Beyond AI Performance: A Unified Framework for Analyzing AI Integration in Financial Technology

Arsenii Moshninov
Graduate School of Business
HSE University
Moscow, Russia
akmoshninov@edu.hse.ru

Mikhail Komarov
Graduate School of Business
HSE University
Moscow, Russia
mmkomarov@hse.ru

***Abstract*— Artificial intelligence adoption in fintech is accelerating, yet successful deployment remains uneven in regulated and risk-sensitive environments. The relationship between AI architecture and sustainable organizational integration remains unclear. This study examines how fintech firms deploy classical machine learning, large language models, retrieval-augmented generation, and agentic systems using a combined Technology–Organization–Environment and Socio-Technical Systems framework. We analyzed twelve public cases across technological, organizational, environmental, and interaction dimensions. The presence of large language models did not imply high autonomy. Large-scale deployment was associated with systemic governance and calibrated trust, providing a structured basis for assessing AI deployment beyond model performance.**

Keywords— FinTech; Artificial Intelligence Adoption; TOE Framework; Socio-Technical Systems; Large Language Models

1. INTRODUCTION

1.1 Problem statement: "a model" is not "deployment"

Recent advances in AI, especially LLMs, improved model performance, yet organizational deployment often fails to deliver sustained value. Many AI initiatives underperform because success depends on organizational and socio-technical conditions, not only on model quality [1]. Prior work shows that deployment outcomes are shaped by leadership readiness, workforce skills, and the ability to redesign routines and controls [2]. Mixed-method evidence also suggests that staff skills and change capacity matter more than technological complexity in explaining adoption outcomes [3]. AI should be analyzed as a socio-technical system rather than a standalone technical artifact.

1.2 Why fintech: regulation, scale, and risk sensitivity

Fintech is a useful setting for studying AI deployment because regulation and risk control directly shape system design. Compliance expectations require traceability, auditability, and clear control boundaries, which limits 'deploy first, fix later' approaches [4]. Fintech also runs at high scale and low tolerance for errors. Real-time decisions can create financial loss and regulatory exposure, so organizations often need explicit oversight, access control, and robust data governance for high-impact use cases [5]. This combination of scale, regulation, and risk makes fintech a stress test for AI adoption frameworks. Successful deployment in this domain suggests that both technical architectures and organizational arrangements are robust enough to transfer to other regulated industries.

1.3 Research gap

AI adoption research often separates technological and organizational factors from human-AI work design. TOE structures adoption drivers but under-specifies how AI changes roles, decision authority, and trust in day-to-day operations [2]. STS captures these dynamics, yet many studies lack standard dimensions for cross-case comparison [6]. This creates a gap between macro-level adoption analysis and micro-level socio-technical dynamics. Combining TOE and STS addresses this limitation by enabling systematic cross-case comparison while preserving sensitivity to human-AI interaction.

2. RESEARCH QUESTIONS / CONTRIBUTIONS

The analysis is guided by four research questions:
RQ1: How do fintech organizations differ in technological configurations of AI systems, including LLM, RAG, and agentic architectures?
RQ2: How do these organizations differ in the social and interaction layer?

RQ3: What analytical methodology is required to systematically capture technological, organizational, environmental, and socio-interaction factors of AI deployment?
RQ4: What recurring AI deployment patterns can be identified across fintech cases using a combined TOE–STS perspective?
This study contributes methodologically by integrating TOE and STS into a unified analytical framework, empirically by comparing twelve fintech AI cases, and practically by identifying recurring deployment patterns in regulated environments.

3. METHODOLOGY

3.1 Case Selection

This study analyzes twelve public fintech AI cases: Wise, Monzo, Plaid (Income), Plaid (AI Coding), Revolut, Nubank, Lemonade, Coinbase, SumUp, Ramp, Adyen, and Digits. Cases were selected based on two criteria. Firstly, the source had to provide a clear description of the implemented AI architecture, including model types, infrastructure, or integration patterns. Secondly, the material had to contain observable organizational elements such as training, governance, deployment process, or workflow adaptation. Only cases meeting both conditions were included.

The selected cases span multiple AI application domains, including financial reporting and analytics (Digits [6], Wise [17], Monzo [16]), credit scoring and income verification (Plaid Income [11], Coinbase [10]), risk and compliance (Revolut [14], SumUp [9], Nubank [13]), customer-facing automation (Lemonade [12], Ramp [8]), and engineering productivity (Plaid AI Coding [15], Adyen [7]). The dataset captures variation in AI architectures, ranging from classical ML and sequence modeling to LLM-based systems, RAG pipelines, and agentic infrastructures.

3.2 Classification criteria and labels / Coding Scheme

The coding scheme is structured around a combined TOE–STS framework, it serves as the analytical backbone (Figure 1).

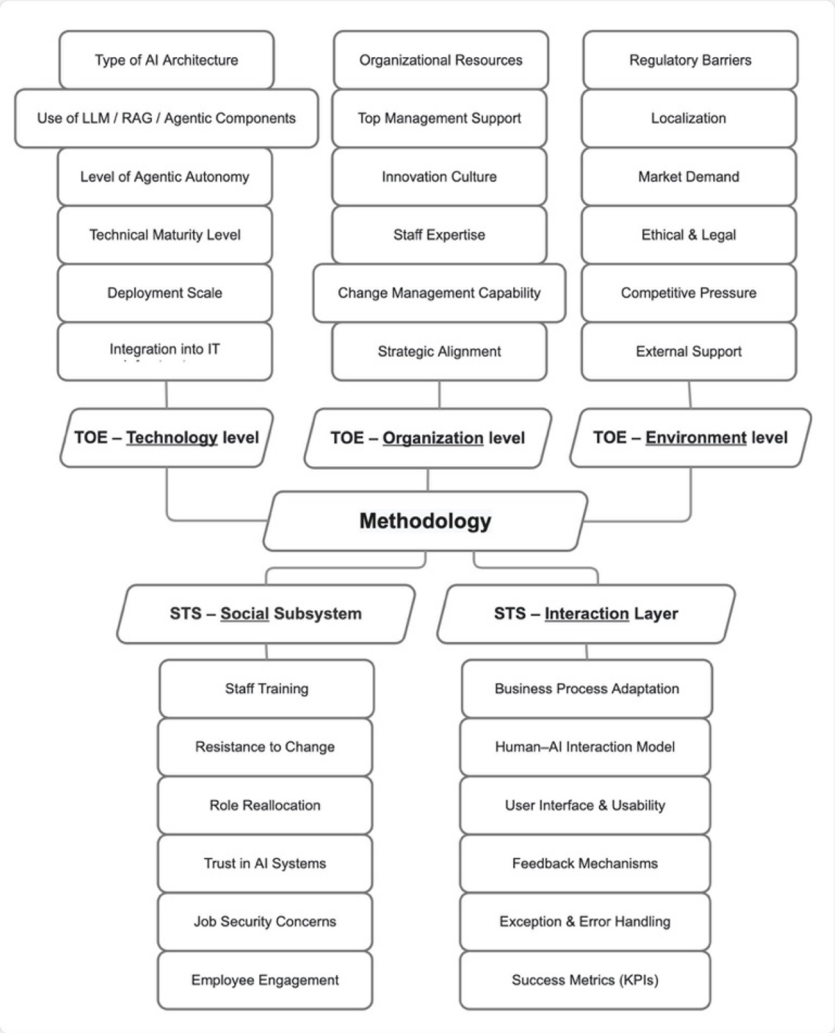

Figure 1. The chart with combined TOE–STS framework

Each criterion is operationalized through predefined labels to ensure comparability. For example, AI Architecture is coded as Classical ML, Hybrid ML, Deep Learning, LLM-based, Hybrid LLM + ML, or Agentic AI. STS dimensions are coded with descriptors such as Targeted/ Continuous training or Human-in-the-loop/AI-first interaction. The complete list of labels and their definitions is presented in Figures 2-6 (Appendix A).

3.3 Classification and Normalization

Rich textual case descriptions were transformed into standardized classification labels through a structured two-stage procedure: Firstly, all twelve cases were uploaded into the Gemini environment via the NotebookLM interface [18]. Prior to classification, the full set of predefined criteria and labels (Figures 2–6, Appendix A) was fixed and embedded directly into the prompt to ensure consistency and prevent category drift. Gemini was then used in a controlled RAG-based mode to extract evidence from each case and map long-form descriptions onto the corresponding TOE and STS labels. A detailed coding protocol describing operational definitions and evidence indicators for each criterion is provided in Appendix C. The protocol ensured traceability by linking classification labels to explicit textual evidence from the source materials.

Secondly, all generated classifications were manually reviewed and validated by the authors. In cases where classification remained ambiguous, final coding decisions were determined through joint author review and consensus based on the most explicit interpretation of the source evidence. To increase coding reliability, both authors independently reviewed the classification results. Inter-coder agreement across coded attributes reached approximately 90%, indicating a high level of coding consistency. Remaining disagreements were resolved through discussion and reference to the original source materials.

3.4 Analysis approach

The analysis followed a cross-case comparative logic. After normalization, cases were examined for recurring configurations across technological, organizational, environmental, and socio-technical dimensions.

4 RESULTS

Classification results are presented in Table I (Appendix B). The sample is not random and it consists of twelve leading fintech cases that meet strict architectural and organizational transparency criteria (see Methodology).

Model proliferation does not automatically translate into delegated autonomy. LLM or RAG components appear in 8 of 12 cases. However, only two cases demonstrate high agentic autonomy. In this sample, high autonomy emerges only when LLM or Agentic LLM is combined with RAG and platform-native integration. Importantly, both High-autonomy cases share the same STS profile: continuous training, AI-first interaction, self-optimizing feedback, systemic exception handling, and calibrated trust.

Platform-Native Architecture suggests Governance Formalization. All platform-native cases demonstrate systemic exception handling. Platform-native integration is frequently associated with global or enterprise-level deployment. However, platform-native status does not automatically imply AI-native UX or self-optimizing feedback. This suggests that infrastructure depth predicts governance formalization, but not necessarily interaction sophistication.

Scaled deployment requires dual institutionalization. Three cases are coded as Mature-at-scale. All three share the following attributes:

- Global deployment,
- Platform-native integration,
- Institutionalized or continuous change management,
- Calibrated trust,
- Systemic exception handling.

This cluster shows that scale in fintech AI is potentially associated with dual institutionalization: technical standardization and organizational normalization.

Calibrated trust characterizes global AI deployment. Global deployments consistently show calibrated trust rather than high trust. No global case relies on unconditional trust. High trust appears more often in enterprise or internal deployments.

Process re-engineering is not determined by model type alone. Business process re-engineering appears in both hybrid ML and LLM-based systems. It is not specific to generative architectures. Some classical or hybrid ML platforms exhibit

deeper operational restructuring than certain LLM cases. This potentially indicates that infrastructural embedding, not model type, drives structural process change.

Governance formalization increases with deployment scope. Manual exception handling is observed only in internally limited deployments. In contrast, systemic exception handling dominates global and large enterprise cases. Rule-based approaches are more typical of mid-level implementations. Although risk intensity was not explicitly coded, broader deployment scope is likely associated with more structured control mechanisms in this sample.

Organizational resources function as baseline conditions rather than differentiating factors. Most cases demonstrate strong resources, strategic alignment, and high AI literacy. Given the non-random, high-performing sample, these factors function as baseline conditions rather than discriminators.

5 Discussion / Limitations / Future Work

The patterns observed in this study are consistent with recent work on AI adoption in information systems research. Staff skills and organizational change capacity are stronger predictors of GenAI adoption than technological features alone [3], [19], [20]. Practitioner evidence on LLM production systems further confirms that RAG pipelines, monitoring, and human oversight are essential for stable deployment, especially in high-risk settings [21]. Likewise, socio-technical research demonstrates that AI adoption unfolds through recursive alignment between technical design, roles, and organizational routines rather than as a linear technical upgrade [22].

From a practical perspective, teams should treat AI autonomy as a governance design problem rather than as a model capability. Decisions about RAG integration, exception handling, and feedback loops should be made together with risk and compliance functions. Continuous training and trust calibration mechanisms need to be designed explicitly, not assumed to emerge from usage.

Theoretically, the study suggests that TOE and STS are most informative when treated as interacting layers rather than parallel explanations. This supports a configurational view of AI adoption, where outcomes depend on bundles of aligned technical and socio-technical elements.

The cases are based on public reports and mainly describe successful implementations, which may create positive bias. The study relies on a relatively small and purposive sample of twelve publicly documented fintech AI deployments. As a result, the findings should be interpreted as pattern-indicative of fintech AI leaders, practical and expert rather than macro-scale statistically generalizable big data study. The coding process, although structured and reviewed, includes subjective interpretation and does not fully allow causal statistical conclusions. As an additional robustness check, selected cases were re-examined under alternative classification interpretations. The overall cross-case patterns remained stable, suggesting that the identified configurations are not associated with individual coding choices.

Future research should incorporate quantitative indicators such as incident rates, latency, and adoption metrics. Configurational analytical approaches such as fuzzy-set qualitative comparative analysis (fsQCA) may also be applied to formally test the deployment configurations identified in this exploratory study. In-depth interviews with technical and compliance teams would further strengthen the empirical grounding of the analysis. It may also develop and validate an interactive RAG-based analytical tool with a reasoning LLM agent and user interface that systematically transforms qualitative case descriptions into structured tables and supports comparative case-level decision assessment.

6 Conclusion

The findings showed that technological diversity in fintech cannot be reduced to a simple transition from classical ML to generative AI. Instead, firms adopt distinct architectural configurations that combine model type, retrieval logic, integration depth, and deployment maturity in different ways. Agentic systems can be interpreted as one possible trajectory and tend to appear in contexts characterized by specific infrastructural and organizational conditions rather than as a natural outcome of model progress.

The analysis demonstrated that differences in AI deployment are equally visible in the social and interaction layer. Organizations vary in how they organize training, distribute decision authority, structure human-AI interaction, and formalize feedback and escalation processes. Large-scale and stable deployments are potentially associated with clearly defined control boundaries, structured learning practices, and explicit trust calibration mechanisms. Social embedding and interaction design appear to function as central components of deployment architecture, not secondary implementation details.

The study proposed a structured methodology that integrates TOE and STS into a unified analytical lens. This approach enabled systematic comparison of heterogeneous cases while preserving sensitivity to socio-technical dynamics. By translating qualitative deployment narratives into comparable categories, the framework made visible the configurations through which technical, organizational, environmental, and interactional factors align.

The analysis identified recurring deployment patterns:

First, the presence of LLM or RAG components does not automatically lead to high autonomy; advanced agentic configurations were observed primarily in cases where retrieval mechanisms, platform-native integration, continuous training, and formalized oversight were combined.

Second, large-scale deployment is frequently associated with systemic exception handling, calibrated trust models rather than unconditional high trust, and institutionalized change management, suggesting that large-scale deployment tends to coincide with governance mechanisms embedded both technically and organizationally. Medium trust emerges as a structural feature of global systems, while high trust appears more often in limited or internal AI systems.

Third, process transformation appears to be more strongly associated with infrastructural embedding and interaction design than with model type. Platform-native integration is strongly linked to governance formalization, but it does not automatically imply advanced user experience or self-optimizing feedback mechanisms.

Finally, organizational resources, strategic alignment, and AI literacy appear to function as baseline conditions across the sample rather than as key differentiators.

Across cases, successful AI deployment is typically associated with stable alignment between architecture, control mechanisms, and clearly defined human roles.

The main contribution of this paper lies in formalizing AI deployment as a multi-layer configuration problem. Methodologically, it advances cross-case AI research by operationalizing socio-technical dimensions within a structured comparative framework. Empirically, it maps distinct deployment archetypes across twelve fintech cases. Practically, it provides a diagnostic lens for assessing alignment between architecture, governance, and organizational practices in high-risk environments. The paper also formulates practice-oriented, proven AI implementation results and insights based on company cases.

The findings are relevant for corporations implementing AI/GenAI systems in the financial industry (and other regulation-sensitive industries) and researchers evaluating AI deployment. The study also informs public officials, policymakers involved in designing legal and regulatory requirements for AI in fintech and other services.

DATA AVAILABILITY STATEMENT / ACKNOWLEDGEMENT

All data used in this study are based on publicly available technical and engineering reports cited in the reference list [6]–[17]. The structured classification results are provided in Appendix B. Large language models (Gemini NotebookLM, Yandex GPT, GigaChat and ChatGPT) were used for language editing and structural refinement. The authors fully reviewed all outputs and took complete responsibility for the research design, analysis, and conclusions. No generative AI tools were used to produce original research findings or intellectual contributions.

REFERENCES

[1] N. Welch, “Overcoming leadership barriers in AI adoption: A socio-technical ambidextrous approach,” RAIS Conference Proceedings, pp. 71–79, 2025. doi: 10.5281/zenodo.16218385.

[2] D. Smit, S. Eybers, A. van der Merwe, R. Wies, N. Human, and J. Pielmeier, “Integrating the TOE framework and DOI theory to dissect and understand the key elements of AI adoption in sociotechnical systems,” South African Computer Journal, vol. 36, no. 2, pp. 159–176, 2024. doi: 10.18489/sacj.v36i2.17679.

[3] L. Hughes, F. Davies, K. Li, S. M. Gunaratnege, T. Malik, and Y. K. Dwivedi, “Beyond the hype: Organisational adoption of generative AI through the lens of the TOE framework—A mixed methods perspective,” International Journal of Information Management, vol. 86, Art. no. 102982, 2026. doi: 10.1016/j.ijinfomgt.2025.102982.

[4] A. Rossi, M. K. Patel, L. Chen, and S. Hofmann, “AI integration in financial services: A systematic review of trends and regulatory challenges,” Humanities and Social Sciences Communications, vol. 12, no. 1, Art. no. 123, Apr. 2025. doi: 10.1057/s41599-025-04850-8.

[5] J. Kramer and L. Vogel, “Regulating AI in financial services: Legal frameworks and compliance challenges,” European Business Law Review, vol. 35, no. 4, pp. 621–654, Aug. 2024. Available: SSRN 5168628.

[6] H. Hapke and C. Howard, “ChatGPT for Accounting: How Digits is Using Generative Machine Learning to Transform Finance,” Digits Blog, Mar. 8, 2023. [Online]. Available: https://digits.com/developer/posts/assisting-accountants-with-generative-machine-learning/

[7] R. P. Ledinski, "Elevating Code Quality Through LLM Integration: Adyen's Venture into Augmented Unit Test Generation," Adyen Knowledge Hub, Mar. 27, 2024. [Online]. Available: https://www.adyen.com/knowledge-hub/elevating-code-quality-through-llm-integration
[8] Ramp Business Corporation, "Engineering Posts – Ramp Builders Blog," 2023–2025. [Online]. Available: https://engineering.ramp.com/industry_classification
[9] N. Castillo, "Evaluating the Performance of an LLM Application that Generates Free-Text Narratives in the Context of Financial Crime," Inside SumUp (Medium), Oct. 11, 2023. [Online]. Available: https://medium.com/inside-sumup/evaluating-the-performance-of-an-llm-application-that-generates-free-text-narratives-in-the-context-c402a0136518
[10] J. McAllister and A. Rainwater, "How Coinbase Builds Sequence Features for Machine Learning," Coinbase Blog, May 13, 2025. [Online]. Available: https://www.coinbase.com/en-ar/blog/how-coinbase-builds-sequence-features-for-machine-learning
[11] W. Yao, J. Nagda, A. Annadi, and R. Sriram, "How We Use Machine Learning to Power Accurate, Real-Time Income Verification," Plaid Blog, Sep. 27, 2023. [Online]. Available: https://plaid.com/blog/machine-learning-income-verification/
[12] L. Twito, "GenAI Lead, Lemonade," presentation at AI Summit @ Reichman University, Sep. 4, 2025. [Online]. Available: https://www.youtube.com/watch?v=XGJOU2sysjg
[13] Nubank Editorial, "Presenting Precog, Nubank's Real-Time Event AI," Nubank Research, Jun. 21, 2023. [Online]. Available: https://building.nubank.com.br/presenting-precog-nubanks-real-time-event-ai/
[14] K. Vasilev, "Reinventing Risk at Revolut," Revolut Tech (Medium), Dec. 10, 2024. [Online]. Available: https://medium.com/revolut/reinventing-risk-at-revolut-77e63c552503
[15] C. Allsopp and J. Carroll, "Transforming Engineers: How We Grew AI Coding Adoption at Plaid," Plaid Blog, May 8, 2025. [Online]. Available: https://plaid.com/blog/ai-coding-adoption-plaid/
[16] E. Garland, "Using Topic Modelling to Understand Customer Saving Goals," Data @ Monzo (Medium), Feb. 17, 2023. [Online]. Available: https://medium.com/data-monzo/using-topic-modelling-to-understand-customer-saving-goals-2bb06f00ce2d
[17] M. James, "Wise Tech Stack (2025 Update)," Wise Engineering (Medium), Feb. 12, 2025. [Online]. Available: https://medium.com/wise-engineering/wise-tech-stack-2025-update-d0e63fe718c7
[18] Google, "NotebookLM," 2025. [Online]. Available: https://notebooklm.google.com/
[19] J. Schwäke, A. Peters, D. K. Kanbach, S. Kraus, and P. Jones, "The new normal: The status quo of AI adoption in SMEs," Journal of Small Business Management, published online Aug. 13, 2024, doi: 10.1080/00472778.2024.2379999. DOI: https://doi.org/10.1080/00472778.2024.2379999.
[20] O. Neumann, K. Guirguis, and R. Steiner, "Exploring artificial intelligence adoption in public organizations: a comparative case study," Public Management Review, vol. 26, no. 1, pp. 114–141, Jan. 2024, doi: 10.1080/14719037.2022.2048685. DOI: https://doi.org/10.1080/14719037.2022.2048685.
[21] A. Mailach, S. Simon, J. Dorn, and N. Siegmund, "Themes of building LLM-based applications for production: A practitioner's view," arXiv preprint arXiv:2411.08574, 2025.
[22] A. Amanollahnejad, S. Fosso-Wamba, M. S. Shabbir, and A. Pakseresht, "Aligning socio-technical systems: Rethinking AI adoption and digital transformation in SMEs," Information Systems Management, early access, 2026, doi: 10.1080/10580530.2025.2612175.

Authors' background

Arsenii Moshninovis currently pursuing the master's degree with the E-business and digital innovation, Graduate School of Business, HSE University. He presented a paper at the ICAIMT 2025 International Conference on AI. In addition to his studies, he is a Product Manager in an IT-fintech company. The area of scientific interest is Machine learning, LLM, Artificial intelligence in application to management and business processes.

Mikhail Komarov is currently a Professor with the Department of Business Informatics, Graduate School of Business, HSE University. He is a Specialist in AI, wireless data transmission and IT. He is the Vice-Chair of the Special Interest Group on IoT at the Internet Society (https://www.hse.ru/staff/mkomarov/)

Supplementary Materials : Appendix A, B and C

The complete classification labels (**Appendix A**), full cross-case results (**Appendix B**), coding protocol (**Appendix C**) are available in the supplementary materials at *https://drive.google.com/drive/folders/1l4FpWDSDaXoRS8vCvRW3x3JeSMtf66qr?usp=sharing*