Channdeth SOK

✉ channdeth.sok@gmail.com · channdethsok · 🌐 http://channdethsok.github.io

Education

2024 – 2025	**M.Sc. Data Science, Institut Polytechnique de Paris**.
2022 – 2025	**Diplôme d'Ingénieur (Master's level), ENSAE Paris**.
2019 – 2022	**Engineering program, Institute of Technology of Cambodia**.

Professional Experience

May 2025 – Nov 2025	**Data Science Intern**, Forvia (Paris Tech Center) Conducting research on detecting hallucinations in Retrieval-Augmented Generation (RAG) systems and agentic AI.
Mar 2024 – Aug 2024	**AI Engineer Intern**, AI-VIDENCE Built a trustworthy conversational agent (chatbot) capable of accurately answering questions based on documents.
Jun 2023 – Sep 2023	**Data Science & AI Intern**, Paris Partners Softwares Developed a tool to assess job applicants and explain decisions based on job descriptions using LLM

Research Publications

[1] C. Sok, D. Luz, and Y. Haddam, *Metarag: Metamorphic testing for hallucination detection in rag systems*, 2025. arXiv: 2509.09360 [cs.CL]. 🔗 URL: https://arxiv.org/abs/2509.09360.

Skills

Languages	English (C1), French (B2), Khmer (Native)
Programming	Python, R, Java, SQL, LaTeX

Miscellaneous Experience

Awards and Achievements

2022	**Eiffel scholarship** — French government scholarship for graduate studies at ENSAE Paris.
2019	**Grade A** — Top 20 of 100,000 candidates in the national baccalaureate exam.

Projects

March 2025 – Apr 2025	**Annealed Sinkhorn for Optimal Transport, Optimal Transport course** Implemented in OTT-JAX the algorithm from L. Chizat (2024), *Annealed Sinkhorn for Optimal Transport: convergence, regularization path and debiasing* Colab notebook available. *Tech: Python, JAX, OTT-JAX, Optimal Transport*
	Legal Document Classification with EuroBERT, ML for NLP course Implemented multilingual legal document classification on the MultiEURLEX dataset (English, French, Spanish). Compared full fine-tuning with PEFT methods (LoRA, BitFit) applied to EuroBERT. *Tech: LLMs, EuroBERT, HuggingFace, PEFT*

Projects (continued)

Jan 2025 – Apr 2025 **Agnostic LLM Information Retriever, VINCI**
Developed a generic information retriever for LLMs, robust for general use cases and customizable for domain-specific tasks. *Tech: LLMs, LangChain, RAG, HuggingFace, RAGAS*

Nov 2022 – May 2023 **Explaining Performance Gaps in Mathematics, Collège de France**
Analyzed DEPP panel data to investigate determinants of mathematics performance gaps among French students. *Tech: Visualization, EDA, Regression*

Sep 2022 – Dec 2022 **Research of Potential Countries for Education Investment, Python for Data Science course**
Built composite indicators to rank countries based on education investment opportunities. *Tech: Visualization, EDA, PCA, Scoring*

Jan 2025 – Apr 2025 **Agnostic LLM Information Retriever, VINCI**