

Angelika Romanou

Rue Saint-Martin 25, Lausanne 1003, Switzerland

18/11/1992

+41 78 251 67 38

angelika.romanou@epfl.ch

[Website](#) ♦ [Github](#) ♦ [Google Scholar](#)

RESEARCH INTERESTS

Research areas: Post-Training/Alignment, Evaluation (Reasoning/Multilingual), Robustness & Prompt Sensitivity, Long-Context Reasoning

My research focuses on building reliable AI systems by closing the loop between evaluation and post-training. I develop evaluation pipelines that target complex reasoning, long-context understanding, and prompt sensitivity, exposing where models fail. These insights drive the design of post-training interventions that directly address identified weaknesses, improve robustness and transfer across languages, with demonstrated gains at scale.

PROFESSIONAL & RESEARCH EXPERIENCE

- 08/2025 - Present **Research Scientist Intern**, Meta FAIR, New York City / USA
- Research on evaluating and improving robustness and prompt sensitivity in LLMs.
- 09/2020 - Present **PhD Candidate**, EPFL - [NLP lab](#), [LSIR lab](#), Lausanne / Switzerland
- Research on complex reasoning & cross-lingual transfer in Large Language Models (LLMs).
 - Teaching Assistant for [Modern NLP](#) & [DIS](#) & [DB-Systems](#) EPFL master courses.
 - Supervisor for students' semester projects and master thesis.
- 06/2019 - 08/2020 **Machine Learning Research Scientist**, EPFL - [lsir.epfl.ch](#), Lausanne / Switzerland
- Research on sequence modeling and natural language processing.
 - Software development for machine learning projects.
- 03/2018 – 05/2019 **Machine Learning Software Engineer**, Beat - [thebeat.co](#), Athens / Greece
- Design, implement, and deploy machine learning features for Beat App or internal tools.
 - Develop reliable and scalable services using micro-services architecture.
- 02/2017 – 02/2018 **Big Data Software Engineer**, Beat - [thebeat.co](#), Athens / Greece
- Build data streaming pipelines in order to serve machine learning applications/services.
 - Develop components that analyze, process, and react to operational feeds in near real-time, optimizing driver allocation, optimizing service pricing, and preventing fraudulent activities.
 - Work with other engineering teams to translate complex models and algorithms into production-grade software systems.
- 03/2014 - 01/2017 **Data Scientist**, Intelen Inc - [intelen.com](#), Athens / Greece
- Demand forecasting and energy grid optimization.
 - Perform data-driven business consulting on clients.
 - Research and development on data-oriented functionalities and services.

EDUCATION

- 09/2020 - Present **Ecole Polytechnique Fédérale de Lausanne (EPFL)**, Lausanne, Switzerland
- **PhD degree:** Computer Science
 - Thesis direction: “*Evaluation of Generative AI Agents in Multilingual and Multimodal Contexts*”
- 09/2016 - 07/2018 **Athens University of Economics and Business (AUEB)**, Athens, Greece
- **MSc degree:** Data Science (2-year degree, 78 ECTS)
 - Thesis: “*Named Entity Recognizer for the Greek Language using Deep Learning Architectures.*”
- 09/2010 - 07/2015 **Athens University of Economics and Business (AUEB)**, Athens, Greece
- **BSc degree:** Department of Management Science and Technology (4-year degree, 240 ECTS)
 - Grade: 8.04/10 “Very Good”
 - Major: Information Systems
 - Thesis: “*An Overview of Information Systems Architectures and their Impact on Strategic Management*”

NOTABLE PUBLICATIONS

- **Brittlebench: Quantifying LLM robustness via prompt sensitivity** (Feb 2026)
Romanou A, Ibrahim M, Bosselut A, Williams A et al., *Arxiv 2026* [[Paper](#)]
- **Apertus: Democratizing Open and Compliant LLMs for Global Language Environments** (Sept.2025)
Apertus Team, *Arxiv 2026* [[Paper](#)]
- **Measuring what Matters: Construct Validity in Large Language Model Benchmarks** (July 2025)
Bean A, Kearns R, Romanou A, Mahdi A et al., *NeurIPS 2025* [[Paper](#)]
- **PERK: Long-Context Reasoning as Parameter-Efficient Test-Time Learning** (July 2025)
Chen Z, Romanou A, Weiss G, Bosselut A, *Arxiv 2025 (Under Submission)* [[Paper](#)]
- **Global-MMLU: Understanding and addressing cultural and linguistic biases in multilingual evaluation** (Nov. 2024) Singh S, Romanou A, Fourrier C, Choshen L, Ippolito D, Hooker S, et al., *ACL 2025* [[Paper](#)]
- **INCLUDE: Evaluating Multilingual Language Understanding with Regional Knowledge** (Sept. 2024)
Romanou A, Foroutan N, Hooker S, Bosselut A, et al. *ICLR 2025 (Spotlight)* [[Paper](#)]
- **Meditron-70b: Scaling medical pretraining for large language models** (Nov. 2023)
Chen Z, Cano AH, Romanou A, Bosselut A, et al. *arXiv preprint* [[Paper](#)]
- **CRAB: Assessing the Strength of Causal Relationships Between Real-world Events** (Jun. 2023)
Romanou A, Montariol S, Paul D, Laugier L, Aberer K, Bosselut A, *EMNLP 2023* [[Paper](#)]
- **Interpreting Language Models Through Knowledge Graph Extraction** (Dec. 2021)
Swamy V, Romanou A, Jaggi M, *eXplainable AI Workshop @NeurIPS2021* [[Paper](#)]
- more on my [Google Scholar](#)

GRANTS & AWARDS

- **Meta Large Language Models Evaluations** [Research Grant](#) (2024)

RECENT INVITED TALKS

- **Panel Talk at AMLD 2026**, Lausanne / Switzerland (in-person)
- **Invited Talk at Cohere Connect conference 2025**, (remote)
- **Invited Talk at AMLD 2024**, Lausanne / Switzerland (in-person)
- **Causality Workshop at AAAI 2024**, Vancouver / Canada (in-person)
- **Poster presentation at EMNLP 2023**, Singapore / Singapore (in-person)
- **iNET Oxford Seminars (Nov 2023)**, Oxford / United Kingdom (in-person)

PROGRAMMING SKILLS

- **Languages:** Python, Scala, SQL
- **ML & LLM Frameworks:** PyTorch, HuggingFace Transformers/Datasets, PEFT/LoRA, scikit-learn
- **Experimentation & Evaluation:** Weights & Biases, TensorBoard
- **Data & Distributed Systems:** Apache Spark, Apache Flink, Apache Hive
- **Infrastructure & Tools:** Git, Docker, Slurm

LANGUAGES

- **English** (fluent), **Greek** (native), **French** (basic)