François Hu

AI Research Scientist | Lecturer in statistical learning

E-mail
hu.faugon@gmail.com

Webpage
https://curiousml.github.io/

Github
curiousML

MLOps tools
Code versioning (Git):
●●●●●◐
Model versioning (MLflow):
●●●●○
Data versioning (DVC):
●●●●◐
ML orchestration (Kedro):
●●●●●◐
CI/CD (Jenkins/Gitlab):
●●●◐○

Programming
Python, R : ●●●●●
C, C++: ●●●●○
SQL, noSQL : ●●●●◐

DataViz
Shiny (R) : ●●●●◐
Plotly/ggplot : ●●●●○
Bokeh (Py) : ●●●◐○

Languages
French: native
English: C1
Chinese: B2
Italien: B2
Spanish: learning (A2)

Soft skills
Autonomous, creative, *agile* mindset, relational intelligence, rigorous

Activities
Audiobook, hiking, badminton, climbing.

Professional Experience

2022 - 2024	**AI Research Scientist**	Milliman R&D, Paris

Data Scientist R&D specialized in Statistical Learning, Fairness and Interpretability

2022 - 2024	**Postdoctoral researcher in statistical learning**	Université de Montréal, (QC)

- **Topics**: Data science, fairness, transparency, ESG, NLP, computer vision.
- **with** *Arthur Charpentier* and *Manuel Morales*.
- **Responsibilities**: (1) Define projects / objectives, (2) create robust models with stat. guarantees and (3) supervise interns, Master's and PhD students

2019 - 2022	**Data scientist**	Société Générale Insurance, La défense

3 years - machine learning, statistics, NLP and computer vision.
Responsibilities: (1) Guide interns in their R&D projects, (2) assist in the training of the data science team in new AI tools, and (3) establish research partnerships

2018 - 2019	**Data scientist - junior**	Société Générale Insurance, La défense

scoring, model deployment and contribution to the (internal) ML platform

Apr 2018	**Data scientist - intern**	Société Générale Insurance, La défense

6 months - statistics and DataViz on telematic data

Education

2019 - 2022	**PhD in applied mathematics - statistics**	ENSAE - Institut Polytechnique de Paris

- **SCOR 2022 Award** for the best actuarial thesis.
- **Thesis title:** Semi-supervised learning in insurance: fairness and labelling
- **Topics:** Statistical learning, algorithmic fairness, NLP, insurance
- **with** Romuald Elie, Caroline Hillairet and Marc Juillard

2016 - 2018	**Master's degree in applied mathematics**	Sorbonne Univ. (ex-UPMC), Paris VI

Statistics and data science

2013 - 2016	**Bachelor's degree in mathematics**	Sorbonne Univ. (ex-UPMC), Paris VI

Intensive bi-disciplinary program in mathematics and computer science

2013 - 2016	**Bachelor's degree in computer science**	Sorbonne Univ. (ex-UPMC), Paris VI

Intensive bi-disciplinary program in mathematics and computer science

2015 - 2021	**Summer schools (and visiting research)**	

- *July 2021 - HEC et ENSAE (France)* : ML in asset management
- *March 2020 - Berkeley University (Visiting research, USA)*: Active learning
- *July 2015 - Brown University (USA)*: MATLAB and numerical analysis

Teaching

2020 - *ongoing*	**Master of Science in Artificial Intelligence Systems**	EPITA

Lecturer in python (42H), machine learning (30H) and bayesian statistics (18H).

2019 - 2021	**Data Science, Statistics and Learning**	ENSAE

TA in python (1A), monte-carlo (2A), intro & advanced machine learning (2-3A).

2019 - 2022	**Applied statistics & econometrics**	ENSAE

This course supplements theoretical training by teaching ENSAE students practical skills in statistical modeling, enabling them to work with real-world data.

2019 - 2022	**Data Science for Actuaries (DSA)**	Institut des Actuaires

Lecturer in NLP & Deep Learning.

Notable publications

2023   **Fairness guarantee in multi-class classification**   arXiv - revision
- Joint work with *Christophe Denis*, *Romuald Elie* and *Mohamed Hebiri*.
- https://arxiv.org/abs/2109.13642

2023   **Sequentially Fair Mechanism for Multiple Sensitive Attributes**   AAAI 2024
- Joint work with *Philipp Ratz* and *Arthur Charpentier*
- *The 38th Annual AAAI Conference on Artificial Intelligence (AAAI-24)*. https://arxiv.org/abs/2309.06627

2023   **Fairness in multi-task learning via wasserstein barycenters**   ECML-PKDD 2023
- Joint work with *Philipp Ratz* and *Arthur Charpentier*.
- *Machine Learning and Knowledge Discovery in Databases: Research Track (ECML PKDD 2023), 2023.* https://arxiv.org/abs/2306.10155

2023   **Parametric fairness with statistical guarantees**   arXiv - revision
- Joint work with *Philipp Ratz* and *Arthur Charpentier*
- https://arxiv.org/pdf/2310.20508.pdf

2023   **Bias and explainability in image recognition**   arXiv - revision
- Joint work with *Philipp Ratz* and *Arthur Charpentier*
- https://arxiv.org/abs/2308.11090

2023   **Solving the labelling problem in Insurance**   arXiv - revision
- Joint work with *Romuald Elie*, *Caroline Hillairet* and *Marc Juillard*.
- https://arxiv.org/abs/2112.09466

2023   **Mitigating discrimination in insurance**   BIAS 2023
- Joint work with *Arthur Charpentier* and *Philipp Ratz*.
- *BIAS 2023, 3rd Workshop on Bias and Fairness in AI, International Workshop of ECML PKDD, 2023.* https://arxiv.org/abs/2306.12912

2024   **Models and data used to predict the density and distribution of Ixodes scapularis (blacklegged tick) in North America: a scoping review.**   The Lancet - Am.
- Joint work with *Yogita Sharma*, *Elda K.E. Laison*, *Tanya Philippsen*, *Junling Ma*, *Jude Kong*, *Sajjad Ghaemi*, *Juxin Liu* and *Bouchra Nasri*

Notable projects

2023 - ...   **Large Language Models (LLM) for sustainable finance**   Algora Lab - MILA
- Joint research focused on algorithmic fairness and NLP (traditional & non-traditional with prompt engineering or conversational AI) techniques to identify Environmental, Social, and Governance (ESG) concepts.
- https://algoralab.ca/project/forum-inventia-jalisco-mexique-with-fair-lac/