

# Robin Brochier

## ML Research Engineer

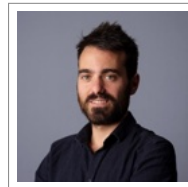
Engineer's degree and Ph.D. in Computer Science

Marseille, France

0033 6 30 41 67 85

contact@robinbrochier.com

www.robinbrochier.com



## Work Experience

- since 2024 **Machine Learning Researcher**, *Novartis*, Basel, Freelance  
Working in the GenAI team on several research projects building LLM agents to tackle electronic health data exploration in clinical trials.
- since 2024 **Founder & Developer**, *CarboTrack*, Marseille, Freelance  
Creation of an agent-based tool suite to help companies to compute their carbon footprint.
- 2024 **Senior Data Scientist - Generative AI**, *Pigment*, Marseille, Freelance  
Development of the assisted modeling feature for *Pigment*, an AI-augmented planning platform built for agility and scale. I coded the full data, LLM (Large Language Model), prompt and RAG (Retrieval Augmented Generation) pipelines in a Python/Flask/FastAPI/OpenAI environment. I also developed several tools to improve the automatic code generation capabilities of the LLMs such as fine-tuning via code grammar parsing with ANTLR and RAG via documentation automatic indexing.  
A video of the result: <https://www.youtube.com/watch?v=J-Qvyo9G8L8>.
- 2022-2023 **Senior Data Scientist - NLP and Graph ML Specialist**, *Farfetch*, Porto  
Machine learning research and development in the *Fashion Knowledge Graph* team at Farfetch, a global e-commerce platform for luxury fashion. I worked on data annotation and ingestion problems with fashion experts. I researched and put in production machine learning algorithms for large scale product classification using CV (Computer Vision), NLP (Natural Language Processing) and GNN (Graph Neural Networks) models (Python, Databricks, Spark, Microsoft Azure, PyTorch, HuggingFace, Neo4j, Stardog, and more).
- 2020-2021 **Postdoctoral Researcher**, *LIS Lab - TALEP Team*, Marseille  
Research in graph machine learning and natural language processing for humanities and social sciences. I worked on combining graph machine learning with large pre-trained Q&A models to help social science researchers to make sense of large corpora. I worked with Python and used the supercomputer *Jean Zay* with Slurm to run multi-GPU neural network training.
- 2016-2020 **Data Scientist**, *Digital Scientific Research Technology*, Lyon  
Research and development for the recommender system of Peerus, a web application dedicated to scientific literature monitoring. I implemented the search engines of the application with Elastic Search and scikit-learn. I implemented and compared via A/B testing various recommender systems. I created an expert retrieval application based on researchers publications and collaborations to help publishers to speed-up the reviewing process.
- 2014-2016 **Data & Embedded Systems Developer**, *Fraunhofer Institut IPK*, Berlin  
Two-years working part-time on several projects: embedded systems programming, web development and data science applied to production management (PHP, Python, C++, Raspberry Pi, Kivy, Scikit-learn).
- 2013-2014 **Web Developer**, *Fraunhofer Institut IPK*, Berlin  
One-year internship designing and building a web application helping firms to safely share their processes together and to reduce the risks in product design (PHP, Symfony, HTML, Javascript).

## Education

- 2017-2020 **Ph.D. in Computer Science**, *ERIC, Université Lyon 2*, Bron  
PhD on representation learning for document networks with application to the scientific literature.
- 2016 **Engineer's Degree**, *CPE Lyon*, Lyon  
Department of electronics, telecommunications and computer science with master's thesis at the TU Berlin on building a robotic simulator for relational reinforcement learning.

## Skills

French	Native speaker.	
English	Good	Cambridge English exam (C1).
German	Good	Gaëthe-Institut exam (B2).
IT	Programming skills	Python, PyTorch, JAX, TensorFlow, Spark, Databricks, C++, Linux, L <sup>A</sup> T <sub>E</sub> X & more.
Communities	<a href="https://github.com/brochier">https://github.com/brochier</a> — <a href="https://stackoverflow.com/users/6084245">https://stackoverflow.com/users/6084245</a> — <a href="https://datascience.stackexchange.com/users/17208/robin">https://datascience.stackexchange.com/users/17208/robin</a>	

## Extra-curricular activities

Hobbies Field Hockey, Cycling, Running, Mountaineering, Gardening, Cooking, Music, Board Games & more

---

## Publications

### Conferences:

- Brochier, R. and Béchet, F. (2021). **Predicting Links on Wikipedia with Anchor Text Information.** *44th International Conference on Research and Development in Information Retrieval, SIGIR 2021.*
- Brochier, R., Guille, A., and Velcin, J. (2020). **Inductive Document Network Embedding with Topic-Word Attention.** *42nd European Conference on IR Research, ECIR 2020.*
- Brochier, R., Guille, A., and Velcin, J. (2019). **Global Vectors for Node Representations.** *In Proceedings of the 2019 World Wide Web Conference. International World Wide Web Conferences Steering Committee.*

### Workshops:

- Brochier, R., Gourru, A., Guille, A., and Velcin, J. (2020). **New datasets and a benchmark of document network embedding methods for scientific expert finding.** *In Bibliometric-enhanced Information Retrieval: 10th International BIR Workshop at ECIR.*
- Brochier, R., Guille, A., and Velcin, J. (2019). **Link Prediction with Mutual Attention for Text-Attributed Networks.** *In Companion of the The Web Conference 2019. International World Wide Web Conferences Steering Committee.*
- Brochier, R., Guille, A., Rothan, B., and Velcin, J. (2018). **Impact of the Query Set on the Evaluation of Expert Finding Systems.** *In BIRNDL 2018 at SIGIR.*

### Thesis :

- Brochier, R. (2020). **Apprentissage de Représentation dans les Réseaux de Documents : Application à la Littérature Scientifique (Representation Learning in Document Networks : Application to the Scientific Literature).** *Université Lumière Lyon 2.*

### Others:

- Brochier, R. (2019). **Representation Learning for Recommender Systems with Application to the Scientific Literature.** *In Companion of the The Web Conference 2019. International World Wide Web Conferences Steering Committee.*
- Brochier, R. (2019). **Apprentissage de Représentation appliqué à la Recommandation pour la Littérature Scientifique.** *In Conférence jointe CORIA-EARIA 2019.*
- Brochier, R., Guille, A., Velcin, J., Rothan, B., and Cioccio, D. (2018). **Peerus Review : Un Outil de Recherche d'Experts Scientifiques. (Peerus Review: a tool for scientific experts finding).** *In EGC 2018.*

---

## Teaching

### University:

- **Java Coding Course (2020 - Aix Marseille Université):** I supervised exercises and tutorials, wrote and corrected exams, simultaneous with face-to-face and remote students. The course focused on object-oriented programming and on the basic libraries for Java.
- **Linear Algebra and Data Analysis (2019 - Université Lyon 2):** I supervised exercises and tutorials, wrote and corrected exams. The course was divided in two parts, first linear algebra with vector spaces, system solving, matrices and diagonalization and then data analysis with straight application of linear algebra using Matlab.

### Mentoring:

- **Data Science Mentor (2021 - OpenClassRooms):** I supported students (in career change) in their data science training.

### Jury:

- **Jury (2024 - Jedha):** I was a jury member for Jedha, a data science bootcamp.