



Francis Kulumba

PH.D STUDENT · DATA SCIENTIST

✉ franciskulumbacs@gmail.com | 🏠 madjakul.com | 📺 Madjakul | 📺 francis-k | 🐦 @Madjakul

Ph.D student in computer and information sciences (NLP). Data Scientist. Working on author names disambiguation in academic papers.

Education

Sorbonne Université

PH.D IN COMPUTER AND INFORMATION SCIENCES

Paris, France

Nov. 2022 - Oct. 2025

- **Subject:** Author name disambiguation in academic papers.
- **Lab.:** ALMAnaCH, Inria Paris.
- **Adviser:** Laurent Romary.

École Supérieur d'Ingénieur Léonard de Vinci (ESILV)

ENGINEER'S DEGREE IN MATHEMATICS AND COMPUTER SCIENCE

Courbevoie, France

Sep. 2017 - Jun. 2022

- Data and Artificial Intelligence.

Chung-Ang University

UNDERGRADUATE EXCHANGE IN MATHEMATICS AND COMPUTER SCIENCE

Seoul, South Korea

Sep. 2019 - Dec. 2019

- College of Software.

Experience

French Ministry of Defense

PH.D STUDENT, DATA SCIENTIST

Paris, France

Nov. 2022 - Oct. 2025

- Implemented custom RAG frameworks.
- Crafted document-based synthetic instruction data for LLMs

French Ministry of Defense

DATA SCIENTIST APPRENTICE

Paris, France

Sep. 2020 - Aug. 2022

- Implemented a custom weakly supervised named entity tagger.

Teaching

2023 - Present

University of Paris I: Panthéon-Sorbonne — Teaching Assistant

Paris, France

Relational Database Tutorials. UML/Python Tutorials.

2024

École pour l'informatique et les techniques avancées (EPITA) — Lecturer

Paris, France

Advanced NLP Lectures.

Skills

Domain Knowledge

Natural Language Processing, Machine Learning, Deep Learning, Language Modeling, Information Extraction, Information Retrieval, Graph Theory

Programming

Python, C++

Scripting

TeX, Bash, Slurm

Softwares and Frameworks

Git, GNU/Linux, Elasticsearch, PyTorch, HuggingFace Transformers, CUDA

Interpersonal

Pedagogy, Vulgarization

Languages

French (Native), **English** (Fluent), **Spanish** (Intermediary), **Korean** (Notions)

Other Interests

Tutoring, Help French high school students with mathematics and computer science.

Reading, https://www.babelio.com/monprofil.php?id_user=1951395

Sport, Used to be on the reserves of the youth French national team of American football.

Publications, Softwares & Corpora

FEATURED PUBLICATIONS

OTHER PUBLICATIONS

- [1] [Francis Kulumba](https://arxiv.org/abs/2407.20595), Wissam Antoun, Guillaume Vimont, and Laurent Romary. ***Harvesting Textual and Structured Data from the HAL Publication Repository***. 2024. arXiv: 2407.20595 [cs.DL]. URL: <https://arxiv.org/abs/2407.20595>.