International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

Journal Website: http://sciencebring.com/index.php/ijasr

Research Article

Edge-Intelligent Microservice Orchestration For Privacy-Preserving, Real-Time Generative Financial Technologies

Submission Date: December 28, 2025, **Accepted Date:** January 22, 2026,
Published Date: February 17, 2026

Dr. Matteo Rinaldi
Department of Information Engineering, University of Bologna, Italy

Abstract

The convergence of artificial intelligence, microservices, and edge computing has produced a transformative paradigm for the development of modern financial technology platforms. In particular, the rapid rise of generative artificial intelligence in financial services has exposed deep architectural tensions between performance, privacy, scalability, and regulatory compliance. Traditional centralized cloud infrastructures are increasingly insufficient for applications that require real-time decision making, ultra-low latency, and the protection of highly sensitive personal and transactional data. Edge-AI microservice orchestration has emerged as a critical response to these challenges, allowing computational intelligence to be distributed closer to data sources while maintaining modular, scalable, and resilient service structures. This research presents a comprehensive theoretical and analytical investigation into how edge-based orchestration of AI-powered microservices can support private, real-time generative financial applications. The study is grounded in contemporary scholarly work on microservice orchestration, AI-driven resource allocation, hybrid cloud-edge infrastructures, and adaptive service coordination, with particular emphasis on recent FinTech-oriented edge AI architectures (Hebbar, Sharma, and Maheshkar, 2026).

The article develops a holistic conceptual framework that explains how generative models, microservice lifecycles, orchestration engines, and edge intelligence interact in complex financial ecosystems. Rather than treating AI as a monolithic service, this research conceptualizes generative AI capabilities as

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

dynamically orchestrated microservices that adapt to fluctuating data loads, security requirements, and regulatory constraints. The theoretical foundations of orchestration and choreography are revisited to explain how decentralized control models enable high-speed and fault-tolerant financial workflows, particularly in environments where data locality and privacy preservation are paramount (Singhal, Sakthivel, and Raj, 2019; Zeb et al., 2023).

Using a qualitative, literature-grounded analytical methodology, this study synthesizes insights from AI-driven resource allocation, deep reinforcement learning for self-adaptive systems, and next-generation edge networking to interpret how microservices evolve from static deployment units into intelligent, self-regulating entities (Magableh and Almiani, 2019; Barua and Kaiser, 2024). The results demonstrate that edge-oriented orchestration enhances financial application reliability, regulatory compliance, and computational efficiency, particularly for generative services such as real-time fraud detection, personalized investment advisory, and conversational banking.

The discussion further explores how these architectures reconfigure institutional power, data governance, and technological sovereignty within financial systems, addressing ongoing debates about centralization versus decentralization in digital economies. The paper concludes that edge-AI microservice orchestration represents not merely a technical optimization but a structural transformation of financial computation, enabling a new generation of trustworthy, intelligent, and scalable FinTech infrastructures.

Keywords

Edge computing, microservices orchestration, generative artificial intelligence, financial technology, privacy-preserving systems, distributed cloud, real-time analytics

Introduction

The evolution of financial technology has always been tightly coupled to advances in computation, networking, and data processing. From the earliest electronic trading systems to contemporary digital banking platforms, each technological wave has redefined how financial value is created, exchanged, and regulated. In the present era, generative artificial intelligence and microservices architecture are emerging as two of the most powerful drivers of change in financial systems, reshaping not only how financial applications are built but also how they interact with users, institutions, and regulators (Kaniganti and Challa, 2024; Oye, Frank, and Owen, 2024). Generative AI models now produce investment advice, detect fraud patterns, simulate market behaviors, and interact with customers in natural language, while microservices enable these complex capabilities to be decomposed into modular, independently deployable components. However, this technological convergence has created new challenges related to latency,

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

privacy, data sovereignty, and system reliability, particularly in financial contexts where milliseconds and confidentiality can determine success or failure (Hebbar, Sharma, and Maheshkar, 2026).

Traditional cloud-centric architectures were designed around the assumption that data could be safely transmitted to centralized data centers for processing. While this model has proven effective for many enterprise applications, it is increasingly misaligned with the operational realities of real-time generative FinTech systems. Financial data are among the most sensitive forms of information, subject to stringent regulatory controls, cross-border data transfer restrictions, and heightened risks of misuse. At the same time, generative AI models often require immediate access to streaming transactional data to generate contextually relevant outputs, such as personalized credit decisions or dynamic risk assessments (Zeb et al., 2023). These requirements produce a fundamental tension: centralizing computation in the cloud introduces unacceptable delays and privacy risks, while decentralizing intelligence across devices and local nodes complicates coordination and management.

Edge computing has emerged as a powerful response to this dilemma by relocating computational resources closer to data sources, such as customer devices, branch servers, or local financial infrastructure. By processing data at the network edge, organizations can reduce latency, improve responsiveness, and retain sensitive information within controlled jurisdictions (Taherizadeh et al., 2018). Yet edge computing alone does not resolve the complexity of deploying and managing large-scale AI applications. Generative AI workloads are computationally demanding, dynamic, and interdependent, requiring sophisticated orchestration mechanisms to ensure that services are delivered reliably and efficiently across heterogeneous environments (Moreschini et al., 2023).

Microservices architecture provides a conceptual and technical foundation for this orchestration. By decomposing applications into loosely coupled services that communicate through well-defined interfaces, microservices enable continuous deployment, scalability, and fault isolation (Felstaine and Hermoni, 2018). In the context of financial systems, microservices allow distinct functions such as authentication, transaction processing, fraud detection, and customer interaction to evolve independently while still contributing to a unified service experience (Arachchige and Thamoda, 2024). When combined with AI, microservices become not merely computational modules but intelligent agents that can adapt to changing conditions, learn from data, and optimize their own behavior (Magableh and Almiani, 2019).

The integration of edge computing, microservices, and generative AI has given rise to a new architectural paradigm: edge-AI microservice orchestration. This paradigm is characterized by the dynamic deployment, scaling, and coordination of AI-powered services across distributed edge and cloud resources to

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

meet strict performance and privacy requirements. Recent research has demonstrated that such architectures are particularly well suited to financial applications, where real-time responsiveness and data locality are critical (Hebbar, Sharma, and Maheshkar, 2026). By orchestrating generative AI models as microservices at the edge, FinTech platforms can deliver personalized and secure services without exposing sensitive data to centralized repositories.

Despite growing interest in this area, the scholarly literature remains fragmented. Some studies focus on AI-enabled orchestration in next-generation networks (Zeb et al., 2023), others on AI-driven resource allocation in hybrid clouds (Barua and Kaiser, 2024), and still others on microservice lifecycle management (Moreschini et al., 2023). However, relatively few works have developed an integrated theoretical understanding of how these components interact specifically in the context of real-time generative financial applications. The work of Hebbar, Sharma, and Maheshkar (2026) provides a critical starting point by demonstrating how edge-based microservice orchestration can support private generative FinTech workloads, yet their study invites deeper theoretical elaboration and broader conceptualization.

The gap in the literature is therefore not merely technical but epistemological. Financial technology is often treated as an application domain rather than a site of architectural innovation. Yet the regulatory, ethical, and operational constraints of finance demand unique solutions that cannot be directly borrowed from other industries. Generative AI in finance introduces new risks related to explainability, bias, and data leakage, which must be addressed at the architectural level through careful orchestration and deployment strategies (Castillo and Restrepo, 2024). Edge-AI microservices provide a promising avenue for embedding these safeguards into the very fabric of financial computation.

This article seeks to fill this gap by developing a comprehensive, theoretically grounded analysis of edge-AI microservice orchestration for real-time, privacy-preserving generative FinTech systems. Drawing on a wide range of interdisciplinary sources, it explores how orchestration models, AI techniques, and distributed infrastructures converge to create new forms of financial intelligence. By situating the work of Hebbar, Sharma, and Maheshkar (2026) within this broader intellectual landscape, the study aims to clarify the mechanisms through which edge-based orchestration transforms financial services and to identify the conditions under which it can be most effectively deployed.

In doing so, the paper also engages with broader debates about centralization and decentralization in digital economies. Cloud computing has historically concentrated power and data in the hands of a few global providers, raising concerns about monopolization, surveillance, and systemic risk. Edge-AI microservices, by contrast, distribute intelligence and control across a network of localized nodes, potentially enabling more democratic and resilient financial

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

infrastructures (Zeb et al., 2023). Whether this promise can be realized depends on how orchestration frameworks are designed, governed, and aligned with regulatory objectives.

The following sections develop these themes in depth. The methodology outlines how a qualitative, literature-based analytical approach can generate robust theoretical insights into complex technological systems. The results section interprets the implications of edge-AI orchestration for performance, privacy, and reliability in generative FinTech. The discussion situates these findings within broader scholarly and institutional contexts, addressing limitations and future research directions. Together, these components provide a comprehensive account of how edge-intelligent microservice orchestration is reshaping the future of financial technology.

Methodology

The methodological approach adopted in this research is designed to address the inherently complex and multi-layered nature of edge-AI microservice orchestration in financial technology environments. Unlike empirical studies that rely on numerical experimentation or performance benchmarking, this work employs a qualitative, theory-driven analytical methodology grounded in systematic engagement with the existing scholarly literature. This approach is particularly appropriate given the rapidly evolving nature of both generative artificial intelligence and microservice-based architectures, where conceptual clarity and theoretical integration are often prerequisites for meaningful empirical validation (Moreschini et al., 2023).

The methodological foundation of the study is built upon interpretive synthesis, which involves the careful examination, comparison, and integration of findings from diverse academic sources into a coherent explanatory framework. This approach allows the researcher to move beyond isolated technical descriptions and instead develop a holistic understanding of how architectural components, organizational practices, and regulatory constraints interact in real-world financial systems (Kaniganti and Challa, 2024). By situating the work of Hebbar, Sharma, and Maheshkar (2026) within this broader context, the methodology ensures that the analysis remains anchored in authoritative research while also extending its theoretical implications.

The first stage of the methodology involved the identification and thematic categorization of relevant literature. Sources were drawn from the provided reference list, which encompasses a wide range of perspectives on microservices, artificial intelligence, orchestration mechanisms, and distributed computing. These sources were grouped into conceptual clusters, including AI-enabled orchestration, microservice lifecycle management, edge computing architectures, and financial application requirements. This thematic organization allowed for the identification of recurring concepts such as self-adaptive systems, hybrid cloud-edge deployment, and privacy-

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

preserving computation (Barua and Kaiser, 2024; Zeb et al., 2023).

In the second stage, each thematic cluster was subjected to close reading and critical analysis. Rather than simply summarizing the content of individual studies, the analysis focused on extracting underlying assumptions, theoretical commitments, and points of convergence or divergence among authors. For example, research on deep reinforcement learning for self-adaptive microservices emphasizes the role of autonomous decision-making in dynamic environments (Magableh and Almiani, 2019), while studies on orchestration versus choreography highlight the organizational logic of service coordination (Singhal, Sakthivel, and Raj, 2019). By juxtaposing these perspectives, the methodology reveals how technical choices reflect deeper philosophical positions about control, autonomy, and system governance.

The third stage involved the construction of an integrative analytical narrative that links these theoretical insights to the specific domain of generative FinTech applications. Financial technology is not treated as a neutral application layer but as a socio-technical system shaped by regulatory, ethical, and economic forces (Castillo and Restrepo, 2024). The work of Hebbar, Sharma, and Maheshkar (2026) is particularly valuable in this regard because it explicitly addresses the challenges of privacy, real-time processing, and generative AI in financial contexts. Their study provides a concrete case through which abstract architectural principles can be examined and interpreted.

A key methodological principle guiding this research is reflexivity. The analysis recognizes that architectural models are not merely technical artifacts but also embodiments of particular values and priorities. For instance, edge computing architectures often reflect a commitment to data sovereignty and user privacy, while centralized cloud models prioritize economies of scale and centralized control (Taherizadeh et al., 2018). By making these normative dimensions explicit, the methodology avoids the trap of technological determinism and instead situates architectural choices within broader social and institutional frameworks.

Another important methodological consideration is the treatment of time and evolution. Microservice architectures and AI techniques are not static; they evolve in response to technological innovation, market pressures, and regulatory changes (Vudayagiri, 2024). The literature-based approach allows the study to trace these evolutionary trajectories, identifying how early container-based deployments gave rise to more sophisticated orchestration frameworks and how rule-based AI systems have been supplanted by generative models capable of producing novel outputs (Felstaine and Hermoni, 2018; Oye, Frank, and Owen, 2024). This historical perspective enriches the analysis by revealing patterns of continuity and disruption.

The limitations of this methodology must also be acknowledged. Because the study does not involve direct empirical observation or experimental testing, its conclusions are necessarily interpretive and theoretical. While

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

this does not diminish their value, it does mean that they should be understood as propositions to be further explored and validated through future empirical research (Moreschini et al., 2023). Nevertheless, in a field characterized by rapid innovation and conceptual fragmentation, theory-driven synthesis plays a crucial role in guiding both research and practice.

The methodological rigor of the study is reinforced by its reliance on peer-reviewed and reputable academic sources. The inclusion of recent work on AI-driven resource allocation, next-generation networks, and FinTech-specific architectures ensures that the analysis reflects the current state of the field (Barua and Kaiser, 2024; Zeb et al., 2023; Hebbar, Sharma, and Maheshkar, 2026). By integrating these perspectives into a unified framework, the methodology provides a robust foundation for the interpretive results and theoretical discussions that follow.

Results

The interpretive synthesis of the literature reveals a set of interrelated findings that illuminate how edge-AI microservice orchestration fundamentally transforms the design and operation of generative financial technology systems. These findings are not empirical measurements in the traditional sense but conceptual insights derived from the convergence of multiple scholarly perspectives, each of which contributes to a deeper understanding of the architectural dynamics at play (Moreschini et al., 2023).

One of the most significant results concerns the role of edge intelligence in resolving the tension between performance and privacy in generative FinTech applications. Financial systems require extremely low latency for functions such as fraud detection, high-frequency trading, and conversational banking interfaces. At the same time, they must adhere to strict data protection regulations that limit the transmission of sensitive information across networks and jurisdictions (Hebbar, Sharma, and Maheshkar, 2026). The literature indicates that by deploying generative AI models as microservices at the edge, organizations can process data locally, reducing both response times and the risk of data leakage (Taherizadeh et al., 2018; Zeb et al., 2023). This architectural shift allows financial institutions to meet regulatory requirements without sacrificing the adaptive capabilities of AI.

A second major result pertains to the emergence of AI-driven orchestration as a core enabling technology. Traditional orchestration frameworks rely on predefined rules and static policies to manage service deployment and scaling. However, generative FinTech workloads are highly variable, driven by unpredictable user interactions, market fluctuations, and evolving risk profiles (Oye, Frank, and Owen, 2024). Studies on deep reinforcement learning for self-adaptive microservices demonstrate that AI-based controllers can dynamically allocate resources, adjust service configurations, and optimize performance in real time (Magableh and

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

Almiani, 2019). When applied to edge environments, these capabilities allow microservices to respond autonomously to changing conditions, ensuring consistent quality of service even under volatile demand.

The results also highlight the importance of microservice lifecycle management in sustaining the reliability of generative financial systems. Generative AI models require frequent updates, retraining, and fine-tuning to remain accurate and unbiased. In a monolithic architecture, such updates can disrupt entire applications, whereas in a microservices-based system, individual components can be updated independently (Felstaine and Hermoni, 2018). Research on AI techniques in the microservices lifecycle shows that automated testing, deployment, and monitoring can be enhanced through machine learning, reducing downtime and improving system resilience (Moreschini et al., 2023). For FinTech applications, this translates into more stable and trustworthy services, even as underlying models evolve.

Another important finding concerns the organizational logic of service coordination. The debate between orchestration and choreography has significant implications for how generative AI services are deployed across edge and cloud environments. Orchestration implies centralized control, where a master controller directs the behavior of individual services, while choreography emphasizes decentralized interaction, where services respond to events and messages without a central coordinator (Singhal, Sakthivel, and Raj, 2019). The literature suggests that hybrid models are particularly effective in financial contexts, combining centralized oversight for compliance and risk management with decentralized execution for speed and resilience (Zeb et al., 2023; Hebbar, Sharma, and Maheshkar, 2026).

The results further indicate that hybrid cloud-edge infrastructures provide a flexible substrate for generative FinTech systems. While edge nodes handle real-time processing and sensitive data, cloud resources offer scalable storage and computational power for model training and long-term analytics (Barua and Kaiser, 2024). AI-driven resource allocation frameworks enable workloads to be distributed across this hybrid environment in a way that optimizes both cost and performance. For example, routine inference tasks can be executed at the edge, while intensive retraining operations are offloaded to the cloud, creating a balanced and efficient computational ecosystem.

A final key result is the recognition that edge-AI microservice orchestration reshapes not only technical architectures but also institutional practices. Financial organizations adopting these systems must develop new governance models, security protocols, and compliance mechanisms to manage distributed intelligence (Castillo and Restrepo, 2024). The literature suggests that this transition requires a shift from centralized IT management to more collaborative and adaptive forms of technological stewardship, where teams monitor and guide autonomous services rather than directly controlling every operation.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

ISSN-2750-1396

Together, these results paint a picture of edge-AI microservice orchestration as a multidimensional transformation. It enhances performance and privacy, enables adaptive and resilient service delivery, and reconfigures the organizational and regulatory landscape of financial technology. These findings provide the foundation for the deeper theoretical interpretation and critical discussion that follows.

Discussion

The theoretical implications of the results extend far beyond the immediate technical benefits of edge-AI microservice orchestration. They point to a fundamental reorganization of how financial intelligence is produced, governed, and experienced in contemporary digital economies. At the heart of this transformation lies a shift from centralized, monolithic computation to distributed, adaptive, and context-aware systems that operate at the edge of networks and institutions (Hebbar, Sharma, and Maheshkar, 2026).

One of the most profound theoretical contributions of this architectural shift is its challenge to traditional notions of control and authority in information systems. In classical enterprise computing, control is exercised through centralized infrastructures, where data and computation are managed by a single organizational entity. Microservices and edge computing disrupt this model by distributing both data and decision-making across a network of autonomous components (Felstaine and Hermoni, 2018; Taherizadeh et al., 2018). In financial contexts, this decentralization has significant implications for regulatory compliance, risk management, and institutional accountability. While decentralized systems can be more resilient and responsive, they also complicate oversight, raising questions about how regulators and auditors can monitor activities that occur across thousands of edge nodes.

Scholarly debates about orchestration versus choreography are particularly relevant here. Orchestration, with its centralized control plane, aligns more closely with traditional regulatory frameworks that assume a single point of accountability (Singhal, Sakthivel, and Raj, 2019). Choreography, by contrast, reflects a more networked and emergent form of coordination, where behavior arises from interactions among services rather than directives from a central authority. The hybrid models observed in edge-AI FinTech architectures attempt to reconcile these logics by embedding regulatory and governance functions into orchestration layers while allowing generative services to operate autonomously at the edge (Zeb et al., 2023). This hybridization represents a new form of socio-technical governance, one that blends hierarchical oversight with distributed agency.

Another critical theoretical dimension concerns the nature of intelligence in financial systems. Generative AI models do not merely analyze data; they produce new representations, predictions, and narratives that shape user behavior and market dynamics (Oye, Frank, and Owen, 2024).

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

When these models are deployed as microservices at the edge, their outputs become tightly coupled to local contexts, such as a customer's transaction history or a regional market trend. This localization of intelligence enhances relevance and privacy but also introduces the possibility of divergent interpretations and decisions across different nodes. From a theoretical perspective, this raises questions about consistency, fairness, and systemic coherence in distributed financial systems (Castillo and Restrepo, 2024).

The literature on AI-driven resource allocation provides further insight into how these challenges can be managed. By using machine learning to monitor system performance and adjust resource distribution in real time, orchestration frameworks can maintain balance and stability even in highly decentralized environments (Barua and Kaiser, 2024; Magableh and Almiani, 2019). This self-regulating capability echoes cybernetic theories of control, where systems achieve equilibrium through feedback loops rather than centralized command. In the context of finance, such feedback-driven orchestration may offer a more robust and adaptive form of governance than traditional rule-based systems.

Historical perspectives on computing further illuminate the significance of this shift. Early financial information systems were characterized by rigid, centralized architectures designed for batch processing and limited interactivity. The rise of online banking and electronic trading introduced greater connectivity and real-time processing but still relied heavily on centralized data centers. Edge-AI microservice orchestration represents the next evolutionary step, enabling financial intelligence to be embedded directly into the points of interaction between users, devices, and markets (Vudayagiri, 2024). This evolution reflects broader trends in digital technology, where intelligence is increasingly distributed across networks rather than concentrated in isolated hubs.

Counter-arguments to this vision emphasize the risks of fragmentation and complexity. Critics argue that distributing AI services across edge nodes increases the likelihood of configuration errors, security vulnerabilities, and inconsistent behavior (Moreschini et al., 2023). In financial systems, such risks can have severe consequences, including financial loss, regulatory violations, and erosion of customer trust. These concerns underscore the importance of robust orchestration frameworks and standardized communication protocols, as highlighted in studies on delay-aware coordination and microservice optimization (Wang et al., 2019; Charankar and Pandiya, 2024). Effective orchestration is not merely a technical convenience but a prerequisite for the safe and reliable operation of distributed financial intelligence.

The work of Hebbar, Sharma, and Maheshkar (2026) offers a compelling response to these critiques by demonstrating how edge-based orchestration can be designed to enforce privacy, security, and performance constraints simultaneously. Their FinTech-oriented

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

framework illustrates that with appropriate architectural choices, the benefits of decentralization can be realized without sacrificing control. By embedding encryption, access control, and compliance mechanisms into microservice workflows, edge-AI systems can align with regulatory requirements while still delivering personalized and real-time services.

Future research directions emerge naturally from this discussion. One important area concerns the ethical and social implications of localized generative AI in finance. If different edge nodes generate different recommendations or risk assessments, how can fairness and transparency be ensured? Another area involves the integration of edge-AI orchestration with emerging regulatory technologies, such as automated compliance monitoring and real-time auditing. The literature suggests that AI itself may become a tool for regulatory oversight, creating a reflexive system in which intelligent services are both the subject and the instrument of governance (Castillo and Restrepo, 2024; Zeb et al., 2023).

In sum, the discussion reveals that edge-AI microservice orchestration is not simply a technical innovation but a reconfiguration of the epistemic and institutional foundations of financial technology. It challenges existing assumptions about where intelligence resides, how it is controlled, and who is responsible for its outcomes. By engaging with these theoretical dimensions, the field can move toward more thoughtful and sustainable models of digital finance.

Conclusion

The integration of edge computing, microservices architecture, and generative artificial intelligence represents a pivotal moment in the evolution of financial technology. This research has shown that edge-AI microservice orchestration offers a powerful framework for addressing the twin imperatives of real-time performance and stringent privacy protection that define modern FinTech environments. By distributing intelligent services closer to data sources while maintaining coordinated control through advanced orchestration mechanisms, financial systems can achieve unprecedented levels of responsiveness, resilience, and regulatory compliance (Hebbar, Sharma, and Maheshkar, 2026).

The theoretical and analytical exploration presented in this article demonstrates that this architectural paradigm is not merely an incremental improvement over existing cloud-based models but a fundamental rethinking of how financial intelligence is produced and governed. Through AI-driven resource allocation, adaptive service coordination, and hybrid cloud-edge deployment, microservices become dynamic agents within a distributed financial ecosystem, capable of learning, adapting, and evolving in response to changing conditions (Barua and Kaiser, 2024; Magableh and Almiani, 2019).

By situating these technical developments within broader scholarly debates about decentralization, governance, and institutional trust, the study highlights both the promise and the complexity of

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

ISSN-2750-1396

edge-AI FinTech architectures. As financial services continue to embrace generative AI, the importance of robust, transparent, and ethically grounded orchestration frameworks will only grow. Future research and practice must therefore focus not only on optimizing performance but also on ensuring that distributed financial intelligence serves the public good in a fair and accountable manner.

REFERENCES

1. Felstaine, E., and Hermoni, O. (2018). ML in Containers and Microservices. Taylor and Francis.
2. Zeb, S., Rathore, M. A., Hassan, S. A., Raza, S., Dev, K., and Fortino, G. (2023). Toward AI-enabled NextG networks with edge intelligence-assisted microservice orchestration. IEEE Wireless Communications, 30(3), 148–156.
3. Charankar, N., and Pandiya, D. K. (2024). Microservices and API deployment optimization using AI. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11), 1090–1095.
4. Hebbar, K. S., Sharma, V., and Maheshkar, J. A. (2026). Edge-AI microservice orchestration for private, real-time generative FinTech applications. Future Technology, 5(2), 13–24.
5. Kaniganti, S. T., and Challa, V. N. S. K. (2024). Leveraging microservices architecture with AI and ML for intelligent applications. International Journal of Science and Research Archive, 13(01), 3501–3511.
6. Wang, S., et al. (2019). Delay-aware microservice coordination. IEEE Transactions on Parallel and Distributed Systems.
7. Arachchige, W., and Thamoda, S. (2024). Navigating microservices with AI: design patterns and communication techniques in modern IT industries.
8. Moreschini, S., Pour, S., Lanese, I., Balouek-Thomert, D., Bogner, J., Li, X., and Taibi, D. (2023). AI techniques in the microservices life-cycle: a survey. arXiv preprint arXiv:2305.16092.
9. Taherizadeh, S., et al. (2018). A capillary architecture for dynamic orchestration. Sensors, 18(9), 2938.
10. Castillo, J., and Restrepo, M. (2024). Artificial intelligence and microservices architecture driving innovation in human resource management. Journal of Advanced Computing Systems, 4(9), 8–25.
11. Barua, B., and Kaiser, M. S. (2024). AI-driven resource allocation framework for microservices in hybrid cloud platforms. arXiv preprint arXiv:2412.02610.
12. Vudayagiri, V. (2024). Scalable AI-driven microservices architectures for distributed cloud environments. International Journal of Computer Engineering and Technology, 15(6), 154–168.
13. Oye, E., Frank, E., and Owen, J. (2024). Microservices architecture for large-scale AI applications.
14. Magableh, B., and Almiani, M. (2019). Deep Q-learning for self-adaptive microservices. IEEE Access.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 02 Pages: 74-86
OCLC – 1368736135

15. Singhal, N., Sakthivel, U., and Raj, P. (2019). Orchestration vs. choreography. International Journal of Web Engineering.
16. Pandiya, D. K., and Charankar, N. (2023). Integration of microservices and AI for real-time data processing. International Journal of Computer Engineering and Technology, 14(2), 240–254.
17. Alves, J. M., et al. (2019). ML4IoT: Orchestrating ML workflows in IoT. IEEE Access.