Cambridge, United Kingdom
07547340408
jonathan.cr1302@gmail.com
linkedin.com/in/jonathancrabbe
jonathancrabbe.github.io
Born Feb. 1996, Belgium

Jonathan Crabbé

I am looking for a research internship where I can put my curiosity, skills and knowledge at the service of expanding our understanding of artificial intelligence. I'd like to achieve this by developing robust methods that have a real-world impact by empowering human beings.

Education

2020-NOW

PhD in Applied Mathematics

University of Cambridge

Synopsis: Is it possible to *explain the predictions* of complex machine-learning models? This crucial question lies at the core of my thesis. My work consists in developing *post-hoc explainability methods* that create an *interface between complex models and human users*. To illustrate, let us consider the example of a medical machine learning model that recommends a treatment for a patient. By using post-hoc explainability, we can answer crucial questions such as "What part of this patient's data motivates the model's recommendation?" or "Are there similar patients previously seen by the model for which this treatment worked?".
Supervisor: Mihaela van der Schaar
Publications: During my first year of PhD, I have published 3 papers introducing new post-hoc explainability methods in NeurIPS and ICML, the two best conferences in my field.

2018-2019

MASt in Applied Mathematics and Theoretical Physics

University of Cambridge

Description: This master is one of the most challenging courses of math and physics in the world. This year of master taught me to formulate rigorous reasonings by using cutting-edge mathematics. To succeed in the bellow relevant courses, I had to master complex mathematical paradigms such as *differential geometry*, *group theory* as well as *functional analysis*. I believe that this formal knowledge perfectly complements the practical skills that I have acquired during my bachelor.
Grade: Distinction (highest grade)
Relevant 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. During this year, I have learned to *adapt* to an environment that is *out of my comfort zone* in a short amount of time. This course was much more theoretical and rigorous than my previous engineering course. Further, it is designed for the best students from the prestigious French preparatory school, a very involved and targeted teaching system. To achieve an excellent grade in this course, I had to quickly *overcome the difficult*

transition between two completely different teaching systems.
Grade: Mention Bien

2014-2017

Bachelor's in engineering
Université Libre de Bruxelles

Description: This bachelor's in engineering provided me with a wide *set of tools* to solve practical problems with a *principled* and *pragmatic approach*. This was the occasion for me to develop a real interest for using *computer science to solve real-world problems*. In addition, I discovered a real passion for *teaching* as a teaching assistant during my third year.
Grade: The Highest distinction, first of my promotion (among 450 students in first year)
Relevant 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%

Experience

2019 – 2020

Research Assistant
Université Libre de Bruxelles

Description: 1 year of full-time research in the department of theoretical and mathematical physics of ULB. I have worked on Quantum Field Theories (QFT) defined on a manifold with boundary. The purpose of my research was to interpret some topological contributions to the partition function of the QFT in terms of the modes spanning the Hilbert space in which the fields are defined. During this year, I was able to clearly identify the modes giving rise to a topological contribution to the partition function. In doing so, I have learned the *autonomy*, *perseverance* and *conviction* required to be a well-rounded researcher. I've also been teaching assistant for the first-year course of physics for pharma students. This was a very insightful challenge as physics is typically not the main interest of these students. I've had to learn to convince an audience to catch their attention. During this year, I understood that I needed to work in a field that has a *real-world impact*.
Transferable skills: autonomy, perseverance, conviction, convince an audience, desire to have a real-world impact

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 more than 80,000 subscribers and this number is quickly increasing. Please visit clipedia.be for more information.
Transferable skills: communication, pedagogy, ability to adapt a speech to a wide audience

FEB 2020 – MAY 2020

Research Intern
University of Cambridge

Description: Part-time research internship in the van der Schaar Lab. This was my first exposure to concrete machine learning research. In a very short amount of time, I had to improve state of the art machine learning methods. More concretely, I have developed a new interpretability method to translate black-box machine learning models into closed-form symbolic expressions. This led to a paper published in NeurIPS 2020 and a PhD offer that I accepted.

Transferable skills: identify possible improvements in existing works, implement those improvements, write a scientific paper, articulate a reasoning to convince other researchers

FEB 2018 – JUN 2018

Research Intern
Imperial College London

Description: 4 months of full-time research in Modified Theories of Gravitation. During this first experience as a researcher, I have learned to deal with the exploration-exploitation trade-off familiar to all researchers. Since I had to produce results by the end of the internship, I had to find the right equilibrium between learning and producing new ideas. My work during this internship focused on a class of quantum field theories, called K-essence theories, that have causal singular behaviours. To make sense of these singularities, I have built an extension (ultraviolet completion) of these theories. To validate my model, I have conducted numerical simulations with Wolfram Mathematica. This led to a scientific report that improved the understanding of K-essence theories.
Transferable skills: formulate a problem, design numerical simulations to verify theoretical claims, manage the time on a project with a clear deadline

Skills

- **Coding**: I have implemented many projects in the form of repositories/packages over the course of my time as a student and a researcher. I have a strong knowledge of Python (including Pytorch, Tensorflow, Scikit-Learn, Pandas, Jupyter and Numpy), C++, Java, HTML and CSS.
- **Modelling**: I have trained machine learning models on various datasets (including time series, tabular data and images) for a wide range of tasks (including classification, regression, clustering and denoising). Furthermore, my physicist background gives me useful knowledge (Fourier, Laplace and stability analysis) in projects that involve dynamical systems, like time series forecasting.
- **Reasoning**: With my MASt in applied mathematics, I have learned to articulate a sound reasoning by leveraging powerful mathematical formalisms such as probability theory, statistics, measure theory, real and complex analysis, group theory, linear algebra and differential geometry.
- **Proving**: My mathematical toolbox allows me to articulate rigorous proofs and arguments. This has been extremely helpful in writing paper and convincing other researchers.
- **Creating**: Beyond what I have learned in lectures, I believe that my various experiences as a researcher (2 internships, 1 assistantship, 1 PhD thesis) taught me to think out-of-the box. Moreover, my unconventional education (both engineering, applied mathematics and theoretical physics) allows me to look at a problem with many different angles.
- **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**: Most of my projects as a researcher were done in collaboration with several co-authors. This taught me the importance of clearly defining the tasks in a group, meeting regularly and keeping an organized workflow (via Github, Slack or other platforms).
- **Working autonomously**: During my 2 research internships, I have worked in almost complete autonomy. This made the first internship both challenging and insightful. My improvements are reflected in the results of my 2nd internship: a paper published and a PhD offer.
- **Supervising**: I have personally supervised 2 research internships (1 Cambridge MPhil student and 1 Imperial MSc student). This taught me to design a clear research agenda and provide guidance in line with this agenda.
- **Managing time**: I believe that my time management skills are well reflected by my publication record (3 papers published in top machine learning conferences over the first year of my PhD)
- **Languages**: French (native), English (IELTS Band 8)

Research interests

- Explainable Artificial Intelligence
- Representation Learning/Understanding
- Time Series Modelling
- Noise/Outlier Detection
- Robust Machine Learning
- Machine Learning for Medicine

I am currently working on my PhD thesis in the van der Schaar lab, a Machine Learning lab from the University of Cambridge led by Mihaela van der Schaar. This lab is among the most prolific machine learning lab in the world. In this stimulating environment, I am learning to become a leading researcher in machine learning.

My research focuses on *Explainable AI* with a special focus on explaining the *latent representations* that are involved in state-of-the-art machine learning models. The purpose of this research agenda is to formalize and implement an *interface between complex state-of-the-art machine learning models and human beings*. This interface would allow human beings to understand how the data is processed by machine-learning models to perform a given task, which is *not yet possible* due to the inherent complexity of those models.

I have a particular focus in making sense of *time series models*: Why do they work? How do they make predictions? How can we improve their *robustness*? I believe that those questions are crucial and very related to my primary Explainable AI focus, since time series data is pervasive in *high-stakes domains such as medicine and finance*. Clearly, we want to understand the models we are dealing with in those settings.

I believe that my *nonconventional education* (bachelor's in engineering followed by a master's in applied mathematics) allows me to look at those problems from *many different angles*. My knowledge of cutting-edge mathematics allows me to formulate and motivate a problem with solid theoretical arguments. My engineering toolbox allows me to pragmatically bring formal ideas to the real world with robust and neat implementations.

Publications

- **NeurIPS 2021**: Crabbé, J., Qian, Z., Imrie, F., & van der Schaar, M. (2021). Explaining Latent Representations with a Corpus of Examples. Advances in Neural Information Processing Systems.
- **ICML 2021**: Crabbé, J., van der Schaar, M. (2021). Explaining Time Series Predictions with Dynamic Masks. In Proceedings of the 38th International Conference on Machine Learning, PMLR 139:2166-2177.
- **NeurIPS 2020**: Crabbé, J., Zame, W. R., Zhang, Y., & van der Schaar, M. (2020). Learning outside the black-box: the pursuit of interpretable models. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), Advances in Neural Information Processing Systems (pp. 17838--17849). Curran Associates, Inc.

Conferences

- **NeurIPS 2021**: spotlight paper (top 3% paper) on explaining latent representations
- **Stanford MedAI Series**: long presentation on Explainable AI for time series
- **OxML 2021**: full summer school attendance
- **ICML 2021**: spotlight paper on explaining time series models
- **NeurIPS 2020**: poster paper on translating black-boxes into closed-form expression

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)

Teaching

- **Research Supervisor**: Supervision of one Cambridge MPhil thesis in machine learning and one research internship in theoretical physics (2 × 4 months)
- **Physics Teaching Assistant**: Responsible for 1st year pharma physics exercise sessions (1 year)
- **Pedagogy project Clipedia**: Creation of videos for a major Belgian pedagogy project (4 years)
- **Engineering Teaching Assistant**: Responsible for weekly Q&A sessions for 1st and 2nd year engineering students (1 year). Subjects: Linear Algebra and Geometry, Analysis, Probability and Statistics, General Physics, Quantum Mechanics, Electricity, Classical Mechanics, Continuum Mechanics.