

Pierre Lepagnol, PhD.



AI Scientist & Engineer

pierre@lepagnol.net | +33 6 72 18 80 01 | Paris, France

[linkedin.com/in/pierre-lepagnol](https://www.linkedin.com/in/pierre-lepagnol) | github.com/PierreLepagnol | [hf.co/PierreLepagnol](https://huggingface.co/PierreLepagnol)

<https://pierre.lepagnol.net>

Skills

Programming Languages Python, C, English with OpenCode

Product AI & Evaluation SFT, RLVR, RLHF, LLM Evaluation, Information Retrieval

ML & LLM Systems PyTorch, Hugging Face, vLLM, LangChain, Unsloth

Optimization & Reliability Small Language Models, Robustness Analysis, Few-Shot Benchmarking

Infra & Data Docker, Linux, Git, SLURM, PySpark, SQL

Languages French (Native), English (Professional), German (Basic)

Experience

PhD Student, LISN Lab

2022/02-present

- Designed and built LLM evaluation pipelines across classical NLP tasks (classification, NER, sentiment analysis).
- Studied model reliability through large-scale systematic experiments with a focus on reproducible evaluation.
- Worked on Jean-Zay Supercomputer (1,456 x H100, 416 x A100, 1,832 x V100)
- Published four international papers across LREC-COLING 2024, Interspeech 2025, LREC 2026.

AI Research Consultant, SCIAM

2022/02-present

- Fine-tuned small language models for NER and analyzed trade-offs between capability, cost, and deployment constraints.
- Architected RAG systems and optimized LLM inference using vLLM and Hugging Face APIs.
- Migrated data analysis and data science projects from SAS to Python.
- Utilized advanced Pandas, PostgreSQL, and Object-Oriented Programming (OOP).

AI Research Engineer - Cybersecurity, French Ministry of Army - Direction Générale de l'Armement

2021/04-2021/09

- Designed and trained a 26M-parameter ResNet50 for side-channel attacks, reaching 85 percent accuracy on held-out data.
- Studied adversarial robustness and hardware deployment constraints for embedded systems in critical environments.
- Deployed optimized inference on Nvidia Jetson Nano under tight latency and memory limits.

Data Scientist, Société Générale

2017/02-2017/09

- Developed customer scoring models using Python, PySpark, and Scikit-learn.
- Performed text clustering using TF-IDF, K-means, embeddings, UMAP, and HDBSCAN.
- Provided internal Data Science training on imbalanced classification, Deep Learning, Autoencoders, and Tabular Data Augmentation.

Education

PhD in Natural Language Processing, LISN - Paris-Saclay University

2023-2026

MSc in Applied Maths and Statistics, University of Rennes/INSA/ENSAI

2019-2021

Publications

Small Language Models are Good Too: An Empirical Study of Zero-Shot Classification

LREC-COLING, 2024

Leveraging IR to Enhance SLU Prompts in Few-Shot Learning

Interspeech, 2025

Détection des contaminations de LLM par extraction de données : une revue de littérature pratique

TALN, 2025

Format Matters: A Critical Evaluation of Output Formats for Prompting LLMs in SLU and NER

LREC, 2026

Small generative models for industrial applications

My thesis manuscript, 2026

Projects

LLM Reliability and Data Leakage Watch

Studied contamination detection and benchmark integrity to better characterize when evaluation results can be trusted.

NER Evaluation Pipeline

Open-sourced reproducible training and evaluation tooling for named entity recognition benchmarks and model comparisons.

Implementing and documenting scientific papers

Lastly: Cross-LoRA: A Data-Free LoRA Transfer Framework across Heterogeneous LLMs

Fine-Tuning Tiny Models for On-Device Inference

Optimized sub-500M models for constrained deployments and published checkpoints on Hugging Face.

Nom et prénom du doctorant : LEPAGNOL Pierre

Titre de la thèse : Petits modèles génératifs en contexte industriel : Adaptation par prompting avec peu de données

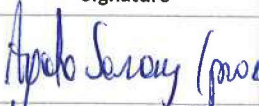
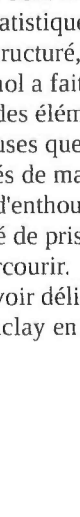

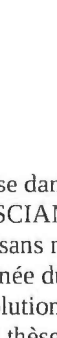
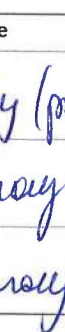
École Doctorale : Sciences et Technologies de l'Information et de la Communication

Pôle : B : Données, connaissances, apprentissage et interactions

Date de la soutenance : 10 mars 2026

Président du jury : Agata SAVARY (À COMPLÉTER)

Membres du jury avec voix délibérative :

Nom	Signature	Nom	Signature
François PORTET		Frédéric BECHET	
Agata SAVARY		Nathalie CAMELIN	
Benoît FAVRE			

Les membres du jury attestent avoir pris connaissance de l'intégralité du rapport.

Ils attestent que le directeur ou la directrice de thèse, ainsi que toute autre personne ayant participé à la direction de la thèse ou l'encadrement du doctorant ou de la doctorante, n'a pas pris part à la décision.

Si le rapport comporte plusieurs pages ou s'il est rédigé sur un document distinct, il devra être paraphé sur chaque page et signé par le Président du jury.

Pierre Lepagnol a mené ses travaux de thèse dans un contexte exigeant. Il s'agit d'une thèse CIFRE intégrant des besoins spécifiques de l'entreprise partenaire SCIAM, notamment liés aux petits modèles de langues, pouvant être utilisés en interne pour une variété de clients, sans nécessiter un apprentissage spécifique. Les travaux de thèse se déroulaient dans une période d'évolution effrénée du domaine de l'intelligence artificielle et spécifiquement du traitement automatique des langues. Cette évolution a généré un très grand nombre d'articles et de résultats directement liés aux questions de recherche posées dans la thèse. Ceci a nécessité une mise à jour permanente de ses connaissances et la remise en question des hypothèses et méthodes. Le jury estime que ce défi a été tenu.

Les questions de recherches posées ont engendré un grand nombre d'expériences, menées de manière bien structurée. Les protocoles d'évaluations, dont la rigueur a été particulièrement appréciée par le jury, incluaient notamment des tests de significativité statistique systématiques. Les analyses des résultats ont été larges et intéressantes.

Le manuscrit de thèse est très bien structuré, bien rédigé dans l'ensemble, et offre un fil conducteur clair.

Lors de la soutenance, Pierre Lepagnol a fait une présentation limpide, synthétique et dynamique, couvrant bien l'ensemble de ses travaux, et contenant des éléments nouveaux par rapport au manuscrit.

Il a également répondu à de nombreuses questions du jury, en apportant beaucoup d'éléments pertinents, même si certains auraient parfois pu être présentés de manière plus synthétique.

Lors de ce dialogue, il a fait preuve d'enthousiasme et d'une solide culture du domaine.

Il a également démontré une capacité de prise de recul par rapport à sa propre méthodologie, aux limites actuelles de la discipline et au chemin qui reste à parcourir.

Pour toutes ces raisons et après en avoir délibéré, le Jury a prononcé à l'unanimité l'admission de Pierre Lepagnol au grade de docteur de l'université Paris-Saclay en Informatique.

