International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Conversational AI Agents for Financial Operations with Escalation-Aware Handoff Protocols: Designing Intelligent Human-AI Collaboration Systems

Gautham Paspala

Submitted:19/02/2026 **Revised**: 01/04/2026 **Accepted:** 11/04/2026

Abstract: Conversational artificial intelligence (AI) provides a model shift from deterministic rule-based process automation to context-aware, always-on learning systems for financial operations. Toward that goal, this article presents a framework for escalation-aware conversational AI in financial operations, including a multi-dimensional signal architecture that leverages linguistic, behavioral, transactional, and relationship signals to make real-time, probabilistic escalation decisions for customers and service agents of financial institutions. Another key concept is the collaboration zone, where artificial intelligence and a human agent are processing in parallel, having distinct skills, and there is no explicit handoff of control between the agents. The curriculum builds on the human agents' reasoning to discover human-like reasoning paths and extend the AI competency frontier. It uses a high rate of automation while also ensuring highly satisfactory customer experiences similar to those of human agents. Other considerations include implementation architecture; the transformation of the workforce; QA and continuous improvement operations; as well as quests for proactive engagement, multimodal interaction, and federated learning; as well as the evolution of autonomous agents.

Keywords: *Conversational Artificial Intelligence, Escalation-Aware Handoff, Human-AI Collaboration, Financial Customer Service, Competency Boundary Expansion*

1. Introduction

The financial services industry interacts with customers billions of times per year through customer service interactions such as checking a bank account balance, disputing a transaction, answering a credit card payment question, and modifying a financial product. The economics of these interactions create a tension in financial services: customers want fast, expert, personalized 24x7 service, and the economics of a fully human-staffed service organization have become increasingly unsustainable. For a major retail bank handling tens of millions of customer contacts, with fixed costs that are reduced and profit margins spread thinner with more customers, structural incentives to automate routine service are hard to ignore.

The first-generation automated response systems, such as IVR systems and rule-based chatbots, have reduced costs at the cost of customer experience, being effective for simple and clearly defined tasks such as balance inquiries and checking branch location, but not so effective when conversations stray from the defined path. For banking chatbots, low adoption, privacy concerns, and the loss of human touch illustrate the main pitfalls that the literature discusses. Both lack of customization of keyword-triggered escalation and improper handling of conversation history are also drawbacks, as they create a binary world where automation either takes over the whole interaction or hands it off to humans without any additional context, forcing customers to repeat instructions.

With LLMs and conversational AI, there is again an opportunity to create systems that can have a natural conversation, that are contextually aware, and that understand the subtleties of the customer's intent, and customers are beginning to expect these capabilities. According to Deloitte's 2025 Connected Consumer Survey of approximately 3500 U.S. consumers, the percentage of consumers

Independent Researcher, USA

who are trialing or regularly using generative AI has increased from 38% in 2024 to 53% in 2025. This suggests that generative AI has hit an inflection point and will be ubiquitous in financial and technology services. But there are challenges to deploying these capabilities in financial services because the conversations are much closer to compliance, fraud, trust, relationships, and human judgment.

Applications, such as when the user's context requires empathy, e.g., when customers are going through financial hardship, or if the chatbot agent is expected to act on behalf of the user, as in the case of suspected elder financial exploitation, legally or ethically require a human-centered solution. For banking chatbots, Grosswieser et al. (2025) found in two studies with a total of 1931 participants that chatbot acceptance is determined by relevance and enjoyment rather than perceived usability [1]. Simultaneously, Deloitte (2025) highlights that 69% of surveyed consumers feel innovation is advancing too quickly, without sufficient attention to risk mitigation—a consideration that is equally relevant for financial institutions balancing automation ambitions with customer trust [2].

The research question is consequently not whether AI should be responsible for all interactions or none of them, but rather how to design systems that decide when and how to handle interactions with the smart collaboration of AI systems and/or human operators. An extended framework for escalation-aware conversational AI in financial operations is presented, along with the concept of a collaboration zone and the learning mechanisms that will progressively extend the performance horizon of the system.

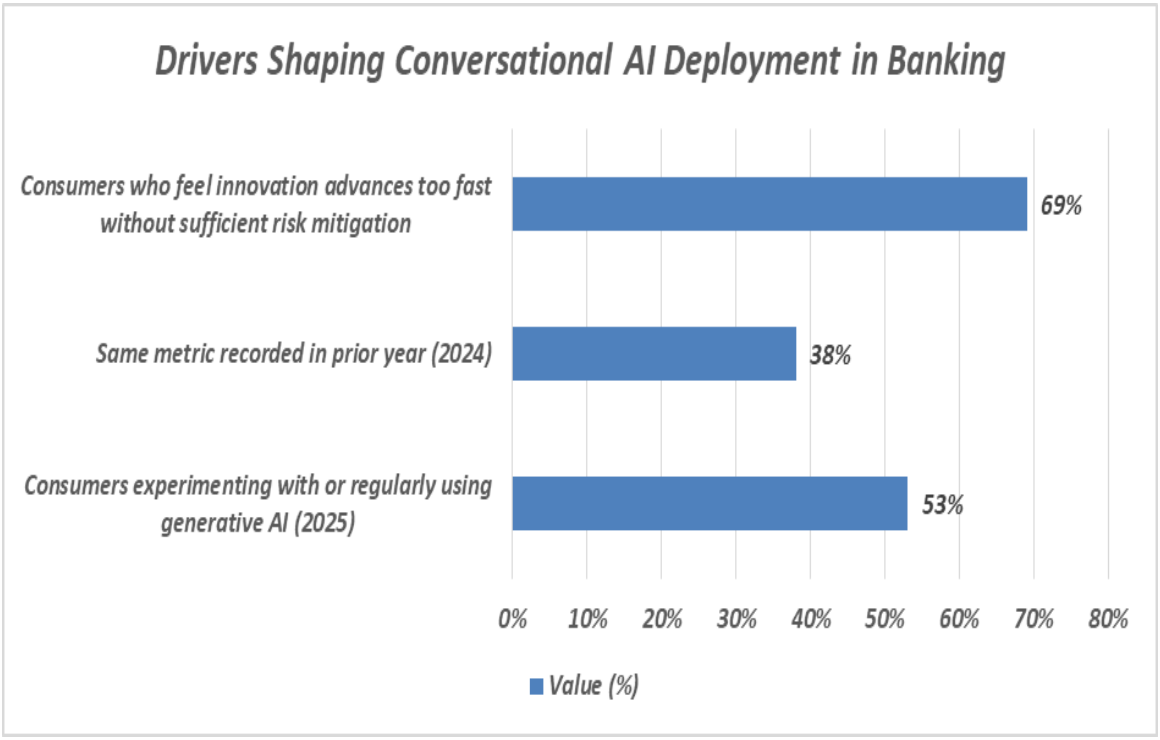

Figure 1: Drivers Shaping Conversational AI Deployment in Banking [1, 2]

2. The Limitations of Current Automation Models

2.1 The Binary Automation Trap

The customary customer service automation structure is that a customer either finds an answer with automated self-service or, if no automated self-service is available, the entire conversation is escalated to a human agent. Worse, these attempts to fit the task to the automation shape service objectives in ways that are unsuccessful and unsatisfactory to most customers. For example, chatbots cycling through solution options that fail to address a subtle dispute are frustrating service experiences and eroding trust in institutions. However, research indicates that such complaints damage brand equity more than wait times for human access, since consumers treat failed

automation as an indication of institutional indifference to customers rather than technological failure [3].

Automation transfer is most effective when automation is a good fit. It is most effective given simple phrases such as "speak to agent" or "human" or expressions of frustration. This implies a huge number of conversations are routed to humans, in spite of the cost. Post-hoc analysis of conversations indicates a high proportion of conversations that have been escalated to human agents involve automation-solvable issues, hinting at the chronic gap between the capability of a technology and its on-the-ground operational performance. Technologies employed at an early stage of the technology life cycle underperform their theoretical potential, a well-researched phenomenon in the technology life cycle literature, and are thus applicable to first-generation financial service chatbots [3].

2.2 The Keyword Escalation Problem

Another example of first-generation automation limitations is rule-based escalation triggers, which may catch obvious escalation words or phrases but miss less obvious escalation indicators and create false positives, all of which overwhelm human agent queues. Missed escalations can occur if the customer uses a low-level and indirect statement to express distress, such as saying that the issue is "affecting my family." False positive escalations can get triggered if trigger phrases are used in benign contexts, causing unnecessary transfers and thereby wasting the potential for automation, as well as the time of human agents.

According to Xiao et al. (2024), who drew on an article on AI applications in China's large enterprises, the biggest user difficulty for clever customer service is answering questions in a standardized fashion, as reported by iiMedia Research; 59.1% of respondents said this was the single biggest friction point. This confirms that keyword-based systems struggle with human language variability [4]. Keyword systems may be easily gamed in this fashion, since customers can rapidly learn that certain phrases will result in a hand-off to a real agent, while confused customers whose keywords are non-normative will instead find themselves stuck in a loop with the robot.

2.3 The Context Continuity Gap

Even if escalation is appropriate, there will be a loss of levels of conversational, emotional, and intentional richness that can be achieved with a human agent, and the conversation might need to include diagnostic questions that may frustrate customers who are unhappy with the automated experience. Xiao et al. (2024) have noted that customer service staff turnover is within 50% every year, creating a structural cause for the context continuity problem, as the human agents with the most experience building and working with partial transferred context are lost [4]. Candidate solutions thus need systems that maintain full conversational state across agent transitions, a design requirement that first-generation automation systems were never intended to meet.

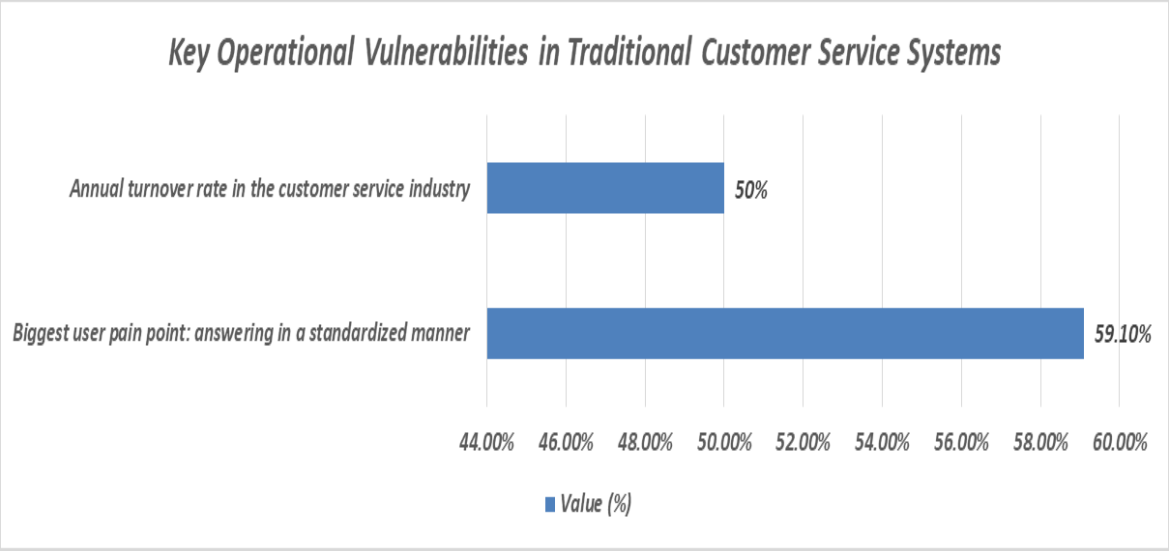

Figure 2: Key Operational Vulnerabilities in Traditional Customer Service Systems [3, 4]

3. Contextual Signal Architecture for Clever Escalation

3.1 Multi-Dimensional Signal Framework

Effective escalation strategies depend on linguistic, behavioral, and transactional as well as relationship-based signals, and the architecture presented in this article utilizes a multi-dimensional signal framework that continuously assesses when to escalate after each interaction. Linguistic signals expand beyond the presence of particular keywords to include signals of sentiment, frustration, complexity, and the drift in customer messaging patterns over time. Recent advances in sentiment analysis use lexicon-based approaches, machine learning classifiers, and transformer-based deep learning architectures. A customer emotion detection model based on transformers has an improved ability to sense subtlety in context that may be lost with other approaches [5]. The identification of changes in sentiment between turns of conversation enables tracking of when continued automation would damage the relationship with the customer.

Behavioral signals include changes in response times, the frequent repetition of certain words, and the distribution of conversation length around an optimal resolution time (see temporal resolution). Transactional signals carry information pertaining to the transaction, such as complexity, amount, importance, pattern of anomaly, and regulation. Relational signals describe account age and recent interactions, life transitions or events, and the relationship health and gross credit risk to the entity. This helps ensure that escalation decision-making is based on the customer relationship rather than the one specific interaction. Zhang et al. (2025) empirically showed that NLP-based smart customer service systems leveraging multi-dimensional contextual signals across linguistic sentiment, behavior, and profile space (i.e., customer profile changes or causal signals) outperform single-signal or keyword-based systems in customer satisfaction, underscoring the signal fusion architectural principle [6].

3.2 Escalation Probability Modeling

The signal framework feeds into an escalation probability model predicting whether human intervention would achieve better outcomes than automation alone. The model is a neural network architecture that is trained on past conversations labeled with one of four outcomes: successful automated resolution, failed automated resolution, successful escalated resolution, or unnecessary escalation. This four-class formulation allows the model to learn when escalation is actually helpful, allowing for a distinction between useful human intervention and over-escalation that creates friction. Rather than outputting a yes/no classification, the model outputs a probability, which can be thresholded to suit the operational context. As noted by Tan et al. (2023), aspect-based sentiment analysis, which identifies sentiment expressed towards aspects of the conversation (as opposed to a document-level polarity), provides considerably more informative features for downstream classifiers and is directly transferable to estimating escalation probability of multi-turn financial service conversations [5].

3.3 Specialized Escalation Triggers

Besides probabilistic modeling, the system also implements triggered interventions, where humans need to intervene in cases, regardless of the model output. High-risk areas of regulation (such as suspected elder abuse, requests for hardship accommodation, inquiries after the death of an account holder, or large fraud claims) and relationship-sensitive and safety-critical interactions have the same kind of exceptions to the policy. Zhang et al. (2025) further established that consumers who interacted with systems that employed transparent escalation paths (where transfer to human agents was on display and relevant) had considerably more positive perceptions of corporate image compared to consumers who interacted with systems that escalated arbitrarily or systems that did not escalate in genuine need, confirming that escalation design is inextricably linked to brand trust in financial service AI deployments [6].

4. The Collaboration Zone Architecture

4.1 Beyond Binary Handoff

The collaboration zone is the opposite of automation-only versus human-only. In the former case, explanatory interactions are simply passed on to humans when they become complicated, whereas in the collaboration zone, AI and humans combine the separate external and internal capabilities they have at their disposal. In this mode

the AI agent holds the conversation while a human agent ratifies the decision-making of the AI or interjects in instances in which human judgment or empathy is necessary. Marcineková et al. (2025) found that, in a study of an AI chatbot agent in micro-enterprise customer service, hybrid models of automation where chatbots automate low-level layers of interaction and humans supervise and/or escalate therefrom consistently outperformed purely automated or purely human implementations in continuity of service and perceived responsiveness by customers, showing the value of co-existing over solely cohabitating.

When the risk of escalation is between a soft and hard threshold, it is called the collaboration zone. This zone is where human values are allowed to remain relevant, but oversight is not considered an unreasonable burden. Examples include emotional yet simple problems, problems that will require multiple interactions where the AI can collect data while the human devises a resolution strategy, and customer-facing problems where more human oversight is considered desirable but not critical.

4.2 Demonstration of Interaction flow in collaborative mode

When human intervention is needed, i.e., the interaction reaches the collaboration zone, the system generates parallel processes for the AI agent and the human agent. The human agent is sent a collaboration request and receives the interaction context that includes the transcript, customer profile, issue classification, escalation reasons, and suggested replies. If accepted, the agent interface presents the live conversation alongside proposed AI-generated responses with confidence scores, and the agent can approve or reject the proposal or edit or replace its own response. Any change made by the agent is reported back to the system as feedback for future adaptations. The customer only has a conversation with the agent; there is no switch from collaborative to complete automation.

4.3 Authority to Bind and Override

An explicit delegation of authority to act permits collaboration, as AI autonomy is limited to typical information requests, simple procedural tasks, and low-stakes actions. High-stakes financial decisions, deviations from policy, and accepting commitments typically require human review due to their unpredictable risks. Human exclusive authority is for account closures, hardship accommodations, fraud determinations, and regulatory decision-making, regardless of whether the technical capabilities of AI are sufficient for such determinations. For Sahebi and Formosa (2025), AI use in finance creates a new epistemic trust problem: AI generation invisibility weakens the relationship authenticity required for ethical trust-cognizant communication with humans. Thus, explicit authority reservation (i.e., separation of AI and human roles) is an ethical requirement in high-stakes financial services environments [8].

4.4 Smooth Transition Management

Transitions between levels of automation require a handoff that is transparent to the customer. Upward transitions in the collaboration zone are naturally smooth since human agents join the AI and never directly interact with the customer, eliminating the need to restate previous context as in other models. Sahebi & Formosa (2025) further confirm that clear disclosure of AI involvement and of the availability of human oversight are foundational for maintaining epistemic trust in AI-mediated service interactions and that the design of transition communication processes is as consequential as the mechanism for transitioning [8].

Authority Level	Scope
AI Autonomous	Routine responses, balance inquiries, low-risk account actions
Human Approval Required	Financial impact actions, policy exceptions, commitments
Human Exclusive	Closures, hardship accommodations, fraud determinations
Collaboration Zone Trigger	Soft threshold exceeded; AI continues with human observation.
Full Human Takeover	Hard threshold exceeded; AI shifts to support role only
Transparency Disclosure	Templates support appropriate AI involvement disclosure

Table 1: Decision Authority Framework for AI-Human Collaborative Interaction Handling [7, 8]

5. Mechanisms for Learning and Competency

5.1 Reasoning Path Capture

Scaling up AI capabilities depends on not just capturing what decisions humans make, but also how they reason about those decisions. In addition to outcome supervision, capturing reasoning paths provides additional learning opportunities for the AI system. For resolved examples, the human annotator is then presented with a set of categories corresponding to the complexity, sensitivity, knowledge gap, and judgment scope of the example. The paper by Gómez-Carmona et al. (2024) argued that human annotators should play a more active role in human-in-the-loop machine learning processes, contributing to reasoning rather than just labeling, to create AI systems that learn human judgment rather than mere top-level outcomes. One implication is that, when building annotation interfaces for the AI in a financial service, the annotation interface should be part of the agents' workflow, such that the annotation interface defaults to the AI's suggestion (what caused the escalation) and the agent has to change the default only if it is incorrect [9].

5.2 Automated Learning Integration

The paths taken by the reasoning process are fed into automated learning processes, allowing the AI to learn new abilities without being programmed for each scenario. This is achieved through supervised learning on resolution patterns that indicate which situations the AI can and cannot handle. Informal human feedback is information about things or methods the AI does not currently possess. For example, humans may explain things that the AI cannot, and natural language understanding capabilities may harness this unexplained information to build knowledge bases. In contrast, policy learning identifies implicit rules and exception handling in human decisions that may be reviewed and formalized. Fox and Victores (2024) advanced safety science in human-AI systems and called for explicit governance of continuous learning loops between humans and AI to avoid behavioral loopholes, emergent phenomena where an AI system learns from human intervention (e.g., correction), but the governance over the learnings is unknown [10].

5.3 Competency Boundary Expansion

Learning mechanisms continue expanding the competency boundary, the set of scenarios the AI can handle without human assistance, by observing the scenarios escalated to humans and those automatically handled by the AI. If the probability of escalation to a situation type is reducing, the boundary of the competency is expanding. Expansions to the competency close the gaps through knowledge engineering, developing a capability, or adjusting a judgment model. Once the boundary has been validated through controlled testing, expansions to it should be locked in to the production system in order to avoid the negative impact of prematurely automated systems on customers' experiences. Gómez-Carmona et al. (2024) argue that the best human-in-the-loop systems are those where the boundary between machines and humans is treated as malleable, dependent on how capable the AI is found to be, rather than fixed at deployment time. This principle underlies the competency boundary expansion model described here [9].

5.4 Knowledge Feedback Loops

Learning requires feedback loops of AI performance, human feedback, and capability updates. Real-time feedback compares the suggestions made by the AI above human responses below. Outcome-based feedback examines whether the system's resolutions were true at the end of the interaction. Fox and Victores (2024) warn that feedback validation is a safety-critical function of human-AI systems. Short of validation, human-corrected feedback is unsafe, as errors from a single agent or idiosyncratic agent preferences may be incorporated into the AI system. Feedback filtering is a basic safety requirement for any financial services AI system with online learning [10].

6. Implementation Architecture

6.1 Conversational AI Platform Components

There are several key components in the platform that collectively deliver the capabilities described throughout this paper. The first building block is natural language understanding (NLU), which processes customer utterances to identify intent, entities, sentiment, and semantic meaning. This is effectively the perceptual layer upon which all

other building blocks rely. Bocklisch et al. (2017) present an open-source framework for building dialogue systems called Rasa. This is a modular system that decouples the natural language understanding and dialogue management into separate layers, which can be trained independently and can be trained to a specific domain, removing the need to retrain the whole system when one component is changed. Particularly with financial services products, the changes to products and regulations often require specialized knowledge updates [11]. Dialogue Management is an intermediary layer that manages conversation state and context and determines how to respond based on previous conversation and user intent. Conversational AI generates natural language responses balancing factuality, fluency, and brand voice. The action execution layer employs API calls to connect conversational AI to backend systems such as core banking systems, payment services, and service tools.

6.2 Escalation Intelligence Layer

The escalation intelligence layer implements the signal processing and decision logic for contextual handoff for each interaction. The signal extraction modules extract a feature set relevant to escalation from each conversation stream across linguistic, behavioral, transactional, and relational dimensions. The underlying escalation probability model is fed by the extracted features. It outputs a continuous value at each turn. That value can be processed by a thresholding procedure and interpreted as an escalation decision. The optimization procedures interpret the resulting outcomes and recommend configuration adjustments that optimize performance or output quality while obeying guardrails. Bocklisch et al. (2017) argue that production conversational AI systems require strong model versioning and evaluation infrastructure to support safe iterative improvement. This is also true for the escalation intelligence layer, where new models must be validated against historical outcome data before they are deployed in order to avoid regression in the quality of escalation decisions [11].

6.3 Collaboration Zone Platform

The collaboration zone platform enables simultaneous real-time work between AI and a human via control interfaces and workflow orchestration. The agent desktop allows for live conversations with proposed AI answers, approval controls, authority guidelines, and contextual summaries to reduce cognitive load on agent users. The AI suggestion engine generates proposals for responses and subsequent actions, including an optional confidence level. If uncertain, multiple proposals may be made to the user. Authority management systems gate approval-required actions until they are approved manually, either granting or denying these actions. These systems automatically produce an audit trail of all authority decisions.

6.4 Learning Platform Components

Along with the learning platform that allows for the accumulation of new data, model training, and capability deployment, a paper by O'Brien et al. (2022) found that the three most common and impactful forms of self-admitted technical debt in production ML systems are insufficient design documentation, inadequate testing infrastructure, and late refactoring of data pipelines in ML systems. These findings directly inform the learning platform design to avoid amassing undetected forms of technical debt over time that degrade overall system performance through investments in annotation pipelines, model versioning, and automated evaluation frameworks [12]. Analytics and monitoring capture automation rates, escalation patterns, outcome quality, and customer satisfaction. Such data can help identify anomalies before they cascade into systemic degradation.

Platform Component	Primary Function
NLU Layer	Extracts intent, entities, sentiment, and semantic meaning
Dialogue Management Layer	Maintains state and coordinates knowledge and action systems
Response Generation Layer	Produces fluent, compliant, brand-consistent outputs
Action Execution Layer	Connects AI to backend systems; logs all executed actions

Escalation Intelligence Layer	Outputs continuous probability scores with low-latency inference
Collaboration Zone Platform	Agent desktop with AI proposals, approvals, and override controls
Learning Platform	Annotation capture, model training, canary deployment, monitoring

Table 2: Integrated Technical Layers of the Conversational AI Implementation Architecture [11, 12]

7. Operational Considerations

7.1 Agent Workforce Transformation

To account for the impact of escalation-aware AI, the workforce would need to shift emphasis from transaction processing to being specialists in complex cases and managers of AI. This would allow humans to focus more on the sensitive and critical human relationships that automation cannot handle. These trends require continuously higher levels of judgment and soft skills and less procedural manual dexterity. Zirar et al. (2023) systematically reviewed the empirical literature on worker-AI co-evolution in workplaces and reached three conclusions on the skills to be developed and retained by workers in a co-evolving world: technical skills to work successfully with smart systems, human skills to work with other people (e.g., emotional intelligence and teamwork), and conceptual skills to participate in critical thinking and judgments. These are the exact skills required by financial service AI deployments of agents in collaboration zone workplace environments [14]. Therefore, training programs must develop these competencies explicitly, as a direct consequence of the operation of collaboration zone missions, the management of complex and sensitive cases, and de-escalation.

Staffing models must also account for changes to the workload. As the number of human-handled interactions decreases, the workload of the interaction becomes more intense (time and complexity). Performance metrics for such tasks are also shifting from being solely focused on efficiency to also incorporating details like resolution quality, customer satisfaction, and annotation quality for the contributions of human agents.

7.2 Customer Experience Management

Experience design enables a positive customer experience of human-AI work delegation. It should address transparency, expectation management, and feedback. For transparency policies, whether or when to disclose the involvement of AI is a particularly relevant decision. In combination with the desire to be informed, people want human oversight when interacting with AI. This is addressed with expectation management, i.e., letting customers know what they can expect automated handling to solve but that human agents are also available to assist with more complex issues. In a study of AI applications in financial services under the Industry 4.0 framework, Mhlanga (2020) cautions that AI-enabled customer service (chatbots and virtual assistants) can expand access to those who may not otherwise receive support, but only if consumers trust the technology and can connect with a human agent if technology fails [13]. The channel strategy must consider the differing capabilities of artificial intelligence; for instance, text-based channels usually provide higher-quality automation than voice-based channels due to differing levels of processing.

7.3 Quality Assurance and Compliance

Financial services AI has regulatory responsibilities and customer impacts; therefore, they are subject to formal quality assurance processes, providing assurance on the quality of responses, regulatory compliance, quality of escalations, and adherence to the audit trail. Mhlanga (2020) argues that financial service providers deploying AI should also tackle data quality issues, as the predictive capabilities and service quality of an AI system are dependent on the quality of the data going into it. This applies to annotation pipelines, interaction logging, and training data stewardship supporting escalation-aware AI over time [13]. .

7.4 Continuous Improvement Operations

Sectoral efficiency demands perpetual operational regimes dedicated to escalation patterns, model performance monitoring, unscheduled knowledge discovery, and customer feedback consumption. Workers, as Zirar et al. (2023) describe, in this perpetual race with workplace AI, must continue to upskill themselves beyond the current capabilities of clever systems. In a financial services context,

operationalizing this as a permanent cycle means humans are managing today's edge cases and training the AIs to handle tomorrow's edge cases automatically [14]. Financial services firms that can structure this cycle through repeated programmatic continuous improvement efforts, such as escalation reviews, recalibrating predictive models, and constantly backfilling new customer feedback, may benefit from more automation benefits than a one-off deploy-and-fix approach can provide.

8. Performance analysis and business impact

8.1 Automation Rate Achievement

The primary value of the system is to replace customer interactions that would have been conducted by a human agent. The "automation rate" is the proportion of customer interactions that were completed without any human involvement. For baseline automation rates for rule-based chatbots, they are bounded by the finite number of situations where these rules can be applied, limited by conservative escalation thresholds, and restricted when a customer requests a human agent when the chatbot is stuck. Scalability is considerably improved with mature escalation awareness, leading to almost twice the automation rate. This is accompanied by both increased autonomy over cases with mid-level complexity and a reduction of false escalation rates, contextual escalation triggers instead of keyword-based triggers, and the capacity to learn and grow. As noted by Rabhi et al. (2025), this has already been visible in the evolution of the AI landscape in financial services, where the move from process augmentation to end-to-end automation is already here, with AI systems monitoring transactional data in real time to identify patterns and making decisions that previously required human intervention, features that drive higher automation rates for escalation-aware architectures [16].

8.2 Cost Structure Transformation

Higher automation rates decrease per-interaction operating costs, but not in a linear fashion because of the collaboration zone and maintenance of the learning platform. Human-to-machine shifts have a direct impact on labor costs. This often leads to large per-interaction decreases at scale. The cost of a collaboration zone is higher than a no-handoff situation but not as high as pure human handling, and once the collaboration zone is up and running, the savings are compounded. According to Rabhi et al. (2025), automation of process discovery and optimization reduces the cost of operations and increases agility for organizations with AI-based BPM. Beyond labor costs, these benefits include lower rework costs, reduced resolution cycle times, and superior resource allocation within contact center operations [16].

8.3 Customer Experience Impact

As long as customer experience is retained, automation is valuable. For example, based on the J.D. Power 2025 U.S. Retail Banking Satisfaction Study, when the survey was conducted with 109,724 retail banking customers in the United States, overall customer satisfaction with primary retail banking partners was found to be 655 out of 1000 in 2025. The findings from the research also revealed that customer satisfaction improved during recessionary times when banks provided a combination of engagement, human interaction, and effective digital tools. The study found that 85% of customer problems were solved and were the major contributors to customer satisfaction scores in banking. Thus, whether issues are resolved through AI or human interaction, satisfaction at the relationship level is mainly concerned with issue resolution. Experience parity is feasible in typical situations, while escalation to humans for more complex, emotionally charged situations preserves satisfaction in cases involving human judgment or discretion.

8.4 Organizational Capability Development

Beyond efficiency, escalation-aware AI is a long-term planned investment for an organization because AI learning assets (models, knowledge bases, and interaction histories) are forms of institutional memory, which do not disappear with turnover or transitions in personnel. Workforce capability development builds human-AI specialists and problem solvers, whose skills remain valuable as AI capabilities improve. Rabhi et al. (2025) argue that organizations must leverage AI as a planned enabler of continuous innovation, rather than treating it as a one-off investment to improve efficiency. Through embedding AI/ML into their operations, organizations can create novel capabilities and sustain improvement initiatives over time [16].

Impact Metric Description	Value
Retail banking customer satisfaction score (J.D. Power 2025)	655 / 1,000 points
Customers whose problems were fully resolved (J.D. Power 2025)	85%
Retail banking customers surveyed in J.D. Power 2025 study	109,724

Table 3: Business Impact Dimensions of Escalation-Aware AI Deployment [15, 16]

9. Advanced Capabilities and Future Evolution

9.1 Proactive Engagement

The current focus is on customer-initiated contact. Future versions could include AI-initiated proactive contact, wherein the system looks to identify future needs before customers contact customer service. Anticipatory service is the practice of detecting when customers are likely to need service (for example, when unusual transaction patterns are detected, billing changes are confusing, or payments are due and the customer has little balance) and proactively contacting the customer to prevent needs from arising. In their bibliometric and content analysis of AI applications in finance, Bahoo et al. (2024) identify predictive analytics and proactive customer engagement among the most important emerging directions. The model shift from responsive to proactive service reflects a qualitative leap in the proposition of AI from being a cost-saver to being a genuine relationship enhancer [17]. Relatedly, personalized recommendations can ease service and relationship management that shows institutional attentiveness without taxing an institution's human agents, allowing for engagement-level replacements for relationship management that the size of a customer base would otherwise preclude.

9.2 Multimodal Interaction

Text-based conversations are the current focus, and voice and visual conversations will be supported in the future. Voice AI capabilities include automated conversations on the phone (with escalation-aware handoff) and voice signals such as tone, speech velocity, thinking pauses, and emotional tone, which supplement text signals to provide richer context for better decisioning. Visual AI capabilities include analyzing images and documents that customers share. Using images and documents reduces the friction of having to explain things orally. The Financial Stability Board (2017), in its original report on AI and machine learning in the financial services sector, predicted that expanding the modalities through which AI systems interacted with customers and processed information would progressively widen the range of service scenarios amenable to automation, and this is proving to be correct [18]. Multimodal signal fusion, combining cues refined through voice, text, and visual modalities, has proven to be well suited to escalation decisions.

9.3 Cross-Institutional Intelligence

Individual institution AI will be limited by the knowledge available at one institution, but cross-institutional intelligence, which learns from multiple institutions while safeguarding privacy, will rapidly accelerate the advancement of AI. Federated learning enables the model to learn global trends without the need for customer data transfer, allowing all players to contribute to both fraud pattern detection and model improvement. Bahoo et al. (2024) argue that cross-institution collaboration and information sharing are less understood frontiers in the field of financial AI. Areas with the largest unrealized value are ones where the data of one institution alone is not enough to detect rare but high-risk events [17].

9.4 Autonomous Agent Evolution

Interactive systems assist humans in the short term, but future systems might be able to act autonomously when properly constrained. From automated end-to-end interactions in well-defined scenarios through AI-to-AI negotiation in structured institutional contexts to a supervising AI guiding the actions of a human agent in an unpredictable environment, as reliability improves, the scenarios considered for human approval might widen. The FSB (2017) also noted earlier in relation to the evolution towards greater AI-enabled autonomy in financial services that it is important to ensure that regulatory acceptance,

institutional governance, and technical capability advance in parallel and not allow the technical capability to evolve faster than accountability [18].

Conclusion

For binary automation scenarios in financial customer service, the escalation-aware conversational AI framework is an advancement over customary rules-based keyword triggering to human agents through probabilistic multi-dimensional escalation intelligence and concurrent human and machine task collaboration. The architecture of the collaboration zone allows optimal levels of human involvement for the complexity and sensitivity of each case, replacing the trade-off between brittle full automation and inefficiency from excessive human involvement. As continuous learning systems treat every human decision as a training signal, they will progressively expand the artificial intelligence's ability, compounding the benefits of automation over time. In the new world, human agents will be valued for their judgment, empathy, and ability to build and maintain relationships. This will reduce operational cost and automation ratio without corroding customer satisfaction, making those gains sustainable. Furthermore, as AI technologies advance through proactive customer engagements, multimodal interactions, and federated learning, the institutions that embrace advanced escalation-aware architecture will realize a compounding competitive advantage in a world where financial services are increasingly intelligence-enabled.

References

[1] Hannah Grosswieser et al., "Chatbots in Banking: Key Predictors of User Acceptance Across Two Large-Scale Studies—and How Gender and Ease of Use Fall," ACM, 2025. [Online]. Available: https://dl.acm.org/doi/epdf/10.1145/3743049.3748557

[2] Steve Fineberg et al., "In the gen AI economy, consumers want innovation they can trust," Deloitte, Sep. 2025. [Online]. Available: https://www.deloitte.com/us/en/insights/industry/telecommunications/connectivity-mobile-trends-survey.html

[3] Amol C. Adamuthe et al., "An Empirical Analysis of Hype Cycles: A Case Study of Cloud Computing Technologies," IJARCCE, 2015. [Online]. Available: https://www.ijarcce.com/upload/2015/october-15/IJARCCE%2068.pdf

[4] Nan Xiao et al., "Changes and Applications of AI in the Customer Service Industry," ACM, 2024. [Online]. Available: https://dl.acm.org/doi/epdf/10.1145/3703187.3703197

[5] Kian Long Tan et al., "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," MDPI, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/7/4550

[6] Xiang Zhang et al., "A Study on Customer Satisfaction with NLP-Based Intelligent Customer Service Systems and Impact on Corporate Image," ACM, 2025. [Online]. Available: https://dl.acm.org/doi/epdf/10.1145/3778534.3778578

[7] Katarína Marcineková et al., "Implementing AI Chatbots in Customer Service Optimization—A Case Study in Micro-Enterprise," MDPI, Dec. 2025. [Online]. Available: https://www.mdpi.com/2078-2489/16/12/1078

[8] Siavosh Sahebi and Paul Formosa, "The AI-mediated communication dilemma: epistemic trust, social media, and the challenge of generative artificial intelligence," Springer Nature, Mar. 2025. [Online]. Available: https://link.springer.com/article/10.1007/s11229-025-04963-2

[9] Oihane Gómez-Carmona et al., "Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches," ScienceDirect, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542660523003712

[10] Stephen Fox and Juan G. Victores, "Safety of Human–Artificial Intelligence Systems: Applying Safety Science to Analyze Loopholes in Interactions between Human Organizations, Artificial Intelligence, and Individual People," MDPI, 2024. [Online]. Available: https://www.mdpi.com/2227-9709/11/2/36

[11] Tom Bocklisch et al., "Rasa: Open Source Language Understanding and Dialogue

Management," arXiv, 2017. [Online]. Available: https://arxiv.org/pdf/1712.05181

[12] David O'Brien et al., "23 Shades of Self-Admitted Technical Debt: An Empirical Study on Machine Learning Software," ACM, 2022. [Online]. Available: https://dl.acm.org/doi/epdf/10.1145/3540250.3549088

[13] David Mhlanga, "Industry 4.0 in Finance: The Impact of Artificial Intelligence (AI) on Digital Financial Inclusion," MDPI, 2020. [Online]. Available: https://www.mdpi.com/2227-7072/8/3/45

[14] Araz Zirar et al., "Worker and workplace artificial intelligence (AI) coexistence: Emerging themes and research agenda," ScienceDirect, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166497223000585

[15] J.D. Power, "Retail Bank Customer Satisfaction Surges As Banks Ramp Up Customer Support In Uncertain Economic Environment, JD Power Finds," Mar. 2025. [Online]. Available: https://www.jdpower.com/business/press-releases/2025-us-retail-banking-satisfaction-study

[16] Fethi Rabhi, "Editorial: Business transformation through AI-enabled technologies," Frontiers, Mar. 2025. [Online]. Available: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1577540/full

[17] Salman Bahoo et al., "Artificial intelligence in Finance: a comprehensive review through bibliometric and content analysis," Springer Nature, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s43546-023-00618-x

[18] FSB, "Artificial intelligence and machine learning in financial services," 2017. [Online]. Available: https://www.fsb.org/uploads/P011117.pdf