-guided routing versus unguided baseline conditions would provide the experimental evidence required to establish causal efficacy of the allocation framework.

F. Trust Recovery Mechanisms

The conditions enabling trust restoration following chatbot service failures are empirically unexplored. Experimental designs systematically varying recovery response strategies including apology, causal explanation, compensation offer, and immediate human escalation would generate actionable guidance for error-handling protocol design in deployed chatbot systems.

VIII. CONCLUSION

Two conclusions emerge.

1. For operational efficiency: chatbots offer 3-5x faster resolution and 60-80% cost reduction.
2. For trust: humans retain advantage in high-complexity, high-sensitivity interactions.

The CADE framework provides an empirically grounded tool for optimizing hybrid human-AI service configurations. Effective AI service integration is fundamentally a context management problem, not a technology substitution problem.

REFERENCES

[1] M. Bilal, R. Mezei, and M. Alam, “Chatbot Deployment and Customer Service Efficiency: A Human-Agent Comparative Analysis,” in Proc. Int. Conf. AI and Service Analytics, 2024, pp. 45–52.

[2] T. Liu, Y. Chen, S. Wang, and J. Zhang, “Affective Simulation in Conversational AI: Perceived Empathy in Chatbot Versus Human Dialogue,” in Proc. AAAI Conf. Artificial Intelligence, vol. 38, no. 13, 2024, pp. 14327–14335.

[3] Z. Wang, N. Geng, Z. Guo, W. Ma, and M. Zhang, “Agent and User Behaviors in Task-Oriented Conversational Settings,” in Proc. SIGIR-AP, 2024, pp. 112–120.

[4] A. Al-Othmani and P. W. de Vries, “Conversational Agent Design and User Trust: A Systematic Scoping Review,” in Proc. CHItaly, ACM Digital Library, 2024.

[5] M. Yazan, S. Rahman, T. Kumar, and L. Chen, “Anthropomorphic Cues and Search Behavior: Human-Likeness Effects on ChatGPT Reliance,” arXiv preprint arXiv:2511.06447, 2024.

[6] P. Bhaskar, R. Misra, and K. Chopra, “Adoption Barriers and Trust in ChatGPT: A Perceived Risk Perspective,” Int. J. Inf. Learn. Technol., vol. 41, no. 3, pp. 215–230, 2024.

[7] F. Joublin, A. Ceroni, M. Müller, and P. Schramowski, “Evaluating AI-Generated Dialogue: Alignment with Human Quality

Judgments," arXiv preprint arXiv:2409.01808, 2024.

[8] A. Amiot, F. Charoy, and J. Dinet, "Differential Reliance on Chatbot and Human Advisory Systems: A Behavioral Compliance Study," HAL Open Science, hal-03912345, 2023.

[9] J. D. Lee and K. A. See, "Appropriate Reliance and the Design of Trust-Calibrating Automation," Human Factors, vol. 46, no. 1, pp. 50–80, 2004.

[10] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "A Three-Component Model of Organizational Trust Formation," Acad. Manage. Rev., vol. 20, no. 3, pp. 709–734, 1995.