Franck SIGNE TALLA

franck.signe-talla@polytechnique.edu | linkedin.com/in/franck-signe-talla | www.github.com

EDUCATION

Master 2 Data Science (M2DS) Palaiseau, France

Ecole Polytechnique *Sept. 2023 – Now*

- Relevant courses includes : *Advanced Learning for Text and Graph, Reinforcement Learning, Computer Vision and Object recognition, Audio and music information retrieval*

Master's degree in Applied Mathematics and Computer Science Palaiseau, France

Ecole Polytechnique *Sept. 2020 – Sept 2023*

- Relevant coursework includes : *Machine Learning and Deep Learning, Text Mining and NLP, Deep Learning from Theory to Practice, Regression and Classification, Data Structures and Algorithms, Optimization.*

INTERNSHIP EXPERIENCE

DataLab Credit Agricole Paris, France

Data Scientist Intern *April 2023 – August. 2023*

- Developed and designed custom Active Learning strategies tailored specifically for the task of extracting information from scanned documents, drawing inspiration from well-established techniques in various domains like classification.
- Introduced innovative strategies previously unaddressed in the existing literature, highlighting a dedication to adapting and evolving established methods for new and distinctive use cases.
- Conducted comprehensive testing and evaluation of these specialized strategies using the company's proprietary data and publicly available datasets.

EvidenceB Paris, France

Data Scientist Intern *June 2022 – Août. 2022*

- Reviewed the state of the art algorithms used for adaptive spacing in student's learning for the optimization of the long-term retention and implemented a human memory model in learning process based on this **thesis**, usable in the specific case of learning software developed by the company, using Python (Numpy, Pandas, Sklearn).

PROJECTS

Synthetic Dataset Generation Oct. 2023

PEFT, Torch, Transformers

- Participated in a team at NeurIPS LLM challenge 2023, fine-tuning an open-source model with open-source datasets on a single GPU within a one-day timeframe. Contributed to bias reduction by generating synthetic datasets using the self-instruct approach with the Mistral 7B model, as part of the evaluation metric for model performance.

AI generated text Detection fev. 2023 – March. 2023

Hugging Face, Transformers

- Led a project focused on text classification to distinguish AI-generated text from human-generated text using Transformer-based models, including DeBERTa for classification, and GPT-2 and T5 for feature extraction based on perplexity, achieving an accuracy of 90.6%.

Information Retrieval For Question Answering Task Oct. 2022 – Nov. 2022

Code

- Developped a model that predicts the best paragraph that could answer to a given question among a set of paragraphs. For that purpose we used KNN algorithm to rank paragraphs by relevance. We used different **NLP** embeddings such as *TF-IDF* and *Transformer-based* embedding. We evaluated our model using the *top-k accuracy.*

ADDITIONAL INFORMATION

- **LANGUAGES**: French(Native), English

TECHNICAL SKILLS

- **LANGUAGES**: Python, C++.
- **Python Libraries**: Pandas, NumPy, Matplotlib, Sklearn, Keras, Pytorch.