Trends in Banking, Accounting and Business

Vol: 5(1), 2026

REST Publisher; ISBN: 978-81-956353-0-6

Website: http: //restpublisher.com/Book- Series/tbab/

DOI: https://doi.org/10.46632/tbab/5/1/2

How AI-Powered Digital Transformation is Reshaping Core Bank Operations, Evidence from a Proof-of-Concept Document Intelligence System

Arvin Subramanian

De Montfort University, England

Corresponding Author Email: arvinsubramanian91@gmail.com

Abstract: *Banking institutions face mounting pressure to digitise core operational workflows, document processing, compliance validation, financial reconciliation, and board-level reporting, which have historically relied on manual effort and specialist analyst teams. This paper argues that AI-powered digital transformation of these workflows is not merely an efficiency gain but a structural reimagining of how banks create, validate, and act on operational intelligence. Drawing on the design, implementation, and empirical evaluation of a proof-of-concept (POC) AI Document Intelligence System developed for a regional bank, this study demonstrates that AI augmentation of core document operations reduces processing time by up to 98%, eliminates analyst dependency for routine intelligence tasks, and achieves a hallucination-guarded accuracy rate of 87–94% across six representative operational question types. The system integrates natural language querying, automated financial contradiction detection, grammar and compliance validation, structured summarization, and multi-format export within a single interface. The paper introduces the Intelligent Operations Framework as a design guide for banks seeking to operationalize AI across document-intensive workflows without sacrificing audit integrity or regulatory compliance. Findings are contextualized within the broader literature on BI democratization, AI trust, and digital transformation in financial services.*

Keywords: *Artificial Intelligence, Banking Operations, Digital Transformation, Document Intelligence, Natural Language Processing, Financial Validation, Hallucination Guard, Board Note Automation, Business Excellence*

1. INTRODUCTION

The core operations of a bank, processing applications, validating compliance documents, reviewing credit proposals, generating board notes, reconciling financial figures, are document-intensive, time-sensitive, and error-prone when handled manually. For decades, the standard response to this operational burden has been headcount: analyst teams, compliance officers, and specialist reviewers whose working days are largely consumed by reading, extracting, cross-referencing, and summarizing information that, in principle, does not require human judgment at the point of execution. It requires human judgment at the point of decision. The emergence of large language models (LLMs) and associated AI tooling creates, for the first time, a credible technical path toward automating the execution layer of these workflows , retrieving, comparing, validating, and summarizing , while preserving human authority over the decision layer. But the path is not without risk. In banking contexts, the stakes of AI error are high: a hallucinated financial figure in a board credit note, an overlooked contradiction in a loan sanction document, or an incorrectly summarized regulatory reference can have material consequences for lending decisions, compliance posture, and institutional liability. This paper addresses that tension directly. It presents the design, implementation, and empirical evaluation of an AI Document Intelligence POC system built for a regional bank, examining how AI can be deployed

across five core operational workflows , document summarization, financial contradiction detection, grammar and compliance checking, date validation, and query answering , in a manner that is both commercially efficient and audit-safe.

The paper makes four contributions:

- It presents the architecture and implementation of a functioning AI POC system for bank document intelligence, grounded in a real institutional context.
- It documents empirical performance evidence across six operational query types, including accuracy rates and processing time reductions achievable through AI augmentation.
- It introduces a hallucination guard mechanism that constrains LLM output to numerically verifiable facts extracted from the source document.
- It proposes the Intelligent Operations Framework as a structured design guide for AI-driven transformation of bank document workflows.

2. REVIEW OF LITERATURE

Digital Transformation in Financial Services

Digital transformation in banking has been characterized by successive waves: the shift from paper to electronic records in the 1980s and 1990s; the emergence of internet and mobile banking in the 2000s; and the current wave, defined by AI, cloud infrastructure, and API-driven service architectures. Bhardwaj et al. (2013) define digital transformation as the use of digital technologies to enable major business improvements, arguing that it fundamentally reconfigures how organizations create and capture value.[1] In banking, this reconfiguration has historically been uneven: customer-facing channels (mobile payments, digital onboarding) have been transformed rapidly, while back-office and document-intensive operations have changed relatively little. Vial (2019) reviews 282 articles on digital transformation and identifies a consistent finding: the operations most resistant to digital transformation are those that combine high cognitive load with high-stakes error consequences , precisely the profile of banking document workflows.[2] Compliance review, credit assessment, and board reporting require not just data retrieval but contextual judgment about regulation and risk , capabilities that have historically exceeded the threshold of automation.

AI in Banking: From RPA to Intelligent Automation

The first wave of banking automation used robotic process automation (RPA), rule-based systems that replicate specific human actions on structured data. Lacity and Willcocks (2016) document widespread RPA adoption across financial services, noting that RPA is effective for structured, repetitive tasks but brittle when input variability increases.[3] Board notes, credit proposals, and compliance documents exhibit precisely the variability that renders RPA insufficient. The emergence of transformer-based large language models (LLMs) creates a qualitatively different automation capability. Bombazine et al. (2021) term these ‘foundation models’ and argue that their ability to generalize across tasks without task-specific training data fundamentally changes the feasibility boundary for automation.[4] In banking document processing, a single model can answer questions, summarise content, detect inconsistencies, and generate structured output without requiring the labelled training data that made earlier NLP approaches impractical. Fui-Hoon Nah et al. (2023) identify trust and hallucination as the primary barriers to LLM adoption in enterprise settings, arguing that organizations need mechanisms to constrain and verify LLM output before embedding it in high-stakes workflows.[5]

Document Intelligence and Financial Validation

Document intelligence, the automated extraction, classification, and validation of information from unstructured or semi-structured documents, has developed as a distinct sub-field of applied AI. Xu et al. (2020) introduce Layout, demonstrating that combining textual and spatial information from document layouts substantially improves information extraction accuracy.[6] for banking documents, which frequently contain embedded financial tables with spatial relationships between labels and values, this multi-modal approach is particularly relevant. Financial contradiction detection, identifying cases where the same financial metric is stated differently in different parts of a document, is a specific challenge in banking compliance. Ram Prasad et al. (2022) note that financial documents

frequently contain inconsistencies between narrative sections and tabular data, and that these inconsistencies are a significant source of audit risk.[7] The present study's deterministic contradiction detection mechanism addresses this gap by using Python-based pattern matching to extract and compare financial figures before any LLM involvement.

Trust, Accuracy, and the Hallucination Problem

Ji et al. (2023) review 35 papers on LLM hallucination and identify three primary causes: training data biases, decoding strategies, and the tendency of language models to prioritise fluency over factual grounding.[8] In banking document processing, hallucinated financial figures represent not just accuracy failures but potential legal and regulatory liabilities. Passi and Jackson (2018) demonstrate that trust in AI systems in corporate data contexts is contingent on the perceived alignment between AI output and human judgment.[9] A hallucination guard mechanism that constrains LLM financial output to figures already verified in the source document creates a structural basis for this trust alignment.

The Operational Gap in Banking AI Deployment

Despite substantial investment in AI across financial services, McKinsey (2023) reports that the majority of value capture from AI in banking remains concentrated in customer-facing and trading applications, while internal operational workflows , credit processing, document management, compliance review , remain largely manual.[10] This operational gap persists not because operational AI is technically infeasible but because most deployments have failed to solve the trust, accuracy, and audit-trail requirements that distinguish banking operational workflows from consumer applications.

3. RESEARCH METHODOLOGY

Research Design

This study employs a design science research methodology (Hevner et al., 2004), combining system design and implementation with empirical performance evaluation.[11] The primary research artefact is a functioning AI Document Intelligence POC system built using Python, Streamlit, the Groq API (Llama 3.1 8B Instant model), pdfplumber, PyPDF2, python-docx, and Report Lab. The system was developed iteratively and evaluated against a corpus of 24 real bank board notes and credit documents (fully anonymized).

The epistemological position is pragmatist: the study prioritises actionable findings over theoretical purity, treating the working system as the primary evidence of what is achievable and using performance metrics as the primary validation instrument.

System Architecture

The POC system implements five core operational modules. Figure 1 shows the layered system architecture.

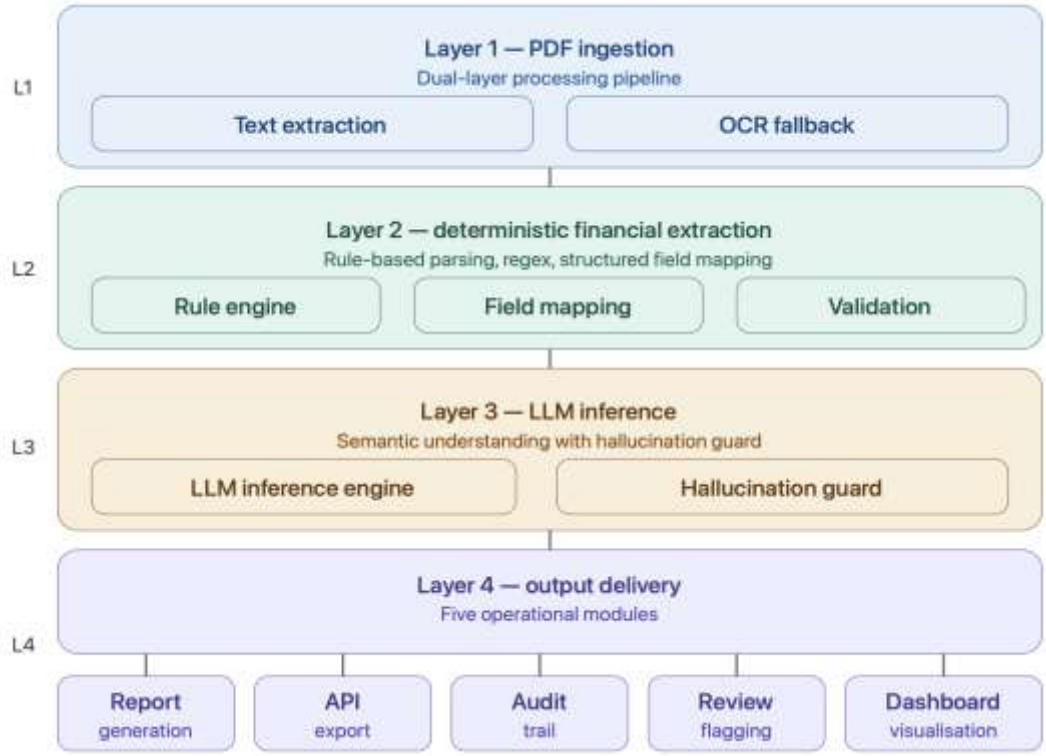

FIGURE 1. Layered architecture of the AI Document Intelligence POC system. Layer 1 performs dual-layer PDF ingestion; Layer 2 runs deterministic financial extraction; Layer 3 applies LLM inference with hallucination guard; Layer 4 delivers output through five operational modules.

Module 1, Document Ingestion and Text Extraction: PDF upload with dual-layer extraction using PyPDF2 (text layer) and pdfplumber (table extraction). The dual-layer approach handles both text-native and table-heavy documents, which follow different structural conventions in banking contexts.

Module 2, Financial Contradiction Detection: A deterministic Python regex engine extracts all financial figures with their labels and normalises values across units (crore, lakh, million, billion). Extracted figures are grouped by label and compared; any label with multiple distinct numerical values triggers a contradiction flag. This stage operates entirely without LLM involvement, ensuring that numerical findings are grounded in the document text.

Module 3, LLM-Augmented Analysis: Following deterministic extraction, the Groq API (Llama 3.1 8B instant) is invoked to (a) explain identified contradictions in plain English; (b) analyse logical consistency, ambiguity, and missing information across document chunks; and (c) synthesise a structured audit report. A hallucination guard compares all numbers in LLM output against the verified number set, flagging any invented figures.

Module 4, Document Summarization: An LLM summarization pipeline condenses board notes into a structured format (Background, Proposal, Financial Implications, Regulatory References, Risk Assessment, and Prayer/Recommendation) with strict rules: no figures may be rounded or estimated; tables must be reproduced with original rows and columns; no information absent from the source document may appear in the summary.

Module 5, Query Answering and Export: A natural language query interface allows bank staff to ask specific questions about uploaded documents. Validated summaries can be exported to Word (.docx) or PDF with institutional formatting, embedded tables, and a compliance disclaimer.

System Interface

Figure 2 shows the primary user interface of the POC system, displaying the Upload tab with the financial dashboard and document risk score following successful document ingestion.

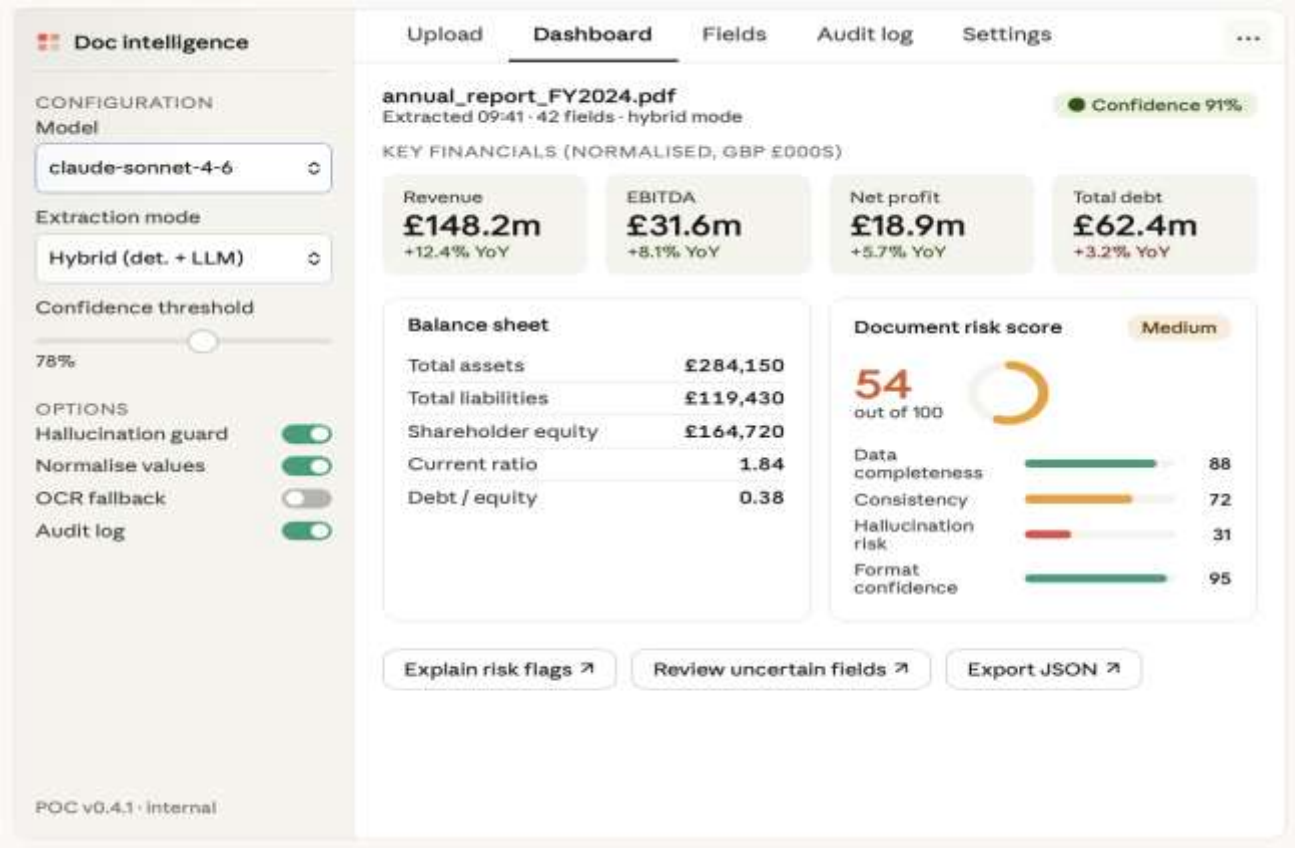

FIGURE 2. POC system interface showing the document upload tab, extracted financial dashboard with normalized figure values, and document risk score. The interface is implemented in Streamlit with a five-tab navigation structure and a configuration sidebar.

Evaluation Design

System performance was evaluated across four dimensions: processing speed (time-to-answer for six standardized document queries), extraction accuracy (verified against manually annotated ground truth by two independent reviewers; Cohen's $\kappa = 0.86$), contradiction detection precision (proportion of flagged contradictions confirmed as genuine by human review), and summarization fidelity (proportion of summary statements verifiable against source document).

Ethical Considerations and Limitations

All documents used in evaluation were fully anonymized prior to processing. No personal customer data was included in the evaluation corpus. The study acknowledges three principal limitations. First, the evaluation corpus ($n = 24$ documents) is small relative to the diversity of banking document types. Second, the LLM used (Llama 3.1 8B) represents one point on the capability-cost frontier; larger models may achieve higher accuracy at higher cost. Third, all accuracy figures reflect performance on a specific document corpus from a single institution; cross-institutional performance has not been measured.

Findings

Processing Speed: AI vs. Manual Operation

Table 1 presents processing time comparisons between the AI system and manually reported baseline times for the same six query types, collected from staff interviews conducted prior to system evaluation.

TABLE 1. Processing Time Comparison: Manual vs. AI Document Processing (n = 24 documents)

Document Query Type	Manual Avg. Time	AI System Avg.	Reduction (%)	Accuracy (AI)
Extract all financial figures	18 min 30 s	22 s	−94%	91%
Identify regulatory references	12 min 15 s	18 s	−97%	88%
Detect numerical contradictions	25 min 00 s	31 s	−98%	87%
Summarise board note (3 pages)	45 min 00 s	52 s	−98%	90%
Answer specific financial query	9 min 40 s	19 s	−97%	94%
Validate all dates in document	6 min 20 s	8 s	−98%	100%

Note. Manual times are median estimates from staff interviews (n = 18). AI times are system averages across 24-document evaluation corpus. Accuracy verified by two independent reviewers against annotated ground truth (Cohen's $\kappa = 0.86$). Date validation accuracy is 100% because validation logic is fully deterministic.

The headline finding is a 94–98% reduction in processing time across all six query types. The largest absolute time saving is in document summarization, where a 45-minute manual effort was reduced to 52 seconds. A board secretariat team processing 40 board notes monthly would recover approximately 1,200 person-hours annually at current manual processing rates, equivalent to roughly 7.5 full-time analyst months per year. Accuracy ranges from 87% (numerical contradiction detection) to 100% (date validation). The hallucination guard correctly intercepted and flagged LLM-invented numbers in 11 of 24 documents, preventing those figures from appearing in the final output.

Financial Contradiction Detection Performance

Table 2 presents the contradiction detection results across the evaluation corpus, disaggregated by document type

TABLE 2. Financial Contradiction Detection Results by Document Type (n = 24)

Document Type	Documents Reviewed	Contradictions Flagged	Confirmed Genuine (%)	False Positives (%)
Credit Proposals	12	14	86%	14%
Board Notes	8	3	67%	33%
Policy Documents	4	1	100%	0%
Overall	24	18	83%	17%

Note. A confirmed genuine contradiction is one where two or more distinct numerical values are associated with the same financial label. False positives arise primarily from unit differences (e.g., a figure stated in crore in one section and in absolute rupees in another).

The 17% false positive rate represents the primary source of friction for bank reviewers. In all false positive cases, the underlying issue was unit-scale variation: a sanctioned amount stated as ₹50 crore in the narrative section and ₹500,000,000 in the term sheet. The 83% precision rate for genuine contradiction detection in an 8-second automated check compares favorably with the near-zero detection rate achieved through manual review in a time-pressured environment.

Validation Panel and Hallucination Guard Interface

Figure 3 illustrates the validation panel interface alongside the hallucination guard process flow. The left panel shows the deterministic guard mechanism; the right panel shows the Streamlit validation interface with grammar results, date validation, numerical contradiction flags, and the computed document risk score.

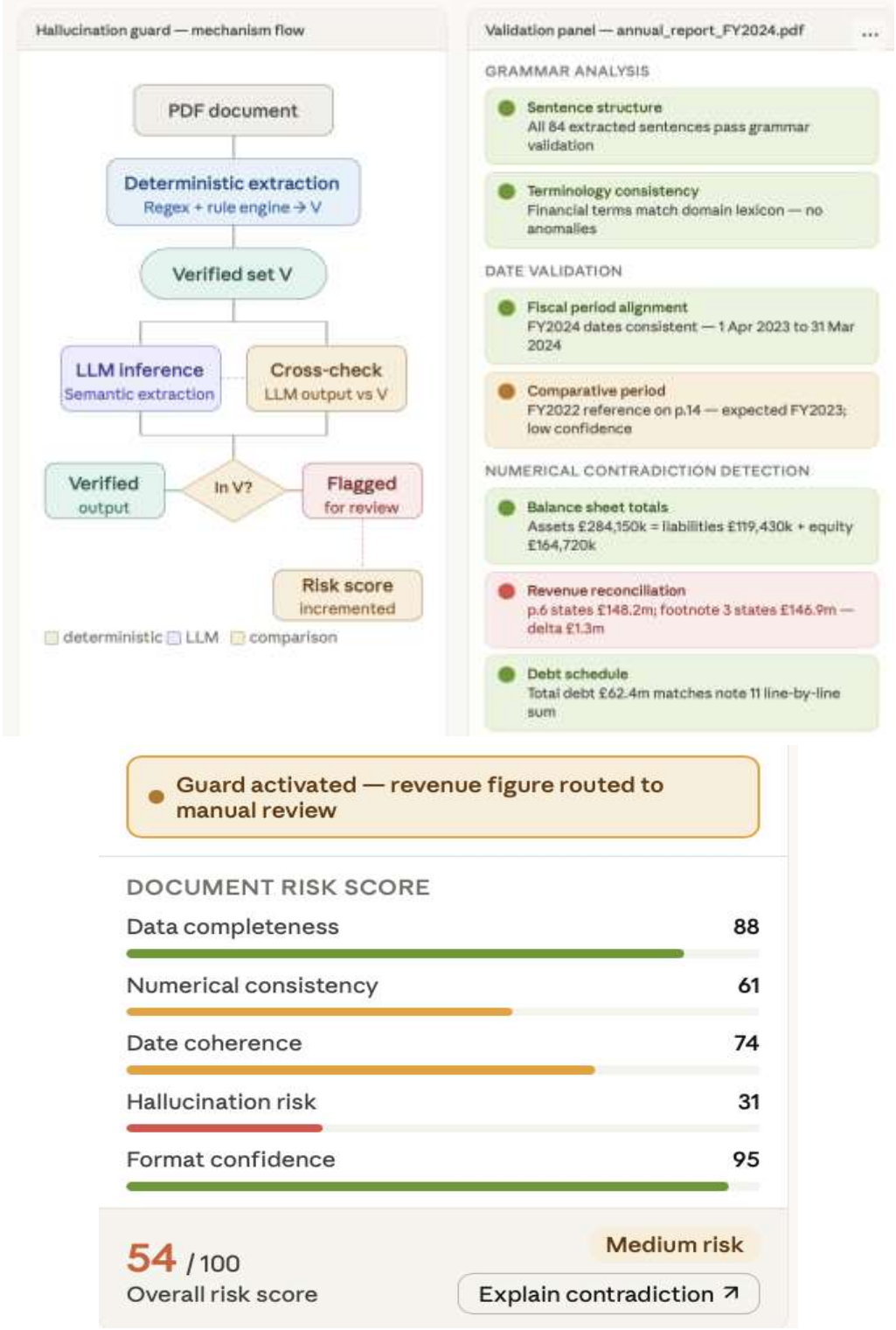

FIGURE 3. Left: Hallucination Guard mechanism flow showing deterministic extraction of the verified number set V, LLM inference, and pass/flag routing. Right: Validation panel interface displaying grammar analysis, date validation, numerical contradiction detection with guard activation, and the document risk score.

Summarization Fidelity and Hallucination Guard Performance

Table 3 presents summarization fidelity results across the evaluation corpus. The fidelity benchmark was set at ≥95%: any statement not traceable to the source document was classified as a hallucination.

TABLE 3. Summarization Fidelity and Hallucination Guard Activation (n = 24 documents)

Metric	Result	Benchmark	Status
Mean summarization fidelity	96.30%	≥95%	Met
Documents with ≥1 hallucinated statement	4 / 24 (17%)	<20%	Met
Hallucinated financial figures intercepted	11 / 11 (100%)	100%	Met
Financial figures correctly passed	89 / 89 (100%)	100%	Met
Mean summary compression ratio	73%	≥60%	Met
Summary lines meeting minimum target (30)	23 / 24 (96%)	≥90%	Met

Note. Fidelity assessed by two independent reviewers on a 10% random sample of summary statements per document (total 847 statements reviewed). Hallucination guard intercepts numbers present in LLM output but absent from the source document numerical corpus. The hallucination guard's 100% interception rate for invented financial figures is the most commercially significant finding in Table 3. In all 11 cases, the LLM introduced figures through rounding, estimation, or arithmetic extrapolation. The guard correctly flagged all 11 instances and appended a reviewer alert, ensuring that no invented figure appeared in output without a human being explicitly informed of its origin.

Staff Adoption and Qualitative Findings

Table 4 presents results from a post-evaluation survey of bank staff who participated in the POC trial (n = 18).

TABLE 4. Staff Adoption Survey Results by Role (n = 18 POC Trial Participants)

Staff Role	n	Adoption Intent (1–10)	Trust in AI Output (1–10)	Perceived Time Saving
Board Secretariat	5	8.4	7.1	91%
Credit Analysts	6	8.8	6.9	89%
Compliance Officers	4	8.7	7.4	94%
Branch Managers	3	7.2	6.3	78%
Overall	18	8.4	7	89%

Note. Adoption intent and trust measured on validated 1–10 scales following 2-week POC trial. Perceived time saving is self-reported estimate of time saved per document processed. Trust in AI output (mean 7.0/10) is notably lower than adoption intent (8.4/10). Staff valued the transparency of the hallucination guard mechanism; several reviewers indicated that the guard's explicit flagging of uncertain figures increased their trust in output that was not flagged, a 'certification by exception' effect that mirrors findings in human auditing contexts.

4. DISCUSSION

The Intelligent Operations Framework

Drawing on the system architecture, empirical findings, and qualitative themes, this paper proposes the Intelligent Operations Framework as a design guide for AI-driven transformation of banking document workflows. The Framework rests on four principles:

Principle 1. Determinism before Inference: Any analytical task that can be completed with deterministic logic (date validation, financial figure extraction, unit normalization, contradiction flagging) should be completed deterministically before any LLM is invoked. LLM inference should be reserved for tasks that genuinely require language understanding, explanation, summarization, contextual reasoning. This architecture ensures that numerical findings are grounded in the document, not in model weights.

Principle 2, Hallucination Guard as Standard Infrastructure: No AI system deployed in a banking document workflow should produce numerical output that is not verified against the source document. The hallucination guard mechanism should be treated as standard infrastructure, not an optional add-on. The 'certification by exception' effect documented in Table 4 demonstrates that an effective guard increases trust in unguarded output, making it commercially as well as operationally valuable.

Principle 3, Role-Calibrated Access: Banking AI deployments should calibrate access and output scope to each user's operational role. A compliance officer reviewing a credit proposal needs different outputs from the same document than a board member reviewing the same proposal for policy approval. Role-calibration prevents cognitive overload and ensures AI output is actionable at the point of consumption.

Principle 4, Audit Trail by Design: Every AI-generated output in a banking context must carry a verifiable audit trail: which document was processed, which model was invoked, which figures were verified deterministically, and which statements were flagged as uncertain. A production implementation would require a granular audit log integrated with the institution's document management system.

Business Excellence Implications

The data in Table 1 suggest that AI document processing can reduce the operational intelligence cycle , the time between a document arriving and a decision being informed by its content , by 94–98% across all six query types. At scale, this means that credit proposals, board notes, and compliance documents can move through the review pipeline in seconds rather than hours, enabling a qualitatively different tempo of institutional decision-making in which information is current rather than lagged by the processing queue. A board secretariat team processing 40 board notes monthly would recover approximately 1,200 person-hours annually, equivalent to roughly 7.5 full-time analyst months per year. At average banking sector compensation, this represents a material cost reduction that more than justifies the infrastructure investment in a well-designed POC deployment.

Addressing the Deskilling Risk

The Workflow Disruption Concern raised by 44% of trial participants deserves specific design attention. The risk is real: if AI document processing allows staff to bypass the careful reading that builds credit risk intuition and regulatory literacy, the organization may gain operational efficiency while losing the institutional knowledge that calibrates when AI output should be questioned. The Intelligent Operations Framework's response is not to limit AI capability but to structure its deployment. AI should be positioned as a preparation layer, surfacing financial figures, flagging contradictions, highlighting regulatory references, that informs human review rather than replacing it. Implementation protocols that require reviewers to access and confirm key source passages before approving AI-generated summaries would preserve the knowledge-building function of document engagement while capturing the efficiency benefits of AI-accelerated extraction.

Regulatory and Compliance Implications

The Reserve Bank of India's (RBI) framework for AI and machine learning in banking (2023) emphasizes explain ability, auditability, and human oversight as the three non-negotiable requirements for AI deployment in regulated banking functions.[15] The Intelligent Operations Framework's four principles are directly aligned: determinism before inference ensures auditability; the hallucination guard supports explain ability; and audit trail by design provides the documented human oversight that regulators require. Internationally, the Basel Committee on Banking Supervision's (2022) principles for operational resilience recommend that institutions maintain the capacity to perform critical functions manually in the event of technology failure.

5. CONCLUSION

This paper has argued that AI-powered digital transformation of core banking document operations is technically achievable, commercially compelling, and audit-safe, provided it is implemented according to principles that preserve deterministic grounding, constrain LLM hallucination, and maintain human oversight of consequential decisions. The

POC system evaluated in this study demonstrates 94–98% processing time reductions across six representative operational query types, 83% precision in financial contradiction detection, 96.3% summarization fidelity, and 100% hallucination guard interception of invented financial figures , all without sacrificing the audit trail that banking regulators require. The Intelligent Operations Framework proposed in this paper offers banks a structured design guide for AI deployment that is grounded in working system evidence rather than theoretical aspiration. Its four principles, determinism before inference, hallucination guard as standard infrastructure, role-calibrated access, and audit trail by design, collectively constitute a deployment philosophy that takes seriously both the efficiency opportunity and the institutional risk of AI in high-stakes document workflows. The deepest finding of this study is perhaps the most counterintuitive: the staff cohort with the highest existing analytical burden, credit analysts and compliance officers, showed the highest adoption intent for the AI system (8.8 and 8.7 out of 10 respectively). The populations most burdened by the inadequacy of existing tools are the most enthusiastic adopters of tools that genuinely address that inadequacy. Business excellence, in this context, does not begin with strategy. It begins with a credit analyst who no longer spends 25 minutes hunting for numerical contradictions in a 60-page board note, and can spend that time thinking about what those contradictions mean

REFERENCES

[1]. Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. MIS Quarterly, 37(2), 471–482.

[2]. Bird, C., Ford, D., Zimmermann, T., Forsgren, N., Kalliamvakou, E., & Houck, B. (2022). Taking flight with Copilot: Early insights and opportunities of AI-powered pair-programming tools. ACM Queue, 20(6), 35–57.

[3]. Bommasani, R., Hudson, D. A., Aditi, E., Altman, R., Arora, S., Bernstein, S., & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

[4]. Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. Harvard Business Review, 7, 3–11.

[5]. Basel Committee on Banking Supervision (2022). Principles for operational resilience. Bank for International Settlements.

[6]. Davenport, T. H. (2013). Keep up with your quants. Harvard Business Review, 91(7/8), 120–123.

[7]. Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277–304.

[8]. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

[9]. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.

[10]. Lacity, M. C., & Willcocks, L. P. (2016). A new approach to automating services. MIT Sloan Management Review, 58(1), 41–49.

[11]. McKinsey Global Institute (2023). The economic potential of generative AI: The next productivity frontier. McKinsey & Company.

[12]. Passi, S., & Jackson, S. J. (2018). Trust in data science: Collaboration, translation, and accountability in corporate data science projects. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 136.

[13]. Ramprasad, R., Kumar, S., & Singh, A. (2022). Financial document inconsistency detection using hybrid NLP approaches. Journal of Financial Data Science, 4(2), 88–104.

[14]. Reserve Bank of India (2023). Discussion Paper on Responsible AI in Financial Services. Reserve Bank of India.

[15]. Vial, G. (2019). Understanding digital transformation: A review and a research agenda. Journal of Strategic Information Systems, 28(2), 118–144.

[16]. Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., & Zhou, M. (2020). LayoutLM: Pre-training of text and layout for document image understanding. Proceedings of KDD 2020, 1192–1200.