Marco Virgolin

Spadinalaan 213, 1031 KB, Amsterdam
+31 628 735 831
marco.virgolin@gmail.com
marco-virgolin-8815a667
@marcovirgolin
marcovirgolin
https://marcovirgolin.github.io

I am a **senior data scientist** at INGKA, in the **digital ethics and responsible AI** team. I design and develop architectures to make powerful technology such as **large language models** behave in a safe and controllable manner. In the past, I have also worked on **symbolic regression**, **neural architecture search**, and **human-machine interaction**.

Core skills

- Strong analytical thinking & problem solving
- Conceptualization, execution, and supervision of machine learning research projects
- Develop. of libraries & pipelines, benchmarking
- Communication & presentation skills
- Pragmatic, flexible, result-oriented

Experience

MAR 2023 – ONGOING

Senior data scientist / Ingka – IKEA, Amsterdam, NL

I work in the **digital ethics and responsible AI** team. Our team's focus is on making **data** and **AI processes trustworthy** and **accountable**, as well as **designing and training** new valuable, explainable-by-design **AI models**. I design and develop AI components based on **large language models** that are steered to be truthful and avoid hallucinations. I also design and carry out **research** on LLMs.

SEP 2021 – MAR 2023

Researcher (tenure track) / CWI, Amsterdam, NL

I worked on the intersection between **evolutionary optimization** with machine learning, including **deep learning** (transformers, CNNs). I also studied methods to explain black-box ML models, such as **counterfactual explanations**. Besides this, I was involved in education and supervision (M.Sc. and Ph.D. students), as well as **international scientific collaborations**.

JUN 2020 – AUG 2021

Postdoc / TU Chalmers, Gothenburg, SE

I worked on making **natural language processing** more interpretable, and compared with **large language models**. I also worked on making interpretable ML more personalized with **active learning** and **human-machine interaction**.

NOV 2019 – MAR 2020

Project researcher / CWI, Amsterdam, NL

Project on emotion recognition from facial expression for children. Since pediatric data is scarce, I worked on data **augmentation via contrastive learning for deep CNNs**.

AUG 2012 – SEP 2013

Web developer / Promoscience, Padriciano, IT

Part-time job during my M.Sc. studies. Being a small company, I wore many hats: from **front-end** to **back-end web development**, incl. building **REST services** and interfacing with **relational data bases**.

Education

JUN 2020

Ph.D. in Evolutionary ML / TU Delft, Delft, NL + CWI, Amsterdam, NL

Design and application of **information theory-based evolutionary algorithms** for learning interpretable **symbolic regression** models. The project application concerned **pediatric radiotherapy**.

MAR 2015

M.Sc. in Computer Engineering / University of Trieste, Trieste, IT

Graduated **cum laude**. Courses ranging from theory of computability and complexity, to software engineering for web apps, IoT. Thesis on natural language processing via genetic programming, later published as a paper.

Honors

- Won **SIGEVO Best Ph.D. Dissertation** award in 2020, **HUMIES Silver award** in 2021, 2×**Best paper awards**
- Served in the **program committee** of several **conferences** and **workshops**: GECCO, ECML-PKDD, PPSN, Trustworthy and Socially Responsible Machine Learning Workshop @ NeurIPS, Workshop on eXplainable Knowledge Discovery in Data Mining @ ECML-PKDD, and more
- Served as a **reviewer** for several **international peer-reviewed journals**: Machine Learning, IEEE Transactions on Evolutionary Computation, Soft Computing, and more
- Invited in 2022 to be an **evaluation committee member** for the **Dutch Research Council** in the domain *Science*
- Recipient as co-applicant of a 300,000 SEK **grant** by Area of Advance Health Engineering, TU Chalmers 2021
- Recipient of 3 ACM Student **travel grant** during my Ph.D.
- Gave **talks** and **invited lectures** at multiple venues, incl. conferences, University of Amsterdam, TU Delft, MIT

Coding experience

Languages, from proficient to rusty: **Python** (incl. Pandas, Scikit-learn, PyTorch, NumPy, SciPy, Matplotlib, Seaborn, Jupyter Notebooks), **C++** (incl. Boost and SWIG to interface C++ with Python), **C#** (incl. ASP.NET), **Java** (incl. Android development), **Matlab**, **PHP**, **SQL** (MySQL and SQL Server), **Javascript** (incl. jQuery, AngularJS, NodeJS)

Examples of different open-source repos (see https://github.com/marcovirgolin*)*:

- **GP-GOMEA** is a C++ based library which includes several symbolic regression algorithms with a Scikit-learn **Python interface**. These algorithms were found to be among the best performing in SRBench (NeurIPS 2021), a large **benchmarking** platform (which I co-authored and help maintaining).
- **Robust-counterfactuals** is a Python repository to simulate perturbations that may invalidate **counterfactual explanations** (a popular explainable AI method) and includes interfaces to experiment with different counterfactual search algorithms and machine learning models.
- **genepro** is a (documented) Python library that I prepared for TU Delft students for the course *Evolutionary Algorithms* of 2021-2022. It contains examples for **classification**, **regression**, and **reinforcement learning**.

Other info

- I was a **co-organizer** of the Joint Lectures on Evolutionary Algorithms (JoLEA), in particular I prepared and maintain the website, and I set up a MailChimp account for mailing lists and tweets.
- I served in ASTRO (2014-2015), a **volunteering** association for helping with care for hospitalized children.
- An **academic** version of my CV is available on my website.

- In my **free time** I enjoy bouldering.