Allaa Boutaleb

Sorbonne University, CNRS, LIP6, F-75005 Paris, France

⚲ Paris, FR   **✉** allaa.boutaleb@lip6.fr   **⚲** allaaboutaleb.com   **in** maeboutaleb   **○** Allaa-Boutaleb

My PhD research at the intersection of Machine Learning, NLP, and Data Management focuses on table representation learning. I'm currently working on techniques for data discovery and integration, such as table union search, table join search and table reclamation, to maximize the utility of large-scale data lakes.

EDUCATION

Sorbonne University of Science & Engineering *Paris, France*

PhD Computer Science	*10/2024 – Present*
MSc Computer Science	*09/2022 – 09/2024*
BSc Computer Science	*09/2021 – 06/2022*

CELSA - Higher School of Communication *Paris, France*

Diploma in Communication & Entrepreneurship	*10/2023 – 07/2024*

ESI - Higher School of Computer Science *Algiers, Algeria*

Preparatory Classes	*09/2018 – 07/2021*

EXPERIENCE

Teaching Assistant   *Paris, FR*
Sorbonne University   *Oct 2024 – Present*

- Instructed undergraduate courses on Python, SQL, and database management.
- Managed the full examination lifecycle, including drafting, proctoring, and grading assessments.

AI Engineer   *Paris, FR*
Sujet AI   *Nov 2023 – Present*

- Curated and open-sourced diverse training datasets for financial LLMs, VLMs, and embedding models.
- Trained and deployed a variety of open-sourced, fine-tuned models for specialized financial tasks.
- Engineered a multi-agent system to analyze financial documents and generate custom reports.

Sujet AI's Hugging Face ☑ — *sujet.ai* ☑

AI Research Intern   *Paris, FR*
RTE France   *Mar 2024 – Sep 2024*

- Deployed a prototype using online topic modeling to detect emerging trends from social media.
- Authored "BERTrend," a research paper accepted at the EMNLP 2024 FuturED workshop.

github.com/rte-france/BERTrend ☑

Research Assistant   *Paris, FR*
LIP6 Laboratory   *June 2023 – Sep 2023*

- Investigated Aligned Neural Topic Modeling (ANTM) to uncover latent data structures.
- Proposed a topic-modeling method to detect paradigm shifts in large scientific archives.

github.com/Allaa-boutaleb/ANTM-Investigations ☑

PUBLICATIONS

A Crtical Re-Evaluation of Table Union Search Benchmarks   TRL @ ACL
Allaa Boutaleb, Bernd Amann, Hubert Naacke, Rafael Angarita   Vienna, AT
                                                                         2025

aclanthology.org/2025.trl-1.7.pdf ☑

HEARTS: Hypergraph-based Related Table Search   RLGMSD
Allaa Boutaleb, Alaa Almutawa, Bernd Amann, Rafael Angarita, Hubert Naacke   Amsterdam, NL
                                                                         2025

openreview.net/forum?id=XgRbxO9pLJ ☑

BERTrend: Neural Topic Modeling for Emerging Trends Detection   FuturED @ EMNLP
Allaa Boutaleb, Jérôme Picault, Guillaume Grosjean   Miami, Florida, US
                                                                         2024

10.18653/v1/2024.futured-1.1 ☑