

FICHE DE PROPOSITION DE SUJET DE THÈSE DE DOCTORAT

Intitulé du sujet

Sécurité et robustesse des grands modèles de langage (LLM) : approches adaptatives de défense et de protection des données dans des environnements souverains

Encadrants : Pr. Mohamed Nabil Saidi, Pr. Imane Chlioui, Pr. Manal El Bajta

Mots-clés : Large Language Models (LLM), cybersécurité, apprentissage automatique, attaques adversariales, robustesse, confidentialité des données, IA de confiance, souveraineté numérique.

1. Contexte et problématique

Les grands modèles de langage (Large Language Models – LLM) connaissent un développement rapide et sont de plus en plus utilisés dans des applications nécessitant le traitement et la génération de données sensibles. Malgré leurs capacités avancées, ces modèles présentent des vulnérabilités susceptibles de compromettre leur sécurité, leur robustesse et la confidentialité des informations traitées.

Parmi les menaces étudiées figurent les attaques adversariales, les tentatives de contournement des mécanismes de sécurité, l'extraction de données mémorisées, l'empoisonnement des données d'entraînement et les attaques par portes dérobées (*backdoors*). Ces menaces soulèvent des défis scientifiques liés à la caractérisation des vulnérabilités, à la conception de mécanismes de défense efficaces et à l'évaluation de la résistance des modèles face à des stratégies d'attaque évolutives.

Dans ce contexte, cette thèse vise à étudier les vulnérabilités des LLM et à proposer de nouvelles approches permettant de renforcer leur sécurité et leur robustesse. Une attention particulière sera accordée aux mécanismes adaptatifs de détection et de mitigation des attaques, à la protection des données et à l'évaluation expérimentale des solutions proposées. Les travaux s'inscriront dans une perspective de souveraineté numérique, en privilégiant des environnements d'expérimentation maîtrisés.

3. Axes de recherche

Axe 1 : Analyse des vulnérabilités et caractérisation des attaques contre les LLM

Étude des principales familles d'attaques, notamment les attaques adversariales, les jailbreaks, les attaques par empoisonnement des données et les mécanismes d'extraction d'informations. Cet axe permettra d'identifier les limites des approches actuelles et de dégager les verrous scientifiques prioritaires.

Axe 2 : Conception de mécanismes de défense et amélioration de la robustesse

Développement de méthodes de défense adaptées aux vulnérabilités identifiées : détection des entrées malveillantes, apprentissage orienté robustesse, adaptation des mécanismes de protection et stratégies de mitigation. L'objectif sera d'améliorer la résistance des modèles tout en préservant leurs performances sur les tâches légitimes.

Axe 3 : Protection des données et évaluation expérimentale de la sécurité

Étude des risques de mémorisation et de divulgation de données, des attaques portant sur la confidentialité et des méthodes de protection associées. Cet axe inclura la définition de protocoles expérimentaux, la comparaison avec des méthodes de référence et l'évaluation des compromis entre sécurité, utilité et coût de calcul.

Les orientations définitives et la sélection des familles d'attaques seront précisées à l'issue d'une revue approfondie de la littérature scientifique.

7. Références bibliographiques indicatives

1. Das, B. C., Amini, M. H., & Wu, Y. (2025). *Security and Privacy Challenges of Large Language Models: A Survey*. ACM Computing Surveys, 57(6), Article 152. <https://doi.org/10.1145/3712001>
2. Liao, Z., Chen, K., Lin, Y., Li, K., Liu, Y., Chen, H., Huang, X., & Yu, Y. (2025). *Attack and Defense Techniques in Large Language Models: A Survey and New Perspectives*. Neural Networks. <https://doi.org/10.1016/j.neunet.2025.108388>
3. Li, M. Q., & Fung, B. C. M. (2025). *Security Concerns for Large Language Models: A Survey*. Journal of Information Security and Applications, 95, Article 104284. <https://doi.org/10.1016/j.jisa.2025.104284>
4. Chen, K., Zhou, X., Lin, Y., Feng, S., Shen, L., et al. (2025). *A Survey on Privacy Risks and Protection in Large Language Models*. Journal of King Saud University Computer and Information Sciences, 37, Article 163. [https://doi.org/10.1007/s44443-025-00177-](https://doi.org/10.1007/s44443-025-00177-1)

5. OWASP Foundation. (2025). *OWASP Top 10 for LLM Applications*. <https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/>
6. Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. <https://doi.org/10.6028/NIST.AI.600-1>

Sujet 1 : Conception de systèmes de recommandation pour les réseaux sociaux : adaptation dynamique à l'évolution des intérêts des utilisateurs à l'aide de l'intelligence artificielle

Directeur de thèse

Mohamed Nabil SAIDI

Co-Directeur de thèse

Khaoula EL Habyb

Présentation du sujet.

Cette proposition porte sur la conception de méthodes d'intelligence artificielle capables de repérer l'évolution des intérêts des utilisateurs et d'actualiser les contenus qui leur sont recommandés sur les réseaux sociaux. Elle s'inscrit dans la continuité méthodologique de travaux en IA et en apprentissage automatique, tout en ouvrant un nouveau domaine d'application.

Contexte et justification

Les systèmes de recommandation des réseaux sociaux s'appuient sur les interactions des utilisateurs pour proposer des contenus pertinents. Comme leurs intérêts évoluent, cette recherche étudiera comment détecter ces changements et adapter les recommandations sans accorder trop d'importance aux interactions ponctuelles.

Problématique

Comment exploiter les méthodes d'intelligence artificielle pour détecter l'évolution des intérêts des utilisateurs et adapter, de manière continue, les recommandations proposées sur les réseaux sociaux ?

Objectif général

Développer et évaluer des modèles de recommandation capables de tenir compte de l'évolution temporelle des intérêts des utilisateurs à partir de leurs interactions avec les contenus des réseaux sociaux.

Axes de recherche

Axe 1 — Analyse des interactions et représentation des intérêts

Identifier les interactions pertinentes et analyser leur évolution afin de modéliser les préférences des utilisateurs en tenant compte de leur récence, de leur fréquence et de leur nature.

Axe 2 — Modélisation de l'évolution des préférences

Comparer les principales approches de recommandation et évaluer l'apport des modèles temporels pour distinguer les préférences durables des intérêts émergents.

Axe 3 — Adaptation des recommandations

Développer une stratégie d'actualisation des profils et des recommandations, puis évaluer son équilibre entre stabilité et adaptation rapide aux nouveaux intérêts.

Axe 4 — Évaluation et responsabilité

Évaluer chronologiquement les modèles selon la pertinence, la réactivité et la diversité des recommandations, tout en examinant les risques liés aux données personnelles et à la répétition des contenus.

Contributions attendues

Une méthode de recommandation qui prend en compte l'évolution temporelle des intérêts des utilisateurs.

Formulation courte du sujet

Cette thèse étudie l'adaptation des systèmes de recommandation des réseaux sociaux à l'évolution des intérêts des utilisateurs. Elle exploite les méthodes d'intelligence artificielle pour analyser les interactions dans le temps, détecter les changements de préférences et actualiser les recommandations. Les approches proposées seront évaluées selon leur pertinence, leur réactivité et la diversité des contenus recommandés.

Sujet 2 : Intelligence artificielle pour l'aide au contrôle et à l'audit dans le secteur financier

Directeur de thèse

Mohamed Nabil SAIDI

Co-Directeur de thèse

Khaoula EL Habyb

Présentation du sujet.

Le secteur financier traite de grands volumes de données et fait face à des risques variés. La fraude, le risque de crédit et les cybermenaces évoluent rapidement. Les méthodes traditionnelles peinent parfois à détecter ces nouvelles menaces. L'intelligence artificielle peut améliorer leur détection et leur analyse. Ce sujet étudiera l'apport de modèles adaptés aux différents risques financiers. Il prendra aussi en compte l'explicabilité, l'équité et les exigences réglementaires.

Problématique de recherche

Comment concevoir et déployer des systèmes d'intelligence artificielle capables d'améliorer la détection des risques et l'aide à la décision dans la finance, tout en garantissant leur performance, leur explicabilité, leur conformité réglementaire et leur fiabilité dans la durée ?

Objectifs de la recherche

Concevoir et valider un cadre méthodologique et un prototype opérationnel d'aide à la décision fondés sur l'intelligence artificielle, destinés à renforcer les capacités de détection de la fraude, d'analyse des risques et d'optimisation des recettes.

Axes de travail

Axe 1 – État de l'art et cadrage réglementaire

Examiner les principales méthodes d'IA appliquées à la finance, notamment les arbres boostés, les réseaux de neurones, la détection d'anomalies, l'apprentissage sur graphes et le traitement du langage naturel. Analyser leur conformité aux exigences réglementaires, notamment le RGPD, le règlement européen sur l'IA et la LCB-FT, puis identifier les cas d'usage prioritaires.

Axe 2 – Données et ingénierie

Préparer des données financières publiques ou synthétiques en les nettoyant et en les anonymisant. Construire un pipeline reproductible intégrant la création de variables pertinentes, le traitement du déséquilibre des classes et la prise en compte de la dimension temporelle.

Axe 3 – Modélisation et aide à la décision

Développement et comparaison de modèles (référence statistique vs. modèles avancés).

Optimisation selon des métriques métier (coût des erreurs, rappel, précision, seuils de décision).

Axe 4 – Explicabilité, robustesse et équité

Évaluer l’explicabilité des modèles à l’aide de méthodes comme SHAP, LIME et les explications contrefactuelles. Tester leur robustesse et auditer les biais afin de proposer des mesures correctives adaptées.

Contributions attendues

Un cadre méthodologique combinant performance prédictive et explicabilité pour l'audit dans le secteur financiers, transférable à d'autres domaines.

Sujet 3 :Vers une évaluation contextualisée de la fiabilité des métriques XAI

Directeur de thèse

Mohamed Nabil SAIDI

Co-Directeur de thèse

Khaoula EL Habyb

Description du sujet :

Les méthodes XAI comme SHAP et LIME expliquent les prédictions des modèles, mais la qualité de ces explications reste difficile à garantir. Les métriques d'évaluation peuvent donner des résultats contradictoires selon les modèles, les données et les protocoles utilisés. Cette recherche étudiera leur fiabilité et les conditions dans lesquelles elles peuvent conduire à des conclusions trompeuses.

Problématique

Comment caractériser les conditions expérimentales dans lesquelles les métriques d'évaluation XAI permettent, ou non, de mesurer correctement les propriétés des explications qu'elles sont censées évaluer ?

Objectifs de la recherche

Évaluer la fiabilité des métriques utilisées pour mesurer la qualité des explications produites par les méthodes d'IA explicable.

Objectifs spécifiques :

- Recenser et caractériser les principales métriques d'évaluation des explications XAI.
- Comparer leurs résultats selon les modèles, les jeux de données et les méthodes d'explication utilisés.
- Analyser les facteurs pouvant expliquer les évaluations contradictoires.
- Identifier les conditions dans lesquelles ces métriques sont fiables ou susceptibles d'induire en erreur.
- Proposer des recommandations pour évaluer les explications de manière plus rigoureuse.

Principaux axes de travail envisages

- Axe 1 : Analyse et structuration des métriques d'évaluation XAI existantes

Réaliser une revue approfondie des métriques d'évaluation des explications, en particulier des métriques de fidélité, stabilité et robustesse, afin d'identifier leurs principes, hypothèses, limites et conditions d'utilisation.

- **Axe 2 : Étude expérimentale de la validité des métriques**

Cet axe consiste à concevoir des expérimentations contrôlées, notamment à partir de données synthétiques et de modèles dont certaines propriétés sont connues, afin d'examiner si les métriques mesurent effectivement les propriétés visées.

- **Axe 3 : Analyse de l'influence des conditions expérimentales**

Cet axe vise à étudier l'influence de facteurs tels que la corrélation entre variables, le bruit, la dimensionnalité, la complexité du modèle ou les stratégies de perturbation sur les résultats des métriques. Des études d'ablation permettront d'isoler l'effet de chaque facteur et d'identifier les situations dans lesquelles les métriques deviennent moins fiables.

- **Axe 4 : Caractérisation des conditions de fiabilité et d'échec**

Cet axe vise à caractériser les conditions de validité et d'échec des métriques XAI, en établissant une cartographie reliant les caractéristiques des données, des modèles et des protocoles expérimentaux au comportement et au niveau de fiabilité des métriques. Cette analyse pourrait conduire à l'élaboration d'un cadre méthodologique permettant d'orienter leur sélection selon le contexte d'utilisation.

Sujet 4 : Hardware-Aware Multi-Objective Neural Architecture Search for Resource-Constrained Edge AI

Directeur de thèse

Mohamed Nabil SAIDI

Résumé du sujet

L'intelligence artificielle embarquée (Edge AI) exige des modèles à la fois performants et compatibles avec des ressources de calcul, de mémoire et d'énergie limitées. Cette thèse propose de développer des approches de recherche automatique d'architectures neuronales (Neural Architecture Search — NAS) multi-objectifs, intégrant explicitement les caractéristiques des plateformes matérielles cibles. Les méthodes étudiées viseront à concilier précision prédictive, latence d'inférence, consommation énergétique, empreinte mémoire et taille des modèles. Les travaux pourront également explorer la quantification et le transfert ou l'adaptation des architectures entre dispositifs hétérogènes, afin de faciliter le déploiement efficace de l'IA sur des systèmes embarqués contraints.

Problématique scientifique

Comment concevoir une méthode NAS capable d'explorer efficacement l'espace des architectures neuronales et d'identifier des compromis de Pareto robustes entre qualité prédictive et coûts de déploiement, tout en tenant compte des contraintes et des différences de comportement des plateformes Edge ?

Objectifs de recherche

- Concevoir un cadre NAS multi-objectifs intégrant la précision, la latence, l'énergie, la mémoire et la taille des modèles.
- Développer des stratégies de recherche et d'estimation des coûts sensibles aux caractéristiques du matériel cible.
- Étudier la quantification et, lorsque pertinent, son intégration dans l'optimisation des architectures.
- Analyser la transférabilité et l'adaptation des architectures sur différentes plateformes Edge hétérogènes.
- Valider expérimentalement les compromis obtenus face à des architectures et méthodes NAS de référence.

Résultats attendus

Un cadre méthodologique et des algorithmes NAS hardware-aware multi-objectifs, des modèles optimisés pour plusieurs cibles Edge, ainsi qu'un protocole d'évaluation reproductible mettant en évidence les compromis entre performance prédictive et efficacité de déploiement.

Mots-clés : Edge AI ; Neural Architecture Search ; optimisation multi-objectifs ; hardware-aware NAS ; quantification ; efficacité énergétique.

Sujet 5 : Securing Sensitive Actions of AI Agents in Public Information Systems

Directeur de thèse

Mohamed Nabil SAIDI

Co-Directeur de thèse

Abderrahim EL Qadi

Résumé du sujet

L'intégration d'agents d'intelligence artificielle dans les systèmes d'information publics ouvre de nouvelles perspectives d'automatisation des services et des processus administratifs. Toutefois, lorsque ces agents disposent de capacités d'action — consultation de données confidentielles, modification de dossiers, validation de procédures ou exécution d'opérations via des API — ils introduisent de nouveaux risques de sécurité, notamment les instructions malveillantes (prompt injection), l'abus de privilèges, les erreurs d'autorisation et les actions non conformes aux règles métiers. Cette thèse vise à concevoir et à évaluer un cadre de sécurisation des actions sensibles des agents d'IA, fondé sur le contrôle d'accès contextuel, la vérification des intentions, l'application de politiques de sécurité et la traçabilité des décisions, en tenant compte des exigences de confidentialité, de responsabilité et d'auditabilité propres aux administrations publiques.

Objectif général

Proposer une architecture de confiance permettant d'encadrer, d'autoriser, de surveiller et d'auditer les actions sensibles réalisées par des agents d'IA au sein des systèmes d'information publics, sans compromettre leur efficacité opérationnelle.

Objectifs spécifiques

- Identifier et formaliser les actions sensibles, les scénarios de menace et les exigences de sécurité des agents d'IA dans les processus administratifs.
- Concevoir un mécanisme d'autorisation dynamique combinant identité de l'agent, contexte, niveau de risque, privilèges minimaux et règles métiers.
- Définir des dispositifs de prévention et de contrôle des actions non autorisées, notamment face aux attaques par prompt injection et aux usages abusifs des outils et API.
- Mettre en place une traçabilité explicable et vérifiable des décisions et des actions, avec des possibilités de validation humaine pour les opérations critiques.
- Développer un prototype et évaluer la solution sur des cas d'usage représentatifs du secteur public, selon des critères de sécurité, de conformité, de performance et d'utilisabilité.

Résultats attendus

Un modèle de menace adapté aux agents d'IA, un cadre d'autorisation et de contrôle des actions sensibles, un prototype expérimental ainsi qu'un protocole d'évaluation reproductible destiné aux systèmes d'information publics.

Sujet 6 : Exploitation de l'IA générative et des grands modèles de langage (LLMs) pour la gestion intelligente des données et l'automatisation des processus métier

Directeur de thèse

Mohamed Nabil SAIDI

Co-Directeur de thèse

Oussama Hamal, Khaoula EL Habyb

Contexte et motivation

Les organisations détiennent des données toujours plus volumineuses et hétérogènes (documents, tableurs, courriels, bases de données, flux de capteurs), mais leur collecte, leur nettoyage, leur contrôle et leur exploitation restent en grande partie manuels ou fondés sur des règles, ce qui est lent et fragile.

L'IA générative et les grands modèles de langage (LLMs) changent la donne : ils lisent des contenus non structurés, écrivent des requêtes et du code, et raisonnent sur des tâches en plusieurs étapes. Associés à la recherche d'information (RAG) et à l'appel d'outils, ils permettent de construire des agents qui exécutent des tâches de gestion de données et de processus de bout en bout.

Leur déploiement en contexte métier soulève toutefois des problèmes ouverts : fiabilité des sorties générées, intégration aux systèmes existants, gouvernance et confidentialité des données, et évaluation. Cette thèse traite ces problèmes pour la gestion intelligente des données et l'automatisation des processus métier.

Questions de recherche et objectifs

La thèse cherche comment des agents à base de LLM peuvent traiter les données et les processus métier de façon fiable, sûre et mesurable.

Questions de recherche

1. Comment combiner les LLMs avec la recherche d'information (RAG), des sources de données structurées et des outils pour extraire, nettoyer, intégrer et interroger de façon fiable des données hétérogènes ?
2. Comment des agents à base de LLM peuvent-ils orchestrer des processus métier en plusieurs étapes (lecture, validation, décision, action) tout en conservant un contrôle humain approprié ?
3. Comment évaluer et garantir la qualité, la robustesse, la sécurité et la confidentialité de tels systèmes, y compris avec des modèles hébergés localement ?

Objectifs

- **Proposer une architecture d'agents pilotés par LLM couvrant l'ingestion, le contrôle qualité, l'intégration et l'accès aux données.**
- **Développer des méthodes combinant LLMs et composants déterministes (validation de schémas, points d'approbation, permissions).**
- **Définir un cadre d'évaluation : exactitude, taux d'hallucination, coût, latence et traçabilité.**
- **Valider l'approche sur des cas d'usage réels issus de plusieurs domaines.**

Axe de travail

Le travail s'organise en cinq lots sur trois ans, de l'état de l'art et des prototypes jusqu'à l'évaluation sur des cas d'usage réels.

- 1. État de l'art et formalisation : panorama des agents LLM, du RAG, de l'appel d'outils et de la gestion de données par LLM ; définition formelle des tâches visées.**
- 2. Ingestion intelligente et qualité des données : extraction par LLM depuis des fichiers et des API, mise en correspondance de schémas, détection d'anomalies, toujours suivies d'une validation déterministe.**
- 3. Automatisation agentique des processus : agents avec appel d'outils, RAG et orchestration de workflows, avec validation humaine pour les actions sensibles.**
- 4. Confiance, sécurité et évaluation : métriques, benchmarks et jeux de test ; étude des hallucinations, de l'injection de prompt, de la confidentialité et des modèles hébergés localement.**
- 5. Études de cas et rédaction : validation sur des cas d'usage réels, publications et rédaction du manuscrit.**

PROPOSITION DE SUJET DE THÈSE

Directrice de thèse : Rajaa SAIDI, Professeure de l'Enseignement Supérieur à l'INSEA

Mail : r.saidi@insea.ac.ma

Titre : Piloter l'entreprise autonome : gouvernance et confiance dans les systèmes de gestion des processus métier agents (A-BPMS).

Title : Steering the Autonomous Enterprise: Governance and Trust in Agentic Business Process Management Systems (A-BPMS).

1. Contexte et Problématique

La gestion classique des processus métier (BPM) et l'automatisation robotisée des processus (RPA) reposent sur des flux de travail rigides, dictés à l'avance ("if this, then that"). L'intégration récente de l'IA générative et des grands modèles de langage (LLM) donne naissance aux Systèmes de Gestion de Processus Métier Agentiques (A-BPMS). Dans ce nouveau paradigme, les flux rigides sont remplacés par des agents IA autonomes orientés vers des objectifs spécifiques.

Ces agents sont capables de percevoir leur environnement opérationnel, de raisonner face à des données non structurées et de s'auto-modifier. Cependant, une autonomie totale en entreprise pose des risques majeurs : dérives décisionnelles, non-conformité légale et opacité (effet "boîte noire"). Pour que les entreprises adoptent l'ABPM, il est indispensable de formaliser l'autonomie encadrée (framed autonomy) : comment laisser l'IA choisir la meilleure trajectoire d'exécution tout en lui imposant des barrières de sécurité et des règles métier inviolables ?

2. Objectifs de la Thèse

L'objectif central de cette thèse est de concevoir un framework théorique et applicatif permettant de modéliser, de piloter et de sécuriser la collaboration entre agents autonomes et humains au sein d'un A-BPMS.

- **Formalisation de l'autonomie encadrée** : Développer un langage de contraintes ou un modèle mathématique capable de traduire des règles de conformité (RGPD, contraintes financières) en directives dynamiques pour les agents.
- **Mécanismes d'auditabilité et d'explicabilité** : Concevoir des architectures de traçabilité permettant de comprendre en temps réel les décisions d'un agent multi-systèmes (par exemple via des techniques avancées de process mining).
- **Optimisation des boucles de rétroaction (Human-in-the-loop)** : Créer des protocoles de délégation où l'agent sait évaluer son propre niveau d'incertitude et passer le relais à un expert humain de manière fluide.

Sujets de thèse proposés par Pr. Rachid BENMANSOUR (Laboratoire SI2M)

Sujet 1 : Modélisation et optimisation des chaînes logistiques exportatrices face aux contraintes carbone

Encadrement :

- Pr. Rachid BENMANSOUR (r.benmansour@insea.ac.ma)
- Pr. Abdeslam KADRANI (akadrani@insea.ac.ma)

Sujet 2 : Modélisation et optimisation des opérations portuaires de vrac : coordination des flux, gestion des capacités et résilience

Encadrement :

- Pr. Rachid BENMANSOUR (r.benmansour@insea.ac.ma)
- Pr. Abdeslam KADRANI (akadrani@insea.ac.ma)

Sujet 3 : Optimisation du déploiement des infrastructures de recharge des véhicules électriques : localisation et dimensionnement.

Encadrement :

- Pr. Rachid BENMANSOUR (r.benmansour@insea.ac.ma)
- Pr. Abdeslam KADRANI (akadrani@insea.ac.ma)

Sujet 4 : Optimisation assistée par apprentissage pour l'ordonnancement de machines parallèles avec ressources partagées et événements dynamiques.

Encadrement :

- Pr. Rachid BENMANSOUR (r.benmansour@insea.ac.ma)

Sujet 5 : Optimisation des systèmes de logistique inverse sous incertitude : une approche par simulation pour une logistique verte au service de l'économie circulaire.

Encadrement :

- Pr. Rachid BENMANSOUR (r.benmansour@insea.ac.ma)
- Pr. MIMOUNI Faycal (fmimouni@insea.ac.ma)

Profil recherché :

Titulaire d'un Master ou d'un diplôme d'ingénieur en mathématiques appliquées, génie industriel, recherche opérationnelle, informatique (ou équivalent).

Compétences et qualités requises :

- Bonnes connaissances en optimisation et en recherche opérationnelle.
- Intérêt pour la programmation et les méthodes d'optimisation.
- Bonne maîtrise de l'anglais scientifique.

Pour toute demande d'information, veuillez contacter : r.benmansour@insea.ac.ma

Laboratoire SI2M -- INSEA
Mathématiques Appliquées : Modélisation & optimisation
Sujets de thèse 2026 proposés par Pr. Kadrani

Axe 1 : Modélisation et aide à la décision dans les domaines industriel et hospitalier
(Encadrement : Prs. Kadrani et Benmansour)

Sujet 1. Développement d'un cadre méthodologique basé sur l'IA pour la maintenance prédictive et le pilotage proactif des systèmes industriels complexes.

Sujet 2. Optimisation (déterministe vs stochastique) de la gestion et de l'ordonnancement des ressources dans les blocs opératoires hospitaliers : Cas Maroc.

→ Pour plus d'informations sur les deux sujets, contacter **Pr. Kadrani** : akadrani@insea.ac.ma

Profil du candidat :

- Ingénieur et/ou Master en Mathématiques appliquées, Génie Industriel ou Recherche Opérationnelle.
- Le candidat devra disposer des compétences suivantes :
 - Modélisation de problèmes de la recherche opérationnelle ;
 - Méthodes d'optimisation (exactes et/ ou approchées) ;
 - Bonne maîtrise d'au moins un langage de programmation (C, C++, Python, etc..)

Axe 2 : Approches théoriques et algorithmiques pour les problèmes d'optimisation dégénérés

Sujet 3. Développer et analyser des méthodes de scalarisation-relaxation permettant de caractériser, d'approcher et de calculer les solutions des problèmes d'optimisation bi-niveaux vectoriels, dans lesquels les décisions d'un niveau supérieur dépendent des solutions d'un problème d'optimisation inférieur.

Sujet 4. Développer un cadre de dualité adapté à la structure particulière des problèmes de la programmation mathématique avec contraintes de complémentarité (MPCC), puis concevoir des méthodes de résolution exploitant les propriétés de cette dualité.

→ Pour plus d'informations sur les deux sujets, contacter **Pr. Kadrani** : akadrani@insea.ac.ma

Profil du candidat :

- Master ou équivalent en mathématiques appliquées, optimisation, ou recherche opérationnelle.
- Bonne maîtrise de l'analyse convexe et de l'optimisation non linéaire.
- Intérêt pour la rigueur théorique et l'implémentation des algorithmes

PhD Research Opportunities in Game Theory

PhD Supervisor: Professor Ahmed Doghmi

Research Laboratory SI2M

National Institute of Statistics and Applied Economics (INSEA)
Rabat, Morocco

Research Areas: Game Theory, Mathematical Economics, Social Choice Theory, and Mechanism Design.

Two PhD research topics are proposed in implementation theory, with applications to economic and political environments.

1. Topic 1: Essays in Game Theory: Implementation via Mediated Strong Equilibrium and Economic Applications

Research Topic Description

Implementation theory studies the mechanisms through which social choice correspondences (SCCs) can be implemented on the basis of individuals' preferences. A central question is to identify conditions under which a social choice correspondence can be implemented in equilibrium, taking into account strategic behavior and the possibility of coalitional deviations.

This PhD project focuses on implementation via mediated strong equilibrium. Its main objective is to establish simple sufficient conditions for implementing social choice correspondences and to investigate the role of mediation in ensuring coalitional stability and expanding the class of implementable correspondences.

Particular attention will be devoted to social choice correspondences that are not implementable in the classical Strong Nash Equilibrium (SNE) framework but may become implementable through mediated mechanisms.

The theoretical results will be explored in two economic environments:

- **Matching markets:** Implementation of matching correspondences, stability of matchings, and the analysis of coalitional deviations.
- **Private-goods economies:** Implementation of allocations, Pareto efficiency, and stability against coalitional deviations.

The aim is to develop new theoretical results on implementation via mediated strong equilibrium and investigate their implications in these two areas of mathematical economics.

Research Objectives

The research will focus on the following objectives:

1. Establish simple sufficient conditions for the implementation of social choice correspondences via mediated strong equilibrium.

2. Identify social choice correspondences that are not implementable via classical Strong Nash Equilibrium and investigate conditions under which they can be implemented through mediation.
3. Apply the theoretical results to matching markets and private-goods economies.
4. Study the role of monotonicity, effectivity functions, and coalitional stability in implementation results.

References

1. Peleg, B., & Procaccia, A. D. (2010). *Implementation by mediated equilibrium*. International Journal of Game Theory, 39, 191–207. <https://doi.org/10.1007/s00182-009-0175-4>
 2. Bouchta, M., Doghmi, A., & Ziad, A. (2026). *Strong Nash implementation and economic applications*. Games and Economic Behavior, 159, 286–309. <https://doi.org/10.1016/j.geb.2026.06.010>
-

2. Topic 2: Essays in Game Theory: New Sufficient Conditions for Bayesian Implementation of Social Choice Correspondences

Research Topic Description

Implementation theory under incomplete information investigates the conditions under which social choice correspondences can be implemented through mechanisms when individuals possess private information about their preferences or the economic environment.

This PhD project focuses on Bayesian implementation. Its main objective is to develop new sufficient conditions for implementing a broad class of social choice correspondences (SCCs) in Bayesian equilibrium.

The research will investigate the strategic properties of Bayesian equilibrium and identify conditions that guarantee implementation across different environments. Particular attention will be devoted to extending existing implementation results and establishing new theoretical conditions applicable to economic and political domains.

Potential applications include resource allocation, exchange economies, collective decision-making, and political mechanisms.

The aim is to contribute to the theory of Bayesian implementation by expanding the class of social choice correspondences for which implementation can be established under incomplete information.

Research Objectives

The research will focus on the following objectives:

1. Develop new sufficient conditions for Bayesian implementation of social choice correspondences.
2. Identify strategic and informational properties that guarantee implementation in Bayesian equilibrium.

3. Establish implementability results for a broad class of social choice correspondences in economic and political environments.
4. Investigate the relationships between existing Bayesian implementation conditions and the new conditions developed in the project.
5. Explore applications to economic allocation problems and political decision-making.

References

1. Palfrey, T. R., & Srivastava, S. (1987). *On Bayesian Implementable Allocations*. The Review of Economic Studies, 54(2), 193–208.
 2. Palfrey, T. R., & Srivastava, S. (1989). *Implementation with Incomplete Information in Exchange Economies*. Econometrica, 57(1), 115–134.
 3. Jackson, M. O. (1991). *Bayesian Implementation*. Econometrica, 59(2), 461–477.
-

3. Candidate Profile

Applicants must hold a Research Master’s degree in Mathematics, Mathematical Economics, Economics, or a closely related field. A solid background in mathematical methods, microeconomic theory, and game theory is expected. Prior knowledge of social choice theory, mechanism design, and implementation theory will be considered an advantage.

Candidates should demonstrate strong analytical skills and the ability to undertake rigorous theoretical research, including mathematical modelling, abstract reasoning, formal proof techniques, and the development of original theoretical results. Intellectual curiosity, motivation, and the ability to work independently on challenging research problems are essential.

Candidates must be prepared to devote themselves full-time to their doctoral research throughout the PhD programme. They should be committed to producing original, high-quality theoretical research, with the ambition of publishing their findings in leading international peer-reviewed journals in game theory, mathematical economics, and social choice theory, such as *Games and Economic Behavior*, *Mathematical Social Sciences*, *Journal of Mathematical Economics*, *International Journal of Game Theory*, *Social Choice and Welfare*, and *Theory and Decision*.

Fondements et architecture d'un système d'exploitation nativement agentique : de l'interaction par applications à l'exécution par intentions

Sujet proposé par : Imade BENELALLAM

Laboratoire : SI2M | **Formation doctorale :** Systèmes d'information, systèmes intelligents

Année universitaire : 2026-2027

Contexte et problématique

Les systèmes d'exploitation actuels organisent principalement l'activité informatique autour des applications, des processus, des fichiers et des commandes. L'émergence d'agents capables de planifier, de mobiliser des outils et de déléguer à des sous-agents ouvre la voie à un autre paradigme : l'utilisateur exprime une intention et le système compose les compétences, les connaissances et les ressources nécessaires à sa réalisation.

Les travaux AIOS, AgentVM, AOHP et Agate explorent différentes parties de cette évolution [1-4]. Ils ne constituent pas une architecture générale stabilisée remplaçant les systèmes existants. La définition des unités d'exécution, de la mémoire contextuelle, de la délégation et des interfaces système demeure un champ de recherche actif [5-6].

Cette thèse étudiera les fondements d'un système d'exploitation conçu nativement pour les agents. Le remplacement à long terme des environnements centrés sur les applications constitue l'horizon scientifique ; le résultat attendu du doctorat est une architecture originale et un prototype expérimental évalué.

Question de recherche

Quelles abstractions et quels mécanismes système permettent de transformer une intention utilisateur en une exécution coordonnée d'agents et de sous-agents, persistante et contrôlable, tout en préservant la cohérence de l'état lorsque les objectifs, les ressources ou les délégations évoluent ?

Objectif général

Concevoir, formaliser et expérimenter une architecture de système d'exploitation nativement agentique, dans laquelle les intentions, les agents persistants, les compétences composables et les objets de connaissance constituent des abstractions système explicites.

Axes de recherche

- 1. Abstractions et modèle d'exécution.** Définir l'intention, l'agent, la compétence et la délégation ; étudier leur composition, leur cycle de vie et la modification d'un objectif en cours d'exécution.
- 2. État et mémoire du système.** Articuler contexte de travail, mémoire persistante et objets de connaissance versionnés ; étudier le partage entre agents, la traçabilité et la reprise après interruption.
- 3. Services système et gestion des ressources.** Concevoir des interfaces structurées pour les outils et les communications ; répartir calcul, mémoire et budgets entre agents ; distinguer les politiques adaptatives des mécanismes d'exécution et d'isolation.

Démarche scientifique et framework expérimental

La recherche commencera par une analyse comparative des architectures existantes afin de préciser les limites des abstractions classiques et d'identifier une contribution originale au-delà de l'assemblage de services. L'hypothèse à examiner est qu'une représentation commune de l'intention, de la délégation et de l'état, exploitable par les services système, améliore la continuité et la cohérence des exécutions par rapport à leur gestion séparée dans chaque application.

Le framework comprendra un gestionnaire d'intentions et de tâches, un service de cycle de vie des agents, une mémoire contextuelle persistante, un registre de compétences, des communications structurées et un journal d'exécution. Les propriétés revendiquées seront explicitées et vérifiées dans le périmètre du prototype.

Une première réalisation sur un OS hôte permettra de tester les abstractions. Une seconde étape intégrera un sous-ensemble de mécanismes dans une machine virtuelle ou sur un micro-noyau, selon les résultats de l'étude initiale, pour examiner leur prise en charge au niveau système. L'architecture devra préciser ce qui relève du raisonnement agentique, du contrôle d'exécution et du support matériel.

Protocole d'évaluation

Les expérimentations porteront sur plusieurs familles de tâches : organisation de connaissances, production de documents et traitements de données. Elles incluront des délégations imbriquées, des modifications d'objectifs, des conflits sur des ressources partagées et des interruptions contrôlées. Les tâches de développement et d'évaluation seront séparées.

Le prototype sera comparé à un agent utilisant des applications et outils classiques, ainsi qu'à une orchestration applicative disposant des mêmes modèles et compétences. Les mesures porteront sur la réussite vérifiée, la cohérence de l'état, la reprise, les interventions humaines, la portabilité et le coût total. Des ablations isoleront l'apport des mécanismes proposés ; les résultats seront répétés et rapportés avec leurs conditions et leur incertitude.

Contributions et résultats attendus

Un modèle architectural et des interfaces documentées ; un ou plusieurs mécanismes système originaux ; un prototype reproductible ; un corpus de scénarios et un protocole d'évaluation. Les travaux préciseront les bénéfices, les limites et les conditions de transition vers un environnement informatique organisé autour des intentions et des agents.

Profil souhaité et mots-clés

Master ou diplôme d'ingénieur en informatique ou domaine proche. Compétences en systèmes d'exploitation, systèmes distribués, programmation système et architectures logicielles ; intérêt pour les agents intelligents et la validation expérimentale.

Mots-clés : systèmes d'exploitation agentiques ; calcul orienté intentions ; systèmes multi-agents ; délégation ; mémoire contextuelle ; abstractions système.

Références bibliographiques indicatives

[1] Ge et al. (2023). LLM as OS, Agents as Apps: Envisioning AIOS, Agents and the AIOS-Agent Ecosystem. Prépublication.

[2] Wu et al. (2026). Position: Digital Agents Require Unified Agent-Native Computers. ICML.

[3] Zhao et al. (2026). AOHP: An Open-Source OS-Level Agent Harness for Personalized, Efficient and Secure Interaction. Prépublication.

[4] Wei et Fang (2026). Agate: Capability Microkernels as a Natural Substrate for AI Agents. Workshop AgenticOS, associé à SOSP.

[5] Liu et al. (2026). An AgentOS Needs a Formal Representation of Agents. Workshop AgenticOS, associé à SOSP.

[6] Liu et al. (2026). The Abstractions Are Slipping: Four OS Mismatches in the Agent Era. Workshop AgenticOS, associé à SOSP.

Co-orchestration des environnements d'apprentissage agentiques : partage du contrôle entre enseignants, apprenants et agents pour un apprentissage autonome et responsable

Sujet proposé par : Imade BENELALLAM

Laboratoire : SI2M | **Formation doctorale :** Systèmes d'information, systèmes intelligents

Année universitaire : 2026-2027

Contexte et problématique

Les agents intelligents peuvent accompagner les apprenants, soutenir les enseignants et intervenir dans les activités collectives. Leur intégration transforme la répartition des décisions pédagogiques : choix des tâches, formes d'aide, rythme des interventions et appréciation des difficultés. Les projets CAIL, ClassAid et FACET explorent différentes modalités de cette coopération [1-3].

Les résultats disponibles dépendent toutefois des dispositifs et des contextes. Tutor CoPilot montre l'intérêt d'une assistance au tuteur humain [4], tandis que d'autres expériences soulignent les risques d'une délégation excessive de l'effort cognitif [5]. Une activité réussie avec l'IA ne garantit pas une compréhension durable ni une capacité à agir sans assistance.

Le défi consiste à organiser une coopération où l'enseignant conserve un jugement pédagogique effectif, l'apprenant exerce sa capacité de décision et le système intervient dans un périmètre explicite. Une supervision intégralement centralisée peut surcharger l'enseignant ; une autonomie non régulée peut favoriser la recherche de réponses au détriment de l'apprentissage.

Question de recherche

Comment concevoir et évaluer un mécanisme de partage et de révision du contrôle entre enseignants, apprenants et agents intelligents, capable d'adapter l'assistance aux objectifs pédagogiques tout en préservant l'autonomie des apprenants, la responsabilité humaine et la soutenabilité de la supervision ?

Objectif général

Développer un modèle de co-orchestration et un framework expérimental permettant de définir, déléguer, expliquer et réviser les décisions d'assistance pédagogique, puis mesurer leurs effets sur les apprentissages durables et la collaboration.

Axes de recherche

- 1. Modélisation des rôles et du contrôle partagé.** Définir les responsabilités, les décisions déléguables et les règles d'intervention ; permettre aux apprenants de demander une révision et aux enseignants de modifier les politiques.
- 2. Régulation de l'assistance.** Concevoir des politiques décidant quand proposer un indice, demander une justification, solliciter un pair, réduire l'aide ou transmettre la situation à l'enseignant.
- 3. Collaboration responsable et explicable.** Rendre les interventions et leurs motifs compréhensibles ; expliciter les incertitudes ; intégrer la contestation, la minimisation des données et l'analyse des différences de bénéfices entre apprenants [6].

Framework expérimental et contribution visée

Le framework réunira un espace enseignant pour définir les objectifs et les règles d'aide, un espace apprenant pour travailler et discuter les interventions, des services agentiques de diagnostic et d'assistance, et un moteur de co-orchestration. Ce dernier gèrera les niveaux de délégation, les demandes de révision et les situations nécessitant une intervention humaine.

Un journal limité aux événements utiles permettra de documenter les décisions, les indices fournis et les reprises de contrôle. Les fonctions de diagnostic et de tutorat pourront être séparées en agents spécialisés si leur utilité est démontrée ; la multiplication des agents ne sera pas une contribution en soi. L'enseignant conservera la responsabilité du jugement pédagogique.

L'originalité sera recherchée dans un mécanisme explicite et révisable de partage du contrôle, tenant compte des demandes des apprenants et du budget d'attention de l'enseignant. La première phase précisera ce positionnement face aux systèmes existants, notamment ClassAid, et formulera des hypothèses testables.

Méthodologie et protocole d'évaluation

Une démarche de conception participative associera enseignants et étudiants pour identifier les décisions à partager et les situations de désaccord. Un pilote sera réalisé dans un enseignement de programmation, de statistique ou d'analyse de données, avant une étude sur plusieurs semaines. L'effectif et le plan expérimental seront définis par une analyse de puissance et les contraintes du terrain.

Trois conditions seront comparées : assistance pilotée par l'enseignant, assistance automatique et co-orchestration proposée. Les contenus, les modèles et les ressources seront comparables. Les critères principaux porteront sur des épreuves indépendantes sans IA, la rétention différée et le transfert à de nouveaux problèmes ; les productions réalisées avec assistance seront évaluées séparément.

Les critères complémentaires couvriront la charge de supervision, les interventions humaines, la qualité de la collaboration, l'autonomie et la confiance calibrée. L'autonomie sera appréciée par plusieurs indicateurs : demandes justifiées, vérification des suggestions et capacité à corriger une erreur, et non par la seule fréquence d'utilisation. Des ablations isoleront l'apport du partage du contrôle et de la régulation de l'aide.

Contributions attendues et profil souhaité

Un modèle de co-orchestration, un mécanisme de délégation révisable, un prototype reproductible et une évaluation en situation d'enseignement. Profil : Master ou diplôme d'ingénieur en informatique ou domaine proche ; compétences en IA, systèmes multi-agents et développement logiciel ; intérêt pour l'interaction humain-machine, les méthodes expérimentales et les sciences de l'apprentissage.

Mots-clés : systèmes agentiques ; co-orchestration pédagogique ; contrôle partagé ; autonomie ; apprentissage durable ; collaboration humain-IA.

Références bibliographiques indicatives

- [1] MIT STEP Lab. Collaborative AI for Learning (CAIL). Présentation du projet, consultée en octobre 2026.
- [2] Zhang et al. (2026). ClassAid: A Real-time Instructor-AI-Student Orchestration System for Classroom Programming Activities.
- [3] Gonnermann-Müller et al. (2026). FACET: Multi-Agent AI Supporting Teachers in Scaling Differentiated Learning for Diverse Students. IJCAI.
- [4] Wang et al. Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise. Prépublication, version révisée en 2026.
- [5] Bastani et al. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. PNAS, 122(26).
- [6] UNESCO (2024). AI competency framework for teachers. Cadre de référence.