



Etablissement Sorbonne Université SIS (Sciences, Ingénierie, Santé)

École doctorale Ecole Doctorale d'Informatique, Télécommunications et Electronique

Spécialité Sciences et technologies de l'information et de la communication

Domaine Scientifique Sciences et technologies de l'information et de la communication

Unité de recherche Institut EURECOM

Encadrement de la thèse Paolo PAPOTTI

Financement Employeur AILY Labs



Début de la thèse le 2 février 2026

Date limite de candidature (à 23h59) 1 février 2026

Mots clés - Keywords

LLM augmentés, Raisonnement sur données d'entreprise

Tool LLMs, Enterprise data reasoning

Description de la problématique de recherche - Project description

Thèse sur des LLM augmentés par outils (SQL/SPARQL/Python) pour raisonner sur des données d'entreprise de façon fiable et auditable. Objectif : optimiser précision/coût/latence et intégrer incertitude, clarification et humain-dans-boucle sur cas d'usage réels.

Thèse sur des LLM augmentés par outils (SQL/SPARQL/Python) pour raisonner sur des données d'entreprise de façon fiable et auditable.

Objectif : optimiser précision/coût/latence et intégrer incertitude, clarification et humain-dans-boucle sur cas d'usage réels.

Thématische / Contexte

Industrial PhD at Aily Labs (Barcelona/Madrid) with EURECOM enrollment under Prof. Paolo Papotti's supervision.

Références bibliographiques

1. Sánchez Pérez, A., Boukhary, A., Papotti, P., Castejón Lozano, L., & Elwood, A. (2025). An LLM-Based Approach for Insight Generation in Data Analysis. Proceedings of NAACL-HLT 2025
2. Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language Models Can Teach Themselves to Use Tools. NeurIPS 2023.

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

AUTRE co-encadrant : Alaa Boukhary, Senior Data Scientist, Sté AILY Labs

Industrial contract.

Weekly meetings, stays at EURECOM and annual review meetings

Conditions scientifiques matérielles et financières du projet de recherche

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

Research will be disseminated through publications and presentations, subject to prior company review to protect confidentiality. Industrial property rights (IP) and any patentable results remain with the company.

Profil et compétences recherchées - Profile and skills required

Master en informatique, solide base en ML/NLP et bases de données (SQL/DBMS, graphes). Compétences : très bon Python/expérimentation, intérêt pour fine-tuning de LLM, text-to-SQL et usage d'outils/agents (RL en plus). Rigueur, autonomie et aisance en environnement industriel avec équipes produit/data.

Profile: Master's in Computer Science with a solid background in ML/NLP and databases (SQL/DBMS, knowledge graphs). Skills: strong Python/experimentation, interest in LLM fine-tuning, text-to-SQL and tool/agent use (RL is a plus). Rigorous, autonomous, and comfortable working with product/data teams in an industrial setting.

Dernière mise à jour le 9 janvier 2026