AVIS DE SOUTENANCE D'UNE THESE DE DOCTORAT

Le Directeur de l'Ecole Nationale des Sciences Appliquées a le plaisir d'informer le public qu'une soutenance de thèse de Doctorat en

«Sciences et ingénierie»

aura lieu le 28/11/2026 à 14H00 à l'ENSA, Kénitra

La Thèse sera présentée par Mme AHAIDOUS KHADIJA

Sous le thème :

Exploitation du Big Data dans l'éducation : Nouvelles compétences pour l'apprentissage, amélioration de l'enseignement et la rétention des étudiants

Devant le jury composé de :

Nom et Prénom	Titre	Etablissement
HMINA NABIL	Président	Université Ibn Zohr, Agadir
GAHI YOUSSEF	Rapporteur	ENSA, Kénitra
BADIR HASSAN	Rapporteur	ENSA, Tanger
HAMLICH MOHAMED	Rapporteur	ENSAM, Casablanca
ZEMZAMI MARIA	Examinateur	ENSAM, Rabat
KHIAT AZEDDINE	Examinateur	ENSET, Mohammedia
TABAA MOHAMED	Co-Directeur de thèse	EMSI, Casablanca
HACHIMI HANAA	Directeur de thèse	ENSA, Kénitra

UNIVERSITE IBN TOFAIL ★ PÔLE DES ETUDES DOCTORALES ★ KENITRA

Nom et Prénom : AHAIDOUS KHADIJA
Date de soutenance : 28/11/2026
Directeur de Thèse : HACHIMI HANAA

Sujet de thèse :

Exploitation du Big Data dans l'éducation : Nouvelles compétences pour l'apprentissage, amélioration de l'enseignement et la rétention des étudiants

Abstract:

Avec la croissance rapide des applications riches en données, la diversité des modèles de données a apporté de nombreux avantages mais aussi de nouveaux défis pour la gestion des bases de données. Les organisations combinent de plus en plus différents types de bases de données.. Les solutions actuelles reposent sur des architectures monolithiques ou des intégrations polyglottes faiblement couplées, soulevant des défis de flexibilité, de cohérence et d'interrogation. Les grands modèles de langage ouvrent une voie prometteuse vers l'interrogation en langage naturel, mais introduisent de nouvelles difficultés : coût en tokens, hallucinations et latence. Un autre défi est la scalabilité de l'orchestration multi-bases.
Cette thèse propose un framework intelligent combinant microservices, sérialisation orientée token et agents LLMs pour le stockage et l'interrogation de données hétérogènes. Nous introduisons HDSM, validé sur la plateforme EMSI XR. Nous évaluons TOON, un format optimisé, atteignant une réduction de 39,6 % des tokens et un gain d'efficacité de 76 % par rapport à JSON. Nous étudions le routage et la décomposition de requêtes par LLMs, révélant un compromis précision-latence et une limite de 69 % d'exactitude en orchestration multi-bases.
Nous proposons enfin HeteroMAS, un système multi-agents où des agents collaborent pour fusionner des preuves distribuées.
Ce framework montre un fort potentiel pour améliorer la gestion des bases de données hétérogènes, du stockage à la décision multi-source, ouvrant des perspectives pour l'IA en contexte polyglotte.

Mots-clés : systèmes polyglottes, microservices, sérialisation orientée token, grands modèles de langage, LLM, routage de requêtes, systèmes multi-agents

Résumé :

With the rapid growth of data-intensive applications, especially in enterprise information systems and educational platforms, the diversity of data models has brought many benefits but also new challenges for database management. Organizations increasingly combine different database types to support modern workloads. Current solutions rely mainly on monolithic architectures or loosely coupled polyglot integrations. However, this heterogeneity raises challenges related to storage flexibility, cross-database consistency and query interrogation. The emergence of LLMs has opened a promising path toward natural language querying, but introduces new difficulties, notably token cost, hallucinations and latency, which must be carefully managed.
This thesis proposes a smart framework for heterogeneous data storage, representation and querying. Our solution relies on the combined use of microservices architecture, token-oriented serialization and LLM-based agents. To ensure flexible, consistent storage, we introduce HDSM, a microservices-based framework validated on the EMSI XR educational platform. To reduce LLM inference costs, we evaluate TOON, a token-optimized serialization format, achieving a 39.6% token reduction and a 76% efficiency gain over JSON. To address query interrogation, we study LLM-based routing and decomposition, revealing an accuracy-latency trade-off and a 69% accuracy limit for complex multi-database orchestration. To overcome this, we propose HeteroMAS, a formalized multi-agent system where. specialized agents collaborate to fuse evidence across heterogeneous sources.

Keywords : polyglot systems, microservices, token serialization, large language models, LLM, query routing, multi-agent systems