Jonathan Crabbé

Cambridge, United Kingdom 🏫

- +44 7547340408 📞
- jonathan.cr1302@gmail.com 💌
- linkedin.com/in/jonathancrabbe in
 - jonathancrabbe.github.io 🌐
 - github.com/JonathanCrabbe 🔿
 - Born Feb. 1996, Belgium <

I am looking for a *research* position where I can put my curiosity, knowledge and passion at the service of pushing the limits of *machine learning* (ML). I have a strong experience (through industry projects / publications) and interest for the following themes.

Data-Centric Al

Geometric Deep Learning

• Time Series Modelling

AI for Science and Medicine

- Robust Machine Learning
- Interpretability
- Generative AI
- Representation Learning
- Deep Ensembles
- Experience

MAY 2023 - SEPTEMBER 2023

Research Intern Apple

Description. 5 months internship in the Machine Learning Research team, where I studied the robustness of state-of-the-art multimodal machine learning models. We published the results in a paper in which we discover outlier features and high polysemanticity in zeroshot CLIP models. We show that only 20% of the latent space directions contribute to the model effective robustness. **Transferable skills.** Robust ML, multimodal ML, creating a large codebase from scratch, training SOTA multimodal models with accelerators, continuous integration, unit testing.

JAN 2023 – MAY 2023

Research Intern

Microsoft

Description. 3 months internship in the AI4Science team, where I contributed to the development of new machine learning approaches for material discovery. I implemented a new type of diffusion for categorical data. My results will be published as part of a big paper.

Transferable skills. Generative AI, diffusion models, scientific applications of AI, contributing to large codebases, training large models with accelerators, continuous integration, unit testing.

JUNE 2022 – SEP 2022

Quantitative Research Intern G-Research

Description. 10 weeks internship where I applied SOTA machine learning approaches to financial data. I have learned to build an entire modelling pipeline to transform raw noisy data into predictive features that can be used to train a machine learning model.

Transferable skills. Data cleaning and labelling, feature engineering, modelling financial markets.

2016 – 2020 Content Creator Clipedia

Description. Several years of involvement in the teaching project Clipedia. I have created pedagogical videos on YouTube to teach math, physics, and chemistry to young students. The YouTube channel has over 100k subscribers and this number is quickly increasing. Please visit <u>https://clipedia.be</u> for more information.

Transferable skills. communication, pedagogy, ability to adapt a speech to a wide audience

Education

2020-2024

PhD in Machine Learning University of Cambridge, Mathematics Department

Description. My thesis has a strong focus on developing and improving machine learning approaches to tackle high-stakes domains, such as medicine. I have published *13 papers* in NeurIPS/ICML/ICLR (7 as first author). Through this line of work, I have contributed to the development of an entire interpretability environment available in the form of a public repository https://github.com/vanderschaarlab/Interpretability. I was supervised by Mihaela van der Schaar.

2018-2019

MASt in Applied Mathematics and Theoretical Physics University of Cambridge

Description. This master is one of the most challenging math course in the world. Succeeding required me to master complex mathematical paradigms such as *differential geometry, group theory* as well as *functional analysis*. This formal knowledge perfectly complements the practical skills that I have acquired during my bachelor.

Grade: Distinction (highest grade)

Sample courses. Black Holes: 89%, Symmetries, Fields & Particles: 88%, Advanced Quantum Field Theory: 87%

2017-2018

M1 in Physics

Ecole Normale Supérieure Paris

Description. This master is a research-oriented course in physics. To achieve an excellent grade in this course, I had to quickly *overcome the challenging transition* between engineering and physics. **Grade.** Mention Bien

2014-2017

Bachelor's in engineering Université Libre de Bruxelles

Description. This bachelor's in engineering taught me to solve practical problems with a *principled* and *pragmatic approach*. I also developed a real interest for using *computer science to solve real-world problems*. In addition, I discovered a passion for *teaching* as a teaching assistant.
Grade. The Highest distinction, *first of my promotion* (among 450 students in first year)
Sample courses. Analysis 1&2: 100% for both, Complements of mathematics and numerical analysis: 95%, Physics: 100%, Oriented Object Programming: 90%, Numerical Analysis: 95%, Probability and Statistics: 95%, Signals and Systems: 95%, Quantum Mechanics: 90%

Skills

- **Coding.** Strong experience in Python ML development. Contribution to large codebases through my positions in the tech industry. My coding process is rigorous, it involves unit testing with high coverage and continuous integration. I have a good knowledge of how to work with cloud computing services, such as Azure, AWS and GCP.
- **Modelling.** Trained large and small machine learning models for a wide range of tasks. Thorough experience training models on medical, financial, vision and graph datasets. I have finetuned SOTA multimodal models in terms of effective robustness.
- **Proving.** Most of my papers have a strong theoretical components with rigorous original proofs. I often use probability theory, statistics, measure theory, functional analysis, group theory, linear algebra and differential geometry results in my theoretical analysis.
- **Presenting.** Through many talks and a strong teaching experience, I have learned to present ideas in an inspiring way for a wide variety of audiences.
- Working in a team. Half of my papers are the result of a collaboration with several co-authors. I have also worked in medium-sized (10 people) teams during several internships.
- Autonomy. The other half of my papers are the result of autonomous work.
- **Supervising.** I have personally supervised 2 research internships (1 Cambridge MPhil student and 1 Imperial MSc student) and helped 3 PhD students to define their research agenda.
- Languages. French (native), English (IELTS Band 8).

Presentations

- **NeurIPS 2023.** 4 papers presented in the main conference.
- **IGARSS 2023.** 1 papers presented in the conference.
- ICLR 2023. 1 paper presented in the main conference.
- **NeurIPS 2022.** 3 papers presented in the main conference.
- ICML 2022. 3 papers presented in the main conference.
- NeurIPS 2021. 1 spotlight paper (top 3% paper) presented in the main conference..
- Stanford MedAl Series. Long presentation on Explainable Al for time series.
- **OxML 2021.** Full summer school attendance.
- ICML 2021. 1 paper presented in the main conference.
- **NeurIPS 2020.** 1 paper presented in the main conference.

Awards

- AVIVA PhD Fellowship to fund my PhD (2020).
- ULB Research Assistant Fellowship to fund my year as a research assistant (2019).
- Wolfson College Jennings Price awarded based on outstanding results for my MASt (2019).
- Labex-ICFP Scholarship to fund my M1, awarded based on academic excellence (2018).

Publications

Reference	Published in	Subjects
J. Crabbé, M. van der Schaar (2023). Evaluating the Robustness of Interpretability Methods through Explanation Invariance and Equivariance.	NeurIPS 2023	Interpretability, Robust ML, Representation Learning, Geometric Deep- Learning, NLP
A. Jeffares, T. Liu, J. Crabbé, M. van der Schaar (2023). Joint Training of Deep Ensembles Fails Due to Learner Collusion.	NeurIPS 2023	Deep Ensembles, Representation Learning, Computer Vision
N. Seedat, J. Crabbé, M. van der Schaar, M. (2023). TRIAGE: Characterizing and auditing training data for improved regression.	NeurIPS 2023	Data-Centric Al, Fairness, Tabular Data
H. Sun, B. van Breugel, J. Crabbé, N. Seedat, M. van der Schaar (2022). Latent Density Models for Uncertainty Categorization.	NeurIPS 2023	Representation Learning, Uncertainty Estimation, Interpretability
A. Vandenhoeke, L. Antson, G. Ballesteros, J. Crabbé, M. Shimoni (2023). Explaining the Absorption Features of Deep Learning Hyperspectral Classification Models	IGARSS 2023	Interpretability, Computer Vision
A. Jeffares, T. Liu, J. Crabbé, F. Imrie, M. van der Schaar (2023). TANGOS: Regularizing Tabular Neural Networks through Gradient Orthogonalization and Specialization	ICLR 2023	Representation Learning, Tabular Data
J. Crabbé, M. van der Schaar (2022). Concept Activation Regions: A Generalized Framework For Concept-Based Explanations.	NeurIPS 2022	Interpretability, Representation Learning, Computer Vision
N. Seedat, J. Crabbé, I. Bica, M. van der Schaar (2022). Data-IQ: Characterizing subgroups with heterogeneous outcomes in tabular data.	NeurIPS 2022	Data-Centric Al, Fairness, Tabular Data
J. Crabbé, A. Curth, I. Bica, M. van der Schaar (2022). Benchmarking Heterogeneous Treatment Effect Models through the Lens of Interpretability.	NeurIPS 2022	Interpretability, Treatment Effect Estimation
J. Crabbé, M. van der Schaar (2022). Label-Free Explainability for Unsupervised Models.	ICML 2022	Interpretability, Generative AI, Representation Learning, Computer Vision
N. Seedat, J. Crabbé, M. van der Schaar (2022). Data-SUITE: Data-centric identification of in- distribution incongruous examples.	ICML 2022	Data-Centric Al, Fairness, Tabular Data
J. Crabbé, Z. Qian, F. Imrie, M. van der Schaar (2021). Explaining Latent Representations with a Corpus of Examples.	NeurIPS 2021	Interpretability, Representation Learning, Tabular Data, Computer Vision
J. Crabbé, M. van der Schaar (2021). Explaining Time Series Predictions with Dynamic Masks.	ICML 2021	Interpretability, Time Series Modeling

NeurIPS 2020

Interpretability, Symbolic Regression

J. Crabbé, Y. Zhang, W. Zame, M. van der Schaar (2020). Learning outside the black-box: the pursuit of interpretable models.