Filippo Ficarra

+41 76 231 56 90 / +39 342 64 75 035 filippoficarra.careers@gmail.com
linkedin.com/in/filippo-ficarra github.com/FilippoFicarra 04.06.2000

Experience

IBM Research Zürich Jan 2025 - Jul 2025

Research Scientist - Intern

- **Structure Elucidation from "Mixture IR Spectra"** | | **Paper**
  - Trained, in collaboration with chemists, a Transformer S2S model to generate SMILES of target molecules from *mixture IR spectra* and the target compound molecular formula.
  - Aligned encoder representations with target molecule spectra to boost accuracy, achieving **∼34% Top-1** (SOTA, outperforming human experts).
- **IBM Granite for Single Atom Catalysts Synthesis** | *DeepSpeed ZeRO-3, LangChain*
  - Built a domain-specific knowledge extraction system, parsing semi-structured SAC literature via LLM agents (LangChain).
  - Instruction Tuning of IBM Granite 3.2 8B with different input feature for SAC synthesis, multitask using different amount of input features and with n-shot for RAG enhancement.
- **Large Scale LLM Training** | *DeepSpeed ZeRO-3*
  - Training setup in multi-node distributed training with DeepSpeed ZeRO-3. Trained 8B to 405B LLMs on Nvidia GH200 in a Slurm cluster on on 2-2048 GPUs.

ETH Zürich Sep 2023 - Sep 2024

Student Teaching Assistant

- **Large Language Models** | | (Mar 2024 - Sep 2024)
  - Taught "Prompting and Calibration" tutorial. Wrote, supervised and corrected the final exam.
- **Information Security Lab** | | (Sep 2023 - Jan 2024)
  - Supervised students during Lab sessions. Tested assignments in advance. Taught tutorials, including the Tamarin tutorial.

ORBYTA TECH Mar 2022 - June 2022

Backend Engineer - Intern | C#, ASP.NET, Entity Framework, MySQL

- Designing and optimizing REST API for a *Business-to-Business* clothing company . Developed the backend in C# with ABP Framework and MySQL database.

JEToP - Student Association Nov 2021 - Aug 2022

Mobile developer | Typescript, React Native

- Front-end development for a podcast application, multimedia player, routing and pages.

Education

ETH Zürich Sep 2022 - Nov 2024

Master in Computer Science | GPA: 5.30/6.00

- **Major** in **Machine Intelligence**.
  Relevant Courses: NLP, **Large Language Models, Computer Vision**, Probabilistic AI.
- **Minor** in **Information Security**.
  Relevant Courses: **Applied Cryptography**, Information Security Lab, Security of Wireless Networks, Network Security.

Politecnico di Torino Sep 2019 - Jul 2022

Bachelor in Computer Engineering | GPA: 110/110 cum laude

- Studies fully funded by **Merit-based Scholarhip**: "Percoso Giovani Talenti".

Research

ETH Zürich Apr 2024 - Oct 2024

Master Thesis - Rycolab | Supervisors: Ryan Cotterell, Alex Warstdat |

- **Title**: A Distributional Perspective on Word Learning in Neural Language Models.
- Research accepted as a long paper at **NAACL 2025** main conference.
- Developed innovative metrics based on LLMs surprisal for quantifying **LLM word learning** trajectories and comparing to **human language acquisition** patterns.
- Trained GPT-2 models from scratch for different amount of text and different linguistic source.
- Findings: LLMs learn differently from humans, relying more on word's frequency in the training set.

ETH Zürich Sep 2023 - Mar 2024

Semester Project - Rycolab | Supervisors: Ryan Cotterell, Alex Warstdat

- **Title:** How multimodality influences word acquisition: do images help language models mimic human behavior?
- Analyzed Age of Acquisition in FLAVA models trained with varying vision-to-text ratios.
- Findings: Higher visual input **negatively** impacts word learning efficiency.

ETH Zürich Dec 2023 - June 2024

Swiss AI initiative – Rycolab, SwissAI, CSCS cluster

- Contributed to the training data team for a large-scale academic multimodal Vision-Video-LM.
- Collaborated on the model architecture team.

Publications

Filippo Ficarra, Ryan Cotterell, and Alex Warstadt. 2025. A Distributional Perspective on Word Learning in Neural Language Models. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 11184–11207, Albuquerque, New Mexico.

Selected Projects

Probabilistic AI Projects | | *Grade: 6/6*

- Gaussian Processes: Pollution prediction with GP regression, Bayesian optimization via LCB acquisition.
- Reinforcement Learning: Actor-Critic algorithm for pendulum swing training.

Computer Vision Projects | | *Grade: 6/6*

- Object detection and tracking: BoW vs. CNN classifier, condensation algorithm for dynamic tracking.
- 3D reconstruction: Structure from Motion using Essential matrix estimation and point triangulation.

Sentiment analysis on Tweets | | *Grade: 5.5/6 | TPUs, Google Cloud*

- Ensemble of finetued *RoBERTa* with sparsity introduced by sparsemax, BERTweet and CLIP text features. Our model achieved **92.62%** of accuracy on the classification task (POS/NEG).

Large Language Models course projects | | *Grade: 6/6*

- Assignment 1: Theoretical aspects of Language modeling. Special focus on RNNs and Transformers.
- Assignment 2: LLMs finetuning approaches, effect of prompting for complex tasks and attacks against watermarking and "secret" text embeddings.

Technical Skills

Programming Languages: Python, JavaScript, C#, Java, C, SQL, Bash.
Technologies: Docker, Git, IBM Blue Vela, IBM Cloud, LaTeX, Linux, PyTorch.
Interests: Large Language Models, AI and LLM Vulnerabilities, LLM Hallucinations, Cryptography, ML optimisation.