Matthieu Futeral

Tél. : +33 6 68 98 09 16
Email : matthieu.futeral@inria.fr

EDUCATION

PhD Candidate - ENS Ulm - Paris, France **2021 – 2024**

Supervised by Ivan Laptev, Rachel Bawden, Benoît Sagot and Cordelia Schmid

Master MVA - ENS Paris-Saclay, IP Paris - Gif-sur-Yvette, France **2020 – 2021**

Coursework:

- Computer vision, Speech, NLP, Digital image processing, Deep learning, Reinforcement learning, Functional Brain Imaging (BCI), Graph in Machine Learning

Relevant projects:

- Kaggle Competition: Fine-grained image classification, rank 2 out of 167 participants.
- Brain Computer Interface: Detect erroneous feedback in a P300-Speller experiment.

ENSAE Paris – Palaiseau, France **2017 – 2021**

Master in Engineering - France's leading school in statistics and economics

Coursework:

- Statistics, Machine learning, NLP, Optimization, Monte-Carlo methods, GPU Programming, Optimal Transport

Relevant projects:

- Exploration and evaluation of word embeddings on Twitter data. A study case on sentiment analysis (NLP, 2018).
- Conditional fake news generation with state-of-the-art language models (NLP, 2020).

Classe Préparatoire – Paris, France **2014 – 2017**

Preparatory courses for entrance into French "Grandes Écoles" - admitted to ENSAE Paris

WORK EXPERIENCE

PhD Candidate | INRIA (Paris, France) **Nov. 2021 – Nov. 2024**

Previously Research Intern (May. 2021 - Oct. 2021)

Willow & Almanach teams

- Natural Language Processing & Computer Vision: Multilingual and Multimodal modeling

Multimodal Machine Translation: How to introduce images and videos into machine translation models so that they produce better translations? [1]

Supervised by Ivan Laptev, Rachel Bawden, Benoît Sagot and Cordelia Schmid

Teaching Assistant (NLP) | ENSAE (Palaiseau, France) **Nov. 2021 – May. 2022**

- I supervised a group of 20 students during the NLP course at ENSAE Paris and I evaluated their research projects.

Research Intern | University of Cambridge (Cambridge, UK) **Jan. 2020 – Jul. 2020**
Department of Public Health and Primary Care - Cardiovascular epidemiology Unit (CEU)

- Multiple imputation of missing data from a very large clinical dataset (UK Biobank).

UK Biobank is a very large longitudinal dataset with biological measures from UK patients over the years. The goal of the internship was to compare classic statistical multiple imputation models with modern neural-based approaches (deep learning) and to propose an approach to impute missing data from UK Biobank.

NLP Research Intern | INRIA (Paris, France) **Jul. 2019 – Dec. 2019**
Almanach Team

- Identify *NArabizi* (informal north-african dialect) language in the crawled web in order to create the first *NArabizi* dataset. [2]
- Evaluate gender and ethnic biases in state-of-the-art language models.

Languages : French (native), English (advanced - 950/990 TOEIC), Spanish (intermediate)
Skills : Python (Pytorch, Scikit-Learn), C, CUDA, Bash, LaTeX

PUBLICATIONS

[1] **Tackling ambiguity with images: Improved Multimodal Machine Translation and Contrastive Evaluation.** pdf code
Matthieu Futeral, Ivan Laptev, Benoît Sagot, Cordelia Schmid, Rachel Bawden (2022). *arXiv preprint* (*Under review*)

[2] **Building a User-Generated Content North-African Arabizi Treebank: Tackling Hell.** pdf
Djamé Seddah, Farah Essaidi, Amal Fethi, **Matthieu Futeral**, Benjamin Muller, Pedro Javier Ortiz Suárez, Benoît Sagot and Abhishek Srivastava (2020). In *The 58th Annual Meeting of the Association for Computational Linguistics,* Seattle, Washington, United States. (ACL 2020)