NDEYE MAGUETTE MBAYE

Personal page | Github | LinkedIn

+33 6 64 67 93 74

Sartrouville (France)

maguycf@live.fr

DATA SCIENTIST

PROFESSIONAL PROFILE

As a PhD graduate in Bioinformatics, I am seeking opportunities as a Data Scientist in a dynamic and collaborative team. I have strong experience in Natural Language Processing (NLP), with a focus on applying advanced methods such as Large Language Models (LLMs) to solve complex problems in breast cancer research.

Machine Learning - Deep Learning - Natural Language Processing - Transformers - Electronic Health Records - Multimodal Learning.

EDUCATION

2019 - 2024

Ph.D. in Bioinformatics, Mines Paris PSL | ED Ingénierie des Systèmes, Matériaux, Mécanique, Energétique.

2016 - 2019

Master in Biomathematics – Bioinformatics, Université Cheikh Anta DIOP de Dakar | Faculté des Sciences et Techniques.

2015 - 2016

Bachelor in Biomathematics – Bioinformatics, Université Cheikh Anta DIOP de Dakar | Faculté des Sciences et Techniques.

SKILLS

- **Programming Languages:** Python, R, Bash.
- **AI tools**: ML libraries (numpas, pandas, scikit-learn etc.), DL frameworks (Pytorch), NLP frameworks (Hugging face, Transformers, spaCy, BeautifulSoup, NLTK), Data visualization (matplotlib, seaborn, ggplot2).
- **Soft skills**: communication, curiosity, adaptability, critical thinking.
- **Languages skills**: French, English.

SELECTED TALKS AT CONFERENCES

- **Women in Machine Learning and Data Science**, more information here | Paris, France | September 20th, 2022.
- **Machine Learning Frontiers for Precision Medecine**, more information here | Munich, Germany | October 18th, 2022.
- **Winter School in Computer Science and Engineering**, more information here | Jerusalem, Israël | January 15th, 2023.
- **Computational Systems Biology of Cancer**, more information here | Paris, France | September 25th, 2023 – ***with the Award of the best talk in Ph.D. Students category.***
- **Personalized Health Conference**, more information here | Basel, Switzerland | March 3rd, 2024.

SELECTED WORK EXPERIENCES

September 2019 - September 2024 | 5 years

Ph.D. Student & MLFPM alumni – Center for Computation Biology | Institut Curie, Mines Paris PSL and INSERM | Paris, France.

- Developed classical ML models for multimodal EHR* to predict breast cancer prognostic status.
- Developed Transformer-based models for multimodal EHR* to predict breast cancer prognostic status.
- Interpreted built models and found potential prognostic factors.

** Multimodal Electronic Health Records: free-text medical reports, laboratory tests and clinical data.*

November 2022 - January 2023 | 3 months

Ph.D. intern – Max Planck institute for psychiatry | Munich, Germany

- Developed Tansformer-based models for multimodal EHR using late integration methods
- Tested different embedding methods.

June - September 2021 | 3 months

Ph.D. intern – IBM Research Lab | Haifa, Israel | June – September 2021

- Explored Tansformer-based models for tabular EHR based on the BEHRT model.

April - December 2018 | 8 months

Bioinformatics research assistant | Laboratoire Commun de Microbiologie | Institut de Recherche pour le Développement (IRD), Institut Sénégalais de Recherches Agricoles (ISRA) and Université Cheikh Anta DIOP de Dakar (UCAD) | Bel Air, Dakar | April – December

- Developed dynamic boolean model of gene regulatory network involved in root development in *Arabidopsis thaliana*.
- Validated built model with other species
- Informed on critical genes involved in lateral roots development and highlighted the differential expressions of genes in that process.

SCIENTIFIC PUBLICATION

- Ndèye Maguette Mbaye, Michael Danziger, Aullène Toussaint, Elise Dumas, Julien Guerin, Anne-Sophie Hamy-Petit, Fabien Reyal, Michal Rosen-Zvi, Chloé-Agathe Azencott. (2024) Multimodal BEHRT: Transformers for Multimodal Electronic Health Records to predict breast cancer prognosis – *submitted* .