

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
École Supérieure en Informatique -08 Mai 1945- Sidi Bel Abbès



# THESIS

To obtain the diploma of **Master**

Field: Computer Science  
Specialty: Information systems and Web

## **AI-Driven MicroService Selection and Placement in Distributed systems**

Presented by: **Emziane Mohamed**  
Submission Date: **September, 2024**

In front of the jury composed of:  
**Ms. BENCHERIF Kheira**, President  
**Mr. MALKI Abdelhamid**, Supervisor  
**Ms. LEHIRECHE Nesrine**, Examiner

Academic Year: **2023/2024**

## 0.1 Abstract

The rapid adoption of microservices in distributed systems has revolutionized how applications are developed and deployed, offering flexibility, scalability, and modularity. However, the dynamic nature of microservice architectures introduces significant challenges in managing the selection and placement of services across distributed and heterogeneous computing environments, such as cloud and edge networks. Efficient microservice selection and placement are essential for minimizing latency, ensuring fault tolerance, and optimizing resource allocation, particularly in large-scale systems where decisions must be made in real-time.

Traditional approaches, which often rely on static or heuristic-based methods, are proving insufficient in handling the complexity and scale of modern distributed systems. This thesis addresses these challenges by investigating the application of artificial intelligence (AI) and deep reinforcement learning (DRL) techniques to automate and optimize the microservice selection and placement process.

A comprehensive literature review is conducted to evaluate the current state of the art, examining trends, methodologies, and the limitations of existing approaches in service selection and placement. This work reviews novel methods leveraging artificial intelligence (AI) and deep reinforcement learning (DRL) to optimize decision-making processes in distributed systems. The reviewed models aim to dynamically adapt service placement strategies, improving overall system performance. Results from simulation-based experiments demonstrate the effectiveness of the AI-driven methods in enhancing resource management and load balancing in microservice-based architectures.

**Keywords:** MicroService, Service Placement, Service Selection, Distributed Systems, Edge Computing, AI-driven optimization, Deep Reinforcement Learning

## 0.2 Résumé

L'adoption rapide des microservices dans les systèmes distribués a révolutionné le développement et le déploiement des applications, offrant flexibilité, évolutivité et modularité. Cependant, la nature dynamique des architectures de microservices introduit des défis majeurs dans la gestion de la sélection et du placement des services à travers des environnements de calcul distribués et hétérogènes, tels que les réseaux de cloud et d'edge computing. Une sélection et un placement efficaces des microservices sont essentiels pour minimiser la latence, assurer la tolérance aux pannes et optimiser l'allocation des ressources, en particulier dans des systèmes à grande échelle où les décisions doivent être prises en temps réel.

Les approches traditionnelles, qui reposent souvent sur des méthodes statiques ou heuristiques, s'avèrent insuffisantes pour gérer la complexité et l'échelle des systèmes distribués modernes. Cette thèse aborde ces défis en étudiant l'application de l'intelligence artificielle (IA) et des techniques d'apprentissage par renforcement profond (DRL) pour automatiser et optimiser le processus de sélection et de placement des microservices.

Une revue complète de la littérature est réalisée afin d'évaluer l'état de l'art actuel, en examinant les tendances, les méthodologies et les limites des approches existantes en matière de sélection et de placement de services. Ce travail passe en revue des méthodes novatrices exploitant l'intelligence artificielle (IA) et l'apprentissage par renforcement profond (DRL) pour optimiser les processus décisionnels dans les systèmes distribués. Les modèles étudiés visent à adapter dynamiquement les stratégies de placement des services, améliorant ainsi les performances globales du système. Les résultats d'expériences basées sur des simulations démontrent l'efficacité des méthodes basées sur l'IA dans l'amélioration de la gestion des ressources et de l'équilibrage de charge dans les architectures à base de microservices.

**Mots-clés :** Microservice, Placement de service, Sélection de service, Systèmes distribués, Edge computing, Optimisation par IA, Apprentissage par renforcement profond