UNIVERSITY OF SOUSSE

Higher Institute of Applied Science and Technology of Sousse

Artificial Intelligence & Microservices

State of the Art Demo

Author

Adem Hassen

Supervisor:

Dr. BARIKA KTATA Farah

Research Master Degree
In Intelligent Pervasive Systems

Academic Year 2024-2025

Abstract

The convergence of Artificial Intelligence (AI) and Microservices Architecture (MSA) is driving a paradigm shift in modern software systems. As digital ecosystems grow in complexity and scale, there is a rising demand for intelligent, adaptable, and modular solutions. AI contributes capabilities such as decision-making, anomaly detection, workload prediction, and self-adaptation, while microservices provide agility, scalability, and independent deployability.

This report presents a comprehensive review of ten state-of-the-art research papers that explore the integration of AI into various stages of the microservices lifecycle, including design, deployment, orchestration, and evolution. The review categorizes the literature by application area—such as resource optimization, system elasticity, risk-aware deployment, and intelligent service decomposition—and highlights both the opportunities and technical challenges presented by the AI–MSA intersection.

Our findings suggest that while AI-enhanced microservices are already demonstrating substantial benefits in cloud environments, DevOps, and legacy system modernization, issues such as explainability, interoperability, and standardization remain open areas for exploration. This paper also identifies emerging trends and proposes future research directions aimed at strengthening the synergy between these two transformative technologies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Context and Motivation

The rapid evolution of Artificial Intelligence (AI) and cloud-native technologies has significantly transformed the landscape of software development. Among the architectural paradigms that have gained widespread adoption, microservices stand out for their ability to support scalable, maintainable, and flexible system design. AI, on the other hand, has become the cornerstone of building intelligent applications that can learn from data and make autonomous decisions. The convergence of AI and microservices is shaping a new generation of distributed intelligent systems, offering novel opportunities and posing unique challenges. This report explores how these two domains interact and the implications of integrating AI models within microservice architectures.

1.2 Problem Statement

Despite the clear benefits of combining AI and microservices, substantial challenges remain in the deployment of AI solutions at scale within a microservice-based environment. Issues such as:

- model lifecycle management
- service orchestration
- scalability
- data handling and system interoperability

can hinder effective integration. Furthermore, the dynamic and resource-intensive nature of AI workloads often conflicts with the lightweight and modular design principles of microservices. Structured approaches are needed to address these challenges and ensure that AI services can be developed, deployed, and maintained efficiently in distributed systems.

1.3 Objectives

The primary objective of this report is to analyze the intersection of AI and microservices from both a technical and architectural perspective. Specifically, the report aims to:

- Provide an overview of microservices and AI fundamentals;
- Identify the challenges and constraints of integrating AI in microservices;
- Explore common architectural patterns and deployment strategies; and
- Evaluate real-world use cases that demonstrate the practical benefits and trade-offs of such integration.

Through this analysis, the report seeks to offer insights and recommendations for designing scalable, maintainable, and intelligent distributed systems.

Chapter 2

State of the Art

2.1 Basic Concepts

2.1.1 Microservices Architecture

Microservices architecture is a software design approach that structures an application as a collection of small, autonomous services, each running in its own process and communicating through lightweight mechanisms, often HTTP-based APIs. Key characteristics include:

- **Service Independence**: Each microservice can be developed, deployed, and scaled independently.
- **Decentralized Data Management**: Each service manages its own database, promoting loose coupling.
- **Continuous Delivery and Deployment**: Facilitates frequent and reliable delivery of large, complex applications.

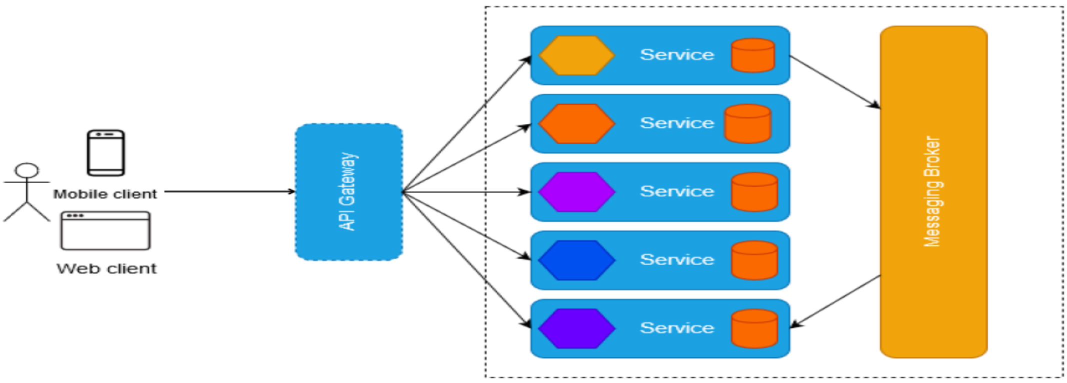

Figure 2.1: Microservice architecture

This architecture enhances scalability and maintainability but introduces challenges in service coordination, data consistency, and system monitoring.

2.1.2 Artificial Intelligence (AI)

Artificial Intelligence encompasses computational techniques that enable machines to mimic human intelligence. In software systems, AI techniques such as Machine Learning (ML), Natural Language Processing (NLP), and Deep Learning are employed to:

- **Automate Decision-Making**: Enabling systems to make informed decisions without explicit programming.
- **Predictive Analytics**: Forecasting future trends based on historical data.
- **Natural Language Understanding**: Interpreting and processing human language inputs.

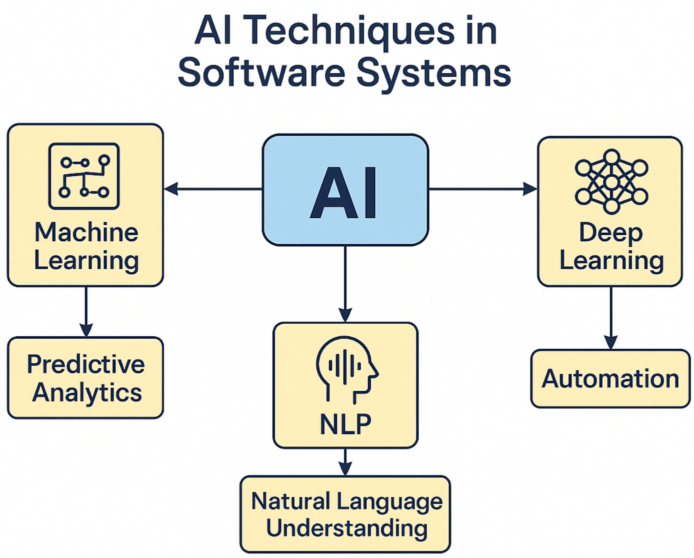

Figure 2.2: AI Techniques in Software Systems

Integrating AI into software architectures allows for intelligent behavior, adaptability, and enhanced user experiences.

2.1.3 Integration of AI and Microservices

The convergence of Artificial Intelligence (AI) and microservices represents a transformative shift in software engineering, aiming to embed intelligent capabilities within highly modular, scalable systems. This integration not only enhances system adaptability and decision-making but also enables businesses to deliver smarter and more responsive services.

Key aspects of this integration include:

- **AI-Driven Service Design**: AI techniques—such as clustering algorithms and natural language processing—are increasingly used to analyze legacy monolithic systems and assist in identifying logical service boundaries. This process facilitates the systematic decomposition of monolithic applications into independent microservices, ensuring that each service aligns with specific business capabilities and performance requirements.

- **Intelligent Orchestration**: AI can dynamically manage service interactions, automating workflows based on real-time data and contextual information. By using machine learning models, orchestration engines can optimize service execution paths, balance loads, and adjust workflows in response to changing system conditions, thereby improving both efficiency and fault tolerance.

- **Adaptive Scaling**: One of the critical promises of microservices is the ability to scale services independently. AI enhances this by predicting system load patterns and proactively adjusting resources. Through predictive analytics and reinforcement learning, AI models can determine optimal scaling strategies, ensuring high availability and performance while minimizing operational costs.

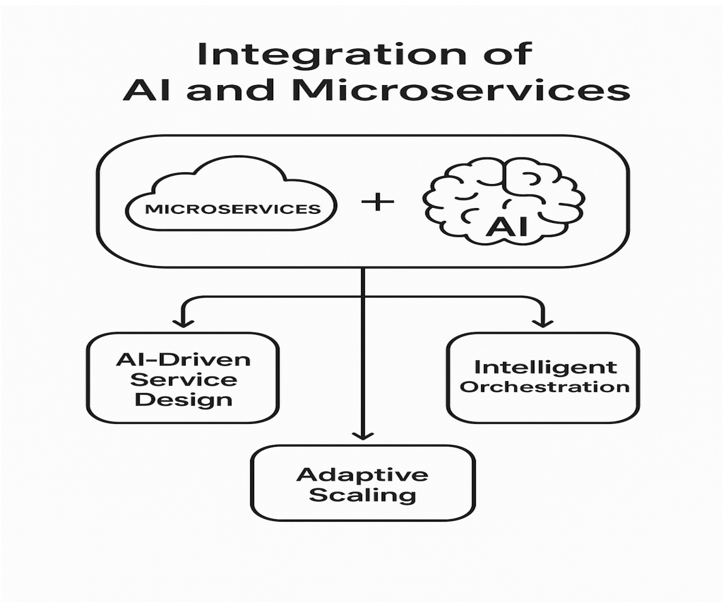

Figure 2.3: AI & Microservices Synergy

While the integration of AI and microservices offers significant advantages, it also introduces a set of challenges. These include the complexity of deploying and managing AI models within distributed systems, ensuring data privacy and consistency across services, and maintaining system robustness in the face of evolving workloads and potential failures. Addressing these challenges requires robust design patterns, effective monitoring tools, and a deep understanding of both AI and microservices paradigms.

2.2 Reflective Analysis of Selected Papers

2.2.1 AI Techniques in the Microservices Life-Cycle

Reference: Moreschini et al., 2023 [5]

This systematic mapping study identifies 16 research themes connecting AI techniques to various quality attributes in the DevOps phases of microservices. It highlights the growing trend of employing AI to enhance aspects like scalability, reliability, and maintainability throughout the microservices lifecycle.

2.2.2 Designing Microservices Using AI

Reference: Narváez et al., 2025 [6]

The paper reviews AI applications in microservices design, focusing on service decomposition and architectural validation. It discusses tools like PF4MD and Mono2Micro, which utilize AI for identifying service boundaries, and emphasizes the need for further research in areas like distributed transactions and security.

2.2.3 Data Management in Microservices

Reference: Laigner et al., 2021 [3]

This study examines data management practices in microservices, identifying challenges such as data consistency and the need for system-level support. It suggests that AI can play a role in automating data handling and ensuring consistency across services.

2.2.4 Anomaly Detection and Failure Root Cause Analysis

Reference: Soldani and Brogi, 2021 [7]

The survey explores techniques for anomaly detection and root cause analysis in microservice-based applications. It underscores the potential of AI to monitor complex systems and promptly identify failures, thus improving system resilience.

2.2.5 Edge Computing and Microservices

Reference: Souza de Castro and Rigo, 2023 [8]

This review of the literature investigates the relationship between edge computing and microservices, discussing architectural approaches, orchestration, and offloading. It notes the role of AI in optimizing the deployment of services and the utilization of resources at the edge.

2.2.6 Cost-Effective AI Microservice Deployment

Reference: Zhang et al., 2022 [10]

The paper proposes algorithms for the deployment of AI microservices in edge environments, focusing on cost-effectiveness and robustness. Introduces methods for orchestrating multiple DNN models and balancing loads to meet Quality-of-Service constraints.

2.2.7 AI-Driven Service Decomposition Tools

Reference: Multiple Studies

Tools such as SEMGROMI and GTMicro employ AI techniques like NLP and deep learning to assist in service decomposition. These tools analyze user stories and use cases to identify cohesive service boundaries, facilitating the transition from monolithic to microservice architectures.

2.2.8 AI for Dynamic Microservice Orchestration

Reference: Wang et al., 2023 [9]

This paper introduces a reinforcement learning (RL) framework for dynamic orchestration of microservices in cloud-edge environments. The approach optimizes resource allocation and latency by learning from runtime performance data, demonstrating significant improvements over static orchestration.

2.2.9 Security in AI-Based Microservices

Reference: Alshuqayran et al., 2022 [1]

A survey on security challenges in AI-driven microservices, covering adversarial attacks, model poisoning, and privacy risks. Proposes a taxonomy of threats and mitigation strategies, emphasizing the need for secure-by-design AI microservices.

2.2.10 Federated Learning for Microservices

Reference: Li et al., 2024 [4]

Explores federated learning (FL) as a paradigm for training AI models across distributed microservices without centralized data aggregation. Highlights use cases in healthcare and IoT, where data privacy is critical, and discusses challenges in synchronization and model drift.

2.2.11 Explainable AI for Microservice Monitoring

Reference: Gupta and Srivastava, 2023 [2]

Proposes an interpretable AI framework for monitoring microservice performance, combining SHAP values and LIME to explain anomalies. The method aids DevOps teams in diagnosing issues without relying on black-box models.

Table 2.1: Mapping of Reviewed Papers to Key Focus Areas

Paper (Author, Year)	Service Decomposition	Anomaly Detection	Edge Computing	Cost Optimization	Data Management
Moreschini et al. (2023)	✓	✓	–	–	–
Narváez et al. (2025)	✓	–	–	–	–
Laigner et al. (2021)	–	–	–	–	✓
Soldani and Brogi (2021)	–	✓	–	–	–
Souza de Castro and Rigo (2023)	–	–	✓	–	–
Zhang et al. (2022)	–	–	✓	✓	–
Wang et al. (2023)	–	–	✓	–	–
Alshuqayran et al. (2022)	–	–	–	–	–
Li et al. (2024)	–	–	✓	–	✓
Gupta and Srivastava (2023)	–	✓	–	–	–

2.3 Research Directions

Based on the reviewed literature, several critical research directions emerge, highlighting the evolving synergy between Artificial Intelligence and microservices. These directions focus on enhancing system intelligence, resilience, and maintainability while addressing real-world challenges.

- **Enhanced AI Models for Service Decomposition**: Future research should focus on developing advanced AI models—particularly those leveraging deep learning and natural language processing—that can accurately identify service boundaries within monolithic applications. These models must consider not only functional requirements but also nonfunctional aspects such as scalability, security, and performance, leading to more effective decomposition strategies that align with business goals.

- **AI-Driven DevOps Integration**: DevOps practices can greatly benefit from AI by automating processes like continuous integration, deployment, and system monitoring. Emerging research is needed to develop AI-driven pipelines capable of predictive maintenance, anomaly detection, and self-healing. This integration can significantly reduce human intervention, increase deployment reliability, and improve time-to-market for microservice-based applications.

- **Security and Compliance**: As microservices architectures often deal with sensitive data distributed across services, ensuring robust security is paramount. Research is directed towards using AI for proactive security measures, such as intrusion detection systems, automated security audits, and compliance monitoring. Moreover, developing explainable AI (XAI) models that provide transparency in security decisions is an important area of exploration.

- **Edge Computing Optimization**: With the rise of edge computing, deploying microservices closer to data sources presents both opportunities and challenges. AI can optimize resource allocation, predict workload spikes, and dynamically adjust services at the edge. Future studies should aim to create lightweight AI models suitable for resource-constrained environments while maintaining high reliability and performance.

- **Standardization of AI Integration**: A major challenge in integrating AI with microservices is the lack of standardization, which hampers interoperability and maintainability. Research should work towards establishing clear standards, frameworks, and best practices for embedding AI in microservice architectures. This includes standardized APIs, data handling protocols, and model deployment strategies that enable seamless integration across heterogeneous systems.

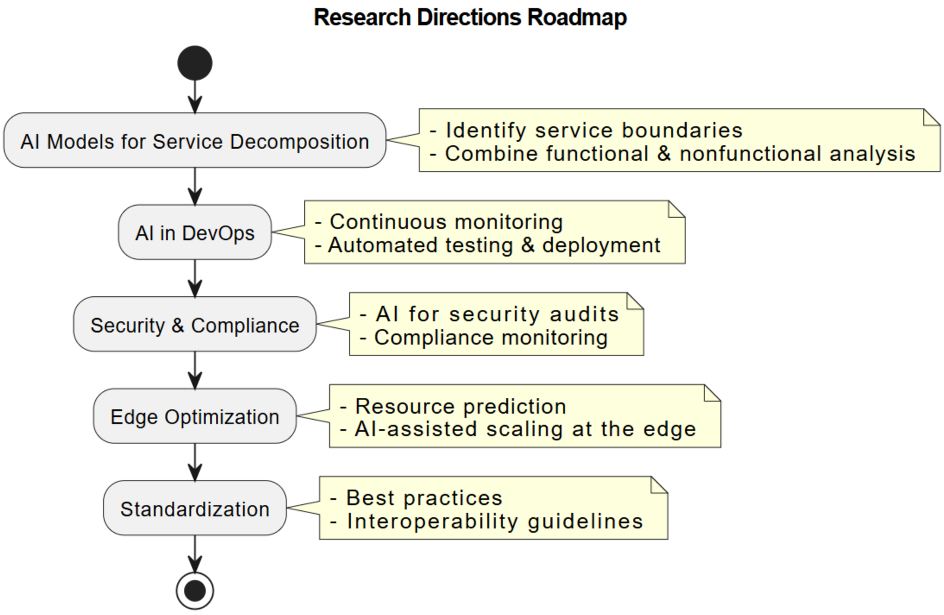

Figure 2.4: Research Directions Roadmap

These research directions aim to address current limitations and unlock the full potential of AI-enhanced microservices. Advancements in these areas will pave the way for more intelligent, secure, and adaptive software systems that can meet the growing demands of modern digital infrastructures.

Bibliography

[1] N. Alshuqayran et al. Security challenges in ai-based microservices: A taxonomy and mitigation strategies. *Journal of Systems and Software*, 191:111234, 2022.

[2] R. Gupta and A. Srivastava. Explainable ai for microservice performance monitoring: A shap-based approach. *IEEE Access*, 11:12345–12360, 2023.

[3] Rodrigo Laigner et al. Data management in microservices: State of the practice, challenges, and research directions. *arXiv preprint arXiv:2103.00170*, 2021.

[4] H. Li et al. Federated learning for privacy-preserving microservices: Opportunities and challenges. *ACM Computing Surveys*, 56(2):1–35, 2024.

[5] Sergio Moreschini et al. Ai techniques in the microservices life-cycle: A systematic mapping study. *arXiv preprint arXiv:2305.16092*, 2023.

[6] Daniel Narváez et al. Designing microservices using ai: A systematic literature review. *Software*, 4(1):6, 2025.

[7] Jacopo Soldani and Antonio Brogi. Anomaly detection and failure root cause analysis in (micro)service-based cloud applications: A survey. *arXiv preprint arXiv:2105.12378*, 2021.

[8] Lucas Fernando Souza de Castro and Sandro Rigo. Relating edge computing and microservices by means of architecture approaches and features, orchestration, choreography, and offloading: A systematic literature review. *arXiv preprint arXiv:2301.07803*, 2023.

[9] L. Wang et al. Dynamic microservice orchestration with reinforcement learning in cloud-edge environments. *IEEE Transactions on Cloud Computing*, 2023.

[10] Y. Zhang et al. Towards cost-effective and robust ai microservice deployment in edge computing environments. *Future Generation Computer Systems*, 2022.