|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Marco Virgolin | |  |  | | --- | --- | | Spadinalaan 213, 1031 KB, Amsterdam |  | | +31 628 735 831 |  | | [marco.virgolin@gmail.com](mailto:marcovirgolin@gmail.com) |  | | [marco-virgolin-8815a667](https://www.linkedin.com/in/marco-virgolin-8815a667/) |  | | [@marcovirgolin](https://twitter.com/MarcoVirgolin) | Twitter Logo, symbol, meaning, history, PNG | | [marcovirgolin](https://github.com/marcovirgolin) | Github, git, logo, social icon - Free download on Iconfinder | | [https://marcovirgolin.github.io](https://marcovirgolin.github.io/) |  | |

I am a **senior data scientist** at INGKA, IKEA. My focus is on the research of architectures & training methods to make **large language models** behave in a safe and controllable manner. Before, I was a scientific researcher in the fields of **explainable AI**, **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 on making **data** and **AI processes** **trustworthy** and **accountable**, as well as **designing and training** new valuable, explainable-by-design **AI models**.

### My current focus is on the research, design, and development of methods to control **large language models**.

### 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** (2020), **HUMIES Silver award** (2021), 2**Best paper awards**
* Published research papers in **top scientific venues** such as ICML, NeurIPS, GECCO
* 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 to be an **evaluation committee member** for the **Dutch Research Council** in the domain *Science* (2022)
* 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 more to less proficient*: **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) --- *Other tech*: **GCP**, **Slurm**, **Git**, **Conda**, **Linux/bash**

*Examples of different open-source repos from academic work (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](https://cavalab.org/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](https://jolea.project.cwi.nl/) (JoLEA), in particular, I prepared and maintained the website, and 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](https://marcovirgolin.github.io/).
* In my **free time** I enjoy bouldering.