Radu Stefan Popa

radupopa028@gmail.com | radu028.github.io | Google Scholar | GitHub

First author or co-first author of three papers in my first seven months of research

Education

University of Bucharest — Bucharest, Romania

B.Sc. in Computer Science, Faculty of Mathematics and Computer Science — *Oct. 2024 – July 2027 (expected)*

- Ranked 1st of 282 students in Year 2, with 10/10 in every course. Year 1: 10th of 289 (9.40/10).
- Coursework: Artificial Intelligence, Probability and Statistics, Linear Algebra, Advanced Algorithms (all 10/10); in progress: Computer Vision, Reinforcement Learning.

Publications

[1] **Radu Stefan Popa**, Adina Camelia Bleotu, Sergiu Nisioi. *Romanian Small Language Models Trained on Native Child-Targeted Media.* NeurIPS 2026 Workshop on Linguistic Principles for Foundation Models (LP4FM). Accepted poster.

Preprints

[2] **Radu Stefan Popa**, Valentin Feroiu, Adina Camelia Bleotu, Sergiu Nisioi. *BLiMP-RO: A Benchmark of Linguistic Minimal Pairs for Romanian.* arXiv preprint, 2026. arXiv:2611.17836.

Research Experience

Romanian Small Language Models — Aug. 2026 – Present

Project lead, advised by Sergiu Nisioi — *University of Bucharest*

- Built a corpus of 10M words from 2,100 hours of Romanian children's video, ten times larger than any Romanian dataset of its kind before it, and trained language models on it from scratch.
- Showed that, at equal size, real speech addressed to young children teaches models almost no grammar (near chance), far less than books, film subtitles, or even children's video.
- Co-built BLiMP-RO, the first Romanian benchmark of minimal pairs written by linguists and tested on native speakers: 50 phenomena, over 6,400 pairs, over 40,000 human judgments.

PaMuk-Bench: Multilingual Vision-Language Benchmark — May 2026 – Present

Co-first author, advised by Radu Tudor Ionescu — *University of Bucharest*

- Co-led the first parallel benchmark to cover five vision-language tasks, including captioning, image generation, and retrieval, in seven languages: 71,400 texts, 13 researchers from six countries.
- Led the captioning study on all four models, three chosen by me: fine-tuning on our data improved each one on every metric in all six languages, even the two pretrained on 39 and 100 languages.
- Devised and built the method that taught three image generators six more languages: it distills each model's own English behavior into its text encoder, closing up to 92% of the gap to English.

Teaching

Undergraduate Tutor, my own weekly classes in the official timetable Oct. 2025 – Present
Selected by the course faculty after earning 10/10 in each course *University of Bucharest*

- Algorithmic Programming (Fall 2025): weekly sessions on DP, backtracking, divide and conquer.
- Data Structures (Spring 2026): weekly C++ sessions on heaps, hashing, AVL and red-black trees.
- Fundamental Algorithms (Fall 2026): weekly graph sessions, from BFS to Edmonds-Karp.

Work Experience

Software Engineer, full-time alongside my studies and research Mar. 2026 – Present
Liamis & HEB International *Bucharest, Romania*

- Develop the custom operating environment that cataract surgeons work in during surgery, where many C# and C++ services run in sync with the surgical hardware.
- Fix race conditions and other bugs in its multithreaded services.
- Built an LLM assistant over 100k+ log entries and Jira incidents for bug triage.

Honors and Awards

PPC Romania research scholarship, 1 of 4 open to all final-year BSc and MSc students 2026 – 2027

University of Bucharest merit scholarship, awarded to the top few percent of my year 2025 – 2027

Skills

ML: PyTorch, Hugging Face (Transformers, Diffusers, PEFT, Accelerate), LoRA, DPO, distillation
Programming: Python, C++, C#, TypeScript
Languages: Romanian (native), English (C2, certified), French (B2, certified)