Proposition de sujet de thèse

« Vers des méthodes d'apprentissage hybride avec garanties de performance pour la supervision et la commande sûre des systèmes cyber-physiques critiques »

Laboratoire d'accueil : Ampère (UMR CNRS 5005), INSA Lyon – Université Claude Bernard Lyon 1

École doctorale : EEA – Électronique, Électrotechnique, Automatique – Université de Lyon

Date de démarrage souhaitée : 01/11/2026

Encadrement : Dr. Mohammed-Farouk Bouaziz & Pr. Eric Zamaï

Financement : Chaire de l'Agence Nationale de la Recherche (Allocation CPJ)

Modalités de candidature : Date limite de candidature au 18 octobre 2026

Contacts : CV-LM et relevés de notes à mohammed-farouk.bouaziz@insa-lyon.fr & eric.zamai@insa-lyon.fr

Lieu : Laboratoire Ampère, équipe AIS, Campus Lyon Tech - La Doua. Villeurbanne

Gratification : voir grille INSA Lyon (contrat doctoral de droit français)

Résumé : Cette thèse propose de développer un cadre méthodologique pour l'intégration sécurisée de l'Intelligence Artificielle dans la supervision des systèmes cyber-physiques critiques. Contrairement aux approches classiques, l'objectif est d'associer aux modèles d'apprentissage des garanties formelles de performance compatibles avec les exigences de l'automatique. Les travaux portent sur la conception de modèles hybrides capables d'estimer leur propre domaine de validité opérationnel. Un superviseur en ligne sera développé pour valider ou rejeter en temps réel les décisions de l'IA, assurant ainsi une reconfiguration sûre face aux défaillances et cyberattaques.

Mots-clés : Automatique, Diagnostic & Pronostic, Intelligence Artificielle, Performance, Modélisation Hybride, Supervision, Systèmes Cyber-Physiques.

Abstract : This thesis develops a methodological framework for the integration of Artificial Intelligence (AI) components into the supervision, diagnosis, and control architectures of critical Cyber-Physical Systems. Moving beyond standard explainable AI, the research aims to provide learning models with formal performance guarantees regarding stability and robustness, aligned with control theory requirements. The project focuses on designing hybrid models (physics-based and data-driven) capable of estimating their operational validity domains. An online supervisor will be implemented to monitor "IA-in-the-Loop" components, validating or rejecting decisions in real-time to ensure safe reconfiguration during failures or cyberattacks.

Keywords : Automatic Control, Diagnosis & Prognosis, Artificial Intelligence, Performance, Hybrid Modeling, Supervision, Cyber-Physical Systems.

Contexte scientifique :

L'évolution industrielle actuelle, portée par les enjeux de la transition énergétique et numérique (Industrie 4.0, Smart-Grids, mobilité électrique), conduit à une complexité croissante des systèmes cyber-physiques. Pour piloter ces systèmes, l'Intelligence Artificielle (IA) et l'apprentissage automatique offrent des perspectives considérables, notamment pour modéliser les phénomènes complexes, optimiser les performances et s'adapter à des environnements incertains. Cependant, les modèles d'apprentissage utilisés pour le diagnostic ou le pronostic des systèmes cyber-physiques fournissent des scores de confiance ou d'incertitude qui ne sont pas directement exploitables par une architecture de commande/supervision en boucle fermée. Or, dans un système critique, une erreur de diagnostic ou de pronostic peut conduire à une décision de maintenance ou de reconfiguration dégradant la sûreté, la disponibilité ou la performance du système. Ainsi, un fossé majeur subsiste et les approches issues de l'IA, souvent perçues comme des "boîtes noires", ne fournissent intrinsèquement aucune garantie sur des propriétés essentielles en automatique telles que la sûreté et la robustesse face aux perturbations. De plus, leur intégration dans des boucles de rétroaction temps réel doit respecter des contraintes fortes.
Pour répondre à cette problématique, ces travaux de thèse se positionnent au croisement de trois communautés (apprentissage fiable, automatique robuste et prognostic and health management) et se proposeront pour développer des méthodes et des outils fondés sur la théorie de l'automatique afin d'encadrer et de sécuriser l'usage de l'IA au sein des infrastructures critiques. En utilisant les concepts de l'automatique, de nouveaux composants d'apprentissage doivent être développés au sein des architectures de diagnostic, de pronostic et de commande pour rendre l'IA intégrable et sans perdre les propriétés attendues en automatique.

Verrous scientifiques :

En continuité avec les thématiques développées au sein du laboratoire Ampère, les travaux de cette thèse s'articuleront autour de trois axes scientifiques :

- Un premier axe de recherche sur l'apprentissage hybride qui vise à construire des modèles IA/hybrides capables de calculer non seulement un résultat de diagnostic/pronostic, mais aussi de dire dans quel domaine ce résultat est valide, avec quelle incertitude, et avec quel risque d'erreur dynamique. Cela peut mobiliser des modèles hybrides (physique/données), des observateurs, des approches probabilistes (réseaux de Petri, réseaux Bayésiens, réseaux de Neurones), ou des modèles à contraintes physiques. En reliant IA, automatique et sûreté, cet axe permet de définir le domaine de validité opérationnel d'une IA dans un système dynamique.
- Un second axe de certification des décisions de l'IA en développant un superviseur qui vérifie en ligne si la sortie de l'IA est admissible. Par exemple : un diagnostic d'attaque, une prédiction de panne ou une solution de commande ne serait acceptée que si elle respecte des contraintes de cohérence dynamique, de robustesse ou de sûreté. Ce superviseur peut s'appuyer sur des invariants, des contraintes physiques, des modèles hybrides, des observateurs, des règles de sécurité, des tests de cohérence ou des garanties probabilistes afin de valider, corriger ou rejeter une décision.
- Un troisième axe de reconfiguration sûre vise à valider les méthodes développées sur des scénarios de défaillance naturelle et d'attaque cyber-physique, afin d'évaluer la capacité du système à maintenir un niveau minimal de performance garanti en situation dégradée ou attaquée.

Ainsi, et à la charnière entre l'automatique et l'IA, cette thèse adressera plusieurs verrous scientifiques novateurs et ambitieux :

- Propriétés de l'IA-in-the-Loop : Comment passer de métriques classiques de l'IA à des garanties mathématiques pertinentes pour l'automatique des systèmes dynamiques (garanties statistiques de validité des sorties IA et garanties de sûreté face aux contraintes dynamiques) ?
- Hybridation et complexité : Comment construire des modèles hybrides qui exploitent les données par apprentissage sans violer les contraintes physiques, énergétiques ou dynamiques du système ?
- Supervision en ligne des décisions IA : Comment concevoir une couche de supervision capable de transformer une sortie issue d'un modèle IA non garantie en une décision admissible pour un système dynamique contraint ?
- Maintien de performance : Comment garantir un niveau minimal de performance lorsque le système sort du domaine nominal, par exemple en cas de dérive capteur, de défaillance actionneur, de perturbation ou d'attaque cyber-physique ?

Objectifs de la thèse :

L'objectif central de cette thèse consiste à convertir des informations statistiques (scores calculés par les modèles d'apprentissage) en contraintes dynamiques vérifiables en ligne, afin de garantir qu'une décision issue de l'IA ne dégrade pas la sûreté ou la performance du système. Un cadre méthodologique doit être proposé afin d'assurer une intégration maîtrisée de composants d'IA dans les architectures de supervision, de diagnostic et de commande des systèmes cyber-physiques. Contrairement aux approches classiques d'IA explicable, l'objectif n'est pas uniquement de rendre les décisions interprétables a posteriori, mais de leur associer des garanties de validité, de robustesse et de sûreté compatibles avec les exigences de l'automatique. Les travaux porteront sur la conception de modèles hybrides combinant connaissances physiques et apprentissage des données, capables d'estimer leur domaine de validité et leur niveau de confiance opérationnel.

En s'appuyant sur ces approches hybrides, une couche de supervision sera ensuite développée afin de vérifier en ligne les décisions produites par l'IA, de détecter les situations hors limites, et de déclencher des mécanismes de reconfiguration sûre. La contribution attendue est un cadre d'IA supervisée par l'automatique, applicable au diagnostic, au pronostic, à la maintenance prédictive et à la cyber-résilience de systèmes critiques, avec validation sur des scénarios réalistes de défaillances, de perturbations et d'attaques cyber-physiques. Ceci afin d'attester de la résilience opérationnelle en temps réel.

Déroulement de la thèse :

Le programme de travail de la thèse combine des volets méthodologiques (modélisation hybride), technologiques (IA certifiable) et expérimentaux (simulation et essais plateforme). Il s'articule en trois étapes majeures déclinées en quatre phases :

Année 1 : Etat de l'art et spécification	M1 – M6 ⇨ Phase 1 : • Analyse bibliographique approfondie au croisement de l'IA, de l'automatique et des systèmes cyber-physiques, et état de l'art sur les métriques de confiance et les garanties de performance pour l'IA. • Identification des systèmes industriels cibles et définition des scénarios critiques (pannes physiques et cyberattaques). M6 – M12 ⇨ Phase 2 : • Spécification de l'interface entre les modèles physiques (équations différentielles, LMI) et les algorithmes d'IA (boîtes noires). • Mise en place de l'architecture de collecte et de prétraitement des données pour l'apprentissage hybride.
Année 2 : Développement méthodologique et algorithmique	M12 – M18 ⇨ Phase 2 - suite : • Développement de la méthodologie pour la fusion modèles & données. • Conception du superviseur en ligne pour valider les sorties de l'IA sous contraintes. • Mise en œuvre des mécanismes de confiance pour les composants « IA-in-the-Loop ». M18 – M24 ⇨ Phase 3 : • Développement d'algorithmes d'apprentissage capables de fonctionner sous contraintes (peu de données, ressources limitées). • Première validation par simulation numérique sur des cas d'étude (ex: infrastructures énergétiques) face à des scénarios de défauts ou de cyber-malveillance.
Année 3 : Intégration, validation expérimentale et valorisation	M24 – M30 ⇨ Phase 3 – suite : • Consolidation des indicateurs de confiance pour garantir la robustesse et la sûreté du diagnostic/pronostic. • Validation sur les plateformes technologiques pour évaluer la résilience opérationnelle face à des perturbations en temps réel M30 – M36 ⇨ Phase 4 : • Évaluation comparative des performances de la solution proposée par rapport aux méthodes classiques. • Valorisation des travaux à travers la rédaction de publications scientifiques dans des revues internationales et la rédaction finale du manuscrit de thèse.

Profil et compétences recherchés :

Le ou la candidat(e) devra être titulaire d'un diplôme d'ingénieur ou d'un Master 2 (ou autre diplôme équivalent) en automatique, génie électrique, informatique industrielle, cybersécurité, data-science, intelligence artificielle ou systèmes cyber-physiques. Des compétences en modélisation, systèmes à événements discrets, Matlab et apprentissage automatique à base de données seront appréciées. Une grande rigueur scientifique, un goût pour le travail multidisciplinaire et une aptitude pour l'expérimentation sur des plateformes technologiques sont attendus.

Rémunération et avantages :

Cette thèse est financée pour une durée de 36 mois dans le cadre d'un contrat doctoral de droit français (Allocation CPJ – ANR – INSA Lyon). Le salaire brut mensuel sera autour de 2 300 € brut par mois pour un contrat de recherche à temps plein. Des activités complémentaires d'enseignement ou de valorisation peuvent être envisagées selon les règles de l'établissement et donner lieu à une rémunération supplémentaire. Le contrat donne accès au système de sécurité sociale français, qui couvre une grande partie des frais médicaux et hospitaliers. Une complémentaire santé peut également être proposée ou recommandée.

Laboratoire et équipe d'accueil :

Le laboratoire Ampère est une unité mixte de recherche (UMR 5005) du CNRS, de l'École Centrale de Lyon, de l'INSA de Lyon, et de l'Université Claude Bernard Lyon 1. Il regroupe 206 personnes, dont 83 doctorants et 75 chercheurs et enseignants-chercheurs. Son activité de recherche repose sur la gestion et l'utilisation rationnelle de l'énergie dans les systèmes en relation avec leur environnement et développe pour cela des activités disciplinaires et transdisciplinaires portées par quatre équipes : Automatique pour l'ingénierie des systèmes (AIS), Conversion de puissance, Stockage de l'énergie et Machines (CoSMa), Matériaux pour le Génie Electrique et ses Composants (MaGEC) et Ingénierie des Systèmes pour la santé et l'Environnement (IS2E). Les travaux de recherche conduits au sein de l'équipe AIS concernent le développement de méthodologies et d'outils visant l'optimisation et la maîtrise du comportement dynamique des systèmes et ce dans de très nombreux domaines d'applications, en collaboration avec les autres équipes du Laboratoire et d'autres laboratoires en Sciences de l'Ingénieur.

Cette thèse s'aligne avec la stratégie du laboratoire Ampère et de son équipe AIS qui vise à explorer les possibilités offertes par l'automatique pour la conception de systèmes complexes, en réponse à des enjeux sociétaux majeurs comme l'énergie propre et l'industrie du futur. Cette thèse s'inscrit au cœur de la thématique RICS (Resilient Industrial Cyber Systems) d'AIS. L'équipe RICS est composée de 15 personnes et a comme axe principal de recherche, la Sûreté et la Sécurité des Systèmes Cyber-Physiques, avec deux volets majeurs :

- Conception de lois de commande : Développer des commandes robustes capables de maîtriser le système physique tout en résistant aux perturbations.
- Gestion des aléas : Concevoir des architectures pour la détection, le diagnostic, le pronostic et la reconfiguration des systèmes face à des défaillances naturelles ou des attaques malveillantes.

Bibliographie :

- Wabersich, K. P., & Zeilinger, M. N. (2021). A predictive safety filter for learning-based control of constrained nonlinear dynamical systems. Automatica. Vol. 129.
- Hewing, L., et al. (2020). Learning-Based Model Predictive Control: Toward Safe Learning in Control. Annual Review of Control, Robotics, and Autonomous Systems. Vol. 3:269-296.
- Dawson, Ch., et al. (2023). Safe Control with Learned Certificates: survey Neural Lyapunov, Barrier and contraction Methods for Robotics and Control. IEEE Trans on Robotics. Vol. 39:3:1749–1767.
- Karniadakis, G., et al. (2021). Physics-informed machine learning. Nature Reviews Physics, Vol. 3:422-440.
- Nghiem, T. X., et al. (2023). Physics-Informed Machine Learning for Modeling and Control of Dynamical Systems. American Control Conference (ACC). San Diego, CA, USA, 2023, pp. 3735-3750.
- Kasilingam, S., et al. (2024). Physics-based and data-driven hybrid modeling in manufacturing: A review. Production & Manufacturing Research. Vol. 12:1:2305358.
- Angelopoulos, A., Bates, S. (2023). Conformal Prediction: A Gentle Introduction. Foundations and Trends in Machine Learning. Foundations and Trends in Machine Learning, Vol. 16:4:494–59.
- Yuan, S., et al. (2026). Conformal machine learning for reliable anomaly detection in industrial cyber-physical systems. Reliability Engineering & System Safety. Vol. 274:112417.
- Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence: Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. Vol. 58:82–115.
- Xu, Z., Saleh, J. H. (2021). Machine learning for reliability engineering and safety applications. Reliability Engineering & System Safety. Vol. 211:107530.
- Wen, Y., et al. (2022). Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective. Measurement. Vol. 187:110276.
- Abdellaoui, S., et al. (2026). Monitoring Cyber-threats in Railway Systems: A Hybrid Framework for Detecting Stealthy Data Tampering Attacks. Reliability Engineering & System Safety. Vol. 266:111747.
- Bhamare, D., et al. (2020). Cybersecurity for industrial control systems: A survey. Computers & Security. Vol. 89:101677.
- Duo, W., et al. (2022). A Survey of Cyber Attacks on Cyber Physical Systems: Recent Advances and Challenges. IEEE/CAA Journal of Automatica Sinica. Vol. 9:5:784–800.