OFFRE D'ENCADREMENT DE THÈSE

THESIS SUPERVISION OFFER

Etablissement **Télécom Paris**

École doctorale **Ecole Doctorale de l'Institut Polytechnique de Paris**

Spécialité **Informatique, données, IA**

Domaine Scientifique **Mathematiques et leurs interactions**

Unité de recherche **Laboratoire de Traitement et Communication de l'Information**

Encadrement de la thèse **Vadim MALVONE**

Début de la thèse le **1 octobre 2026**

Date limite de candidature (à 23h59) **31 mai 2026**

Mots clés - Keywords

IA agentive, Vérification formelle, Raisonnement stratégique, Systèmes multi-agents

Agentic AI, Formal Verification, Strategic Reasoning, Multi-Agent Systems

Description de la problématique de recherche - Project description

L’Intelligence Artificielle Agentive (Agentic AI) est une direction émergente de l’IA contemporaine. Ces systèmes combinent raisonnement basé sur les LLM, planification, mémoire, utilisation d’outils et prises de décision autonomes, agissant comme des agents capables de résoudre des problèmes dans des environnements complexes.

Malgré leurs capacités, ces systèmes manquent de fondations formelles : pipelines faiblement organisés, comportement stratégique imprévisible et absence de garanties de sécurité.

Cette thèse vise à établir les fondations de la vérification formelle des systèmes Agentic AI, en adaptant le model checking à ce nouveau paradigme. L’objectif est de modéliser rigoureusement les agents pilotés par LLM, surveiller leurs comportements stratégiques et étudier la vérification dans des contextes dynamiques et incertains, fournissant ainsi les premiers outils formels pour ces environnements.

Agentic Artificial Intelligence (Agentic AI) is an emerging direction in contemporary AI. These systems combine LLM-based reasoning, planning, memory, tool use, and autonomous decision-making, enabling them to act as agents capable of solving problems in complex environments.

Despite their capabilities, these systems lack formal foundations: they are typically loosely organized pipelines with unpredictable strategic behavior and no safety guarantees.

This PhD aims to establish the foundations of formal verification for Agentic AI systems by adapting model checking to this new paradigm. The goal is to rigorously model LLM-driven agents, monitor their strategic behaviors, and study verification in dynamic and uncertain contexts, providing the first formal tools for these environments.

Thématique / Domaine / Contexte

The PhD focuses on the formal verification of Agentic AI systems, an emerging area of contemporary AI. These systems, based on LLMs and capable of autonomous planning, tool use, and strategic reasoning, currently lack formal foundations and safety guarantees. The research aims to establish rigorous methods to verify the strategic behavior and reliability of these autonomous agents, by adapting and extending formal verification techniques to hybrid symbolic–subsymbolic architectures.

Formal Methods, Verification, and Artificial Intelligence

Classical formal verification techniques, such as model checking [1] and runtime verification (RV) [3], have been extensively applied to software, hardware, and multi-agent systems to ensure compliance with temporal logic specifications [1] and communication protocols [10, 11]. Model checking provides a rigorous framework to determine whether a system satisfies formal specifications and has found numerous industrial applications [2]. However, due to computational complexity and the state-space explosion problem, it struggles to handle highly dynamic, hybrid agent architectures such as those based on large language models (LLMs).

Runtime verification offers a lighter, execution-based approach, monitoring system properties during runtime rather than exploring the entire model [3]. In multi-agent systems, RV has mainly been applied to verifying communication protocols [10, 11] and ensuring that agents comply with predefined interaction rules. Despite these advances, both model checking and RV face limitations when applied to Agentic AI systems, which combine LLM-based reasoning, autonomous planning, tool use, memory, and strategic decision-making in complex environments.

Strategic reasoning is central in multi-agent systems, and temporal logics such as ATL [4] and Strategy Logic [5] have been developed to capture agents’ abilities to enforce specific outcomes. However, strategic verification becomes extremely challenging under imperfect information [6, 7, 8]. Even recent approximation techniques [9] do not fully address the unique challenges posed by LLM-driven autonomous agents.

Currently, no verification framework exists for strategic Agentic AI systems—agents capable of planning, interacting with tools, reasoning under uncertainty, and coordinating with other agentic components. While integrations of model checking and runtime verification have been explored in autonomous cognitive systems [12], none specifically target the emerging field of Agentic AI.

This PhD aims to address this scientific gap by establishing the theoretical and practical foundations for the formal verification of Agentic AI systems, adapting and extending both model checking [1] and runtime verification [3] techniques to rigorously model, verify, and ensure the reliability and strategic behavior of LLM-driven autonomous agents.

Objectifs

The aim of this research is to provide the first formal verification foundations for Agentic AI systems by integrating formal verification techniques with LLM-based reasoning and agentic decision pipelines.

The challenges of this PhD will involve tackling (some of) the following points:

- Studying model checking in monolithic and multi-agent systems to understand which techniques can be adapted to Agentic AI behaviour [1, 4, 5];
- Analysing the internal structure of LLM-driven agents (planning loops, reflections, tool calls, memory) and modelling them rigorously;
- Studying runtime verification to monitor agentic behaviours during execution, including hallucination detection and plan correction [3, 10, 11];
- Developing integrations of model checking and runtime verification tailored to hybrid symbolic–subsymbolic systems [12];
- Exploring decidability and tractability boundaries for verifying strategic behaviour under imperfect information [6, 7, 8, 9];
- Enhancing runtime verification with predictive abilities using design-time verification and runtime monitoring;
- Using runtime monitors to synthesise or refine strategies dynamically in response to unpredictable LLM decisions;
- Extending runtime verification to support strategic reasoning and branching-time properties under stochastic behaviours;
- Evaluating the proposed methods on state-of-the-art Agentic AI frameworks, contributing the first verification tools for practical agentic environments.

Méthode

The research will combine formal modeling of LLM-driven agents, model checking, and runtime verification techniques. It will involve analyzing agent internal structures (planning loops, tool use, memory), designing verification frameworks for hybrid symbolic–subsymbolic architectures, and studying strategic reasoning under uncertainty. The approach integrates theoretical development with practical evaluation on state-of-the-art Agentic AI systems.

Résultats attendus - Expected results

Expected Results:
- Formal models of Agentic AI systems, capturing planning, tool use, memory, and strategic reasoning of LLM-driven agents.
- Verification frameworks and techniques specifically designed for checking the correctness, reliability, and strategic behavior of Agentic AI agents.
- Insights into decidability and tractability of verifying autonomous, stochastic, and multi-agent LLM-based systems.
- Prototypes and software tools enabling practical formal verification of state-of-the-art Agentic AI frameworks.
- Scientific publications and dissemination of methods, contributing the first rigorous foundations for safe and reliable Agentic AI deployment.

Références bibliographiques

[1] E. M. Clarke, O. Grumberg, and D. A. Peled. Model Checking. MIT Press, 1999.
[2] G. Memmi. Integrated Circuits Analysis, System and Method using Model Checking. US Patent 7493247, 2009.
[3] M. Leucker, C. Schallhart. A brief account of runtime verification. J. Log. Algebraic Methods Program., 2009.
[4] R. Alur, T. A. Henzinger, O. Kupferman. Alternating-Time Temporal Logic. JACM, 2002.
[5] F. Mogavero, A. Murano, G. Perelli, M. Y. Vardi. Reasoning About Strategies. TOCL, 2014.
[6] F. Belardinelli, A. Lomuscio, V. Malvone. Abstraction-based verification under imperfect information. AAAI, 2019.
[7] F. Belardinelli, V. Malvone. Three-valued strategic abilities. KR, 2020.
[8] F. Belardinelli, A. Lomuscio, V. Malvone. Approximating perfect recall. KR, 2018.
[9] A. Ferrando, V. Malvone. Verification of Strategic Properties. AAMAS 2023.
[10] D. Ancona, A. Ferrando, V. Mascardi. Parametric runtime verification. AAMAS 2017.
[11] A. Ferrando, D. Ancona, V. Mascardi. Decentralizing MAS monitoring. AAMAS 2017.
[12] A. Ferrando, L. A. Dennis, R. C. Cardoso, M. Fisher, D. Ancona, V. Mascardi. Holistic V&V of Autonomous Systems. ACM TOSEM, 2021.

Contexte du poste : Modalités d'encadrement, de suivi de la formation et d'avancement des recherches du doctorant - Details on the thesis supervision

The doctoral student’s work will be closely supervised, with weekly meetings to discuss progress, address challenges, and plan the next steps of the research. In addition to regular supervision, the student will receive guidance on methodological approaches, literature review, and technical skills, ensuring continuous training and support throughout the PhD. Progress will be tracked systematically, with milestones and deliverables agreed upon to monitor advancement and maintain alignment with the research objectives.

Conditions scientifiques matérielles et financières du projet de recherche

The PhD will be conducted under a doctoral contract with Télécom Paris. The student will have access to the university’s research facilities, computing resources, and software tools required for the project. Regular supervision and guidance will be provided, and all work will follow the institutional safety and ethical guidelines applicable to research involving software systems and AI agents.

Objectifs de valorisation des travaux de recherche du doctorant : diffusion, publication et confidentialité, droit à la propriété intellectuelle,...

The objectives of valorization of the doctoral research are: (i) to disseminate results through publications in leading international conferences and journals in AI, formal verification, and formal methods; (ii) to develop and release open-source software tools for formal verification, fostering community adoption and reproducibility; (iii) to participate in workshops, seminars, and collaborations to maximize the visibility and impact of the research; (iv) to ensure proper management of intellectual property rights, respecting both open-source licenses

for software and publication agreements for research outputs; (v) to contribute to knowledge transfer activities, such as tutorials or educational materials, to support the broader scientific and industrial community.

Collaborations envisagées

Cooperation with Angelo Ferrando, an expert in Runtime Verification techniques and a lecturer at the University of Modena and Reggio Emilia, is planned.

Profil et compétences recherchées - Profile and skills required

Compétences et Qualifications Attendues :
- Master en informatique, mathématiques, intelligence artificielle ou dans un domaine connexe.
- Solide formation en informatique et/ou en mathématiques, avec un accent sur les méthodes formelles, la logique ou l'informatique théorique.
- Bonnes compétences en programmation et expérience en développement logiciel.
- Maîtrise de l'anglais écrit et oral.
- Curiosité, motivation et capacité à explorer de nouvelles directions de recherche, en particulier à l'intersection de l'IA symbolique et subsymbolique.

Compétences Supplémentaires Souhaitables :
- Intérêt pour l'IA autonome, les systèmes multi-agents ou les modèles de langage de grande taille (LLM).
- Connaissances de base en model checking, vérification à l'exécution ou logiques temporelles et stratégiques.
- Ouverture aux approches interdisciplinaires et à la recherche collaborative.

Expected Competencies and Qualifications:
- Master's degree in Computer Science, Mathematics, Artificial Intelligence, or a related field.
- Strong foundation in computer science and/or mathematics, with emphasis on formal methods, logic, or theoretical computer science.
- Solid programming skills and experience with software development.
- Proficiency in written and spoken English.
- Curiosity, motivation, and the ability to explore new research directions, particularly at the intersection of symbolic AI.

Additional Desirable Skills:
- Interest in autonomous AI, agent systems, or large language models.
- Basic knowledge of model checking, runtime verification, or temporal and strategic logics.
- Openness to interdisciplinary approaches and collaborative research.

Dernière mise à jour le 26 novembre 2025