Titre de la thèse / Thesis title : IA de Confiance pour l'Aide à la Décision Clinique : Agents Augmentés par Graphes de Connaissances et Contrats Sémantiques / Trustworthy AI for Clinical Decision Support : Knowledge Graph-Augmented Agents with Semantic Contracts

Laboratoire d'accueil / Host Laboratory : Institut FEMTO-ST / FEMTO-ST Institute

Spécialité du doctorat préparé / Speciality : Informatique / Computer science

Mots-clefs / Keywords : Trustworthy AI; Medical Knowledge Graph; Semantic Contract; LLM

Descriptif détaillé de la thèse / Job description

Context and Problem Statement

The rapid integration of Large Language Models (LLMs) into clinical workflows presents a fundamental tension: while generative agents offer unprecedented capabilities for processing unstructured medical narratives, their stochastic nature conflicts with the deterministic safety requirements of medical practice. Within the European and French regulatory landscape—defined by the EU AI Act, GDPR, and the HDS (*Hébergeur de Données de Santé*) certification—deploying clinical AI is simultaneously a performance challenge and a sovereignty mandate. Healthcare institutions require systems that operate on-premise while providing rigorous, auditable decision support.

Current approaches to clinical AI suffer from three critical shortcomings:

1. **Stochastic hallucination.** LLMs connected directly to Electronic Health Record (EHR) databases may fabricate treatment histories, invent contraindications, or generate clinically plausible but factually incorrect information [1].
2. **Protocol deviation.** Unconstrained agents frequently prioritize statistically likely outputs over formal medical guidelines (HAS, ANSM), producing recommendations that are linguistically fluent but clinically inappropriate [2].
3. **Lack of traceability.** Direct database-to-LLM pipelines provide no symbolic reasoning trail, making it impossible for clinicians to audit the logic behind a recommendation—a requirement under the EU AI Act for high-risk systems [3].

These risks are not theoretical: prescribing errors and adverse drug events affect up to 7% of hospitalized patients [4]. The central scientific position of this thesis is that the path forward is not to "make the LLM smarter", but to **formally constrain it** through a symbolic middleware that structurally guarantees patient safety.

Proposed Research: A Neuro-Symbolic Architecture

This thesis proposes a neuro-symbolic middleware that prevents direct LLM interpretation of raw EHR or medical data. In this design, LLM agents interact exclusively through a **Clinical Semantic API** backed by a formal **Medical Knowledge Graph** and governed by **Semantic Contracts**.

A critical design principle is that this architecture is **task-agnostic**. The same ontology layer and contract catalogue governs data retrieval, discharge report generation, automated clinical coding (CIM-10/CCAM), drug interaction checking, and care pathway compliance auditing. The contracts define *what the agent is allowed to know and do*—not what kind of task it performs. By shifting clinical reasoning into a deterministic knowledge graph, the architecture also addresses the capability gap of smaller, sovereign, on-premise models [5, 11].

The thesis will explore two interrelated research directions. Their precise scope, prioritization, and methodology will be refined during the initial literature review phase, in collaboration with the candidate. The directions below represent a roadmap, not a rigid prescription.

Clinical Knowledge Graph construction. How to automatically build and continuously enrich a medical Knowledge Graph from heterogeneous French clinical data—structured codes (CIM-10, CCAM, ATC), semi-structured EHRs, and unstructured physician notes? The research will investigate how LLMs can support automated information extraction and how the resulting knowledge can be structured into standardized, interoperable representations compatible with established healthcare data models (FHIR, OMOP, SNOMED-CT, LOINC) [6, 7, 10]. Ensuring quality, consistency, and

clinical validity of the constructed KG is a central open question.

Constrained LLM agents via Semantic Contracts. How to design and evaluate LLM agents whose operational boundaries are strictly governed by formal contracts? The thesis will explore formalisms such as **Agent Contracts** [8]—tuples that unify input/output specifications, schema constraints, clinical rules, traceability requirements, and regulatory constraints (RGPD)—and evaluate them across multiple clinical task types (retrieval, report generation, clinical coding, drug interaction checking) to assess the universality of the contract-driven approach. Key evaluation dimensions include hallucination rate reduction, clinical accuracy, traceability of reasoning, and response latency compared to unconstrained baselines.

Validation Strategy:

Evaluation will go beyond standard NLP metrics (BLEU, ROUGE) to focus on clinical reliability and safety. The candidate will define clinically meaningful evaluation protocols, including assessments of clinical consistency against established guidelines, rates of contract violations, completeness of traceability, and comparisons with unconstrained LLM baselines in realistic clinical scenarios. A human-in-the-loop evaluation will assess the system's utility as an augmentative decision-support tool. The final system should preserve the clinician's full decision-making authority; the AI provides a traceable analytical layer rather than autonomous prescriptions.

Références bibliographiques / Bibliography

[1] Z. Ji, N. Lee, R. Frieske, et al., "Survey of hallucination in natural language generation," ACM Computing Surveys, vol. 55, no. 12, pp. 1–38, 2023.

[2] H. Nori, N. King, S. M. McKinney, D. Carignan, and E. Horvitz, "Capabilities of GPT-4 on medical challenge problems," arXiv preprint arXiv:2303.13375, 2023.

[3] European Parliament and Council, "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act)," Official Journal of the European Union, 2024.

[4] D. W. Bates, D. M. Levine, H. Salmasian, et al., "The safety of inpatient health care," New England Journal of Medicine, vol. 388, no. 2, pp. 142–153, 2023.

[5] L. Music, J.-P. Lorré, et al., "OpenLLM-France: Building and training open-source French large language models," Proceedings of LREC-COLING, 2024.

[6] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," Scientific Data, vol. 10, no. 67, 2023.

[7] G. Xiao, E. Pfaff, E. Prud'hommeaux, D. Booth, et al., "FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common Data Model," Journal of Biomedical Informatics, vol. 134, 2022.

[8] Q. Ye and J. Tan, "Agent Contracts: A formal framework for resource-bounded autonomous AI systems," arXiv preprint arXiv:2601.08815, 2026.

[9] L. Miculicich, M. Parmar, H. Palangi, K. Dvijotham, et al., "VeriGuard: Enhancing LLM agent safety via verified code generation," arXiv preprint arXiv:2510.05156, 2025.

[10] L. Murali, G. Gopakumar, D. M. Viswanathan, and P. Nedungadi, "Towards EHR-based medical knowledge graph construction, completion, and applications: A literature study," Journal of Biomedical Informatics, vol. 143, 2023.

[11] X. Zhao, S. Liu, S.-Y. Yang, and C. Miao, "MedRAG: Enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot," arXiv preprint arXiv:2502.04413, 2025.

Profil demandé / Applicant profile

We are looking for a candidate who holds a Master's degree in Computer Science or a related field, with demonstrated experience in at least two of the following areas: Natural Language Processing, Knowledge Representation and Reasoning (Knowledge Graphs, ontologies), Software Architecture, and Machine Learning. Familiarity with healthcare data standards (FHIR, CIM-10, SNOMED-CT) and regulatory frameworks (GDPR, EU AI Act) is a plus but not required. Strong programming skills (Python) and the ability to engage with both formal specification and verification techniques (design-by-contract, runtime monitoring, constraint satisfaction) and empirical evaluation are essential.

Preferred selection criteria:
- Applicants should have experience with AI – neural networks, more particularly LLMs / deep networks, and computer programming frameworks for deep learning using Python.

- Reasonable proficiency in English (written and spoken) is a requirement.

Personal characteristics:
- Interpersonal skills
- Dynamism and rigor
- Teamwork abilities

Financement : MESRI Etablissement

Dossier à envoyer pour le **22 mai 2026**
Début du contrat : 1er Octobre 2026
Salaire mensuel brut : 2300€ brut

Direction de la thèse:/ **Thesis Supervisor**
SALOMON Michel – michel.salomon@umlp.fr

Encadrement de la thèse : **co-directeur(s) et co-encadrant(s)**
AZAR Joseph (co-encadrant) – joseph.azar@umlp.fr

Applicants are invited to submit their application to the PhD supervisors.
Application must contain the following documents:
- CV
- Cover letter
- At least 1 reference letter
- Copy of Master degree if already available
- Copy of final marks and ranks