Razvan Florian Vasile

Cesena, Italy · florian.rvasile@gmail.com · Github · Website · LinkedIn

Education

Master's Degree in Computer Science 2022 - 2025
University of Bologna (Unibo), Bologna, Italy — Awarded Mar 2026

- Completed a Master's degree in Computer Science with a focus on Machine Learning.
- Final grade: 110/110 with honours.

Bachelor's Degree in Computer Science and Mathematics 2014 - 2018
University of Manchester (UoM), Manchester, UK

- Finished a joint honours degree in Computer Science and Mathematics.
- Final grade: First Class Honours.
- Awarded the Manchester Scholarship on entry (A*AA equivalent).

Employment and Experience

Software Engineer 2018 - 2020 / 2020 - 2022
Onit Group / Agronica, Cesena, Italy — European Digital Agriculture 4.0
Python · SQL · NoSQL · IoT · Full-stack developer

- Developed Raspberry Pi IoT integrations in Python for warehouse equipment control.
- Full-stack development across application services supporting SQL/NoSQL databases.

Industrial Placement Year 2016 - 2017
Diamond Light Source, UK — National Synchrotron facility
Python · MATLAB · Scientific software · Accelerator controls

- Project involved a library for accelerator controls, providing a unified interface to accelerator components across previously fragmented MATLAB/Python code.
- Allowed users to perform data collection and experimentation in Python, including steering magnet positions to adjust beam trajectory across different setups.

Peer Assisted Study Sessions (PASS) Leader 2015 - 2016
University of Manchester Teaching Assistant, UK

- Assisted first-year students with programming concepts and coursework problems.
- Organised review sessions and prepared practice materials for exams.

Hornet Representative 2014 - 2016 / 2017 - 2018
University of Manchester IT Services, UK

- Resolved wired and wireless connectivity issues across university halls of residence.
- Visited residents on site to diagnose and resolve network connectivity problems.

Professional Activities

Volunteer October 2025
European Conference on Artificial Intelligence (ECAI), Bologna, Italy

- Supported attendee coordination and speaker logistics at ECAI 2025.

Volunteer March 2026
S.P.R.I.Te. Association, University of Bologna, Cesena, Italy

- Helped organise a University of Bologna event on free and open-source software.

References and more information available upon request.

PROJECTS

Linalg-Zero: Distilling Neurosymbolic Reasoning for Linear Algebra in Small Language Models 2025 – 2026
GitHub · Report · Poster · Video · HF Space

- Built LinAlg-Zero, an end-to-end neurosymbolic pipeline that trains a small tool-using LLM to solve linear algebra through verifiable tool calls under a strict format protocol.
- Developed an entropy-controlled synthetic problem generator and a distillation workflow for multi-turn executable traces, followed by LoRA-based SFT and GSPO RL.
- Achieved 92.63% verifier-checked correctness, 90.26% optimal-trajectory accuracy, and 96.66% format accuracy on a held-out test set, using approximately $75 of cloud compute.

Distributed and Parallel Techniques for Deep Neural Networks 2025
GitHub · Report · Poster

- Implemented GPU training experiments with cuDNN and cuBLAS, covering forward and backward passes, convolution primitives, manual kernels, and linear layers.
- Built reproducible distributed-training examples with PyTorch DDP and Docker to simulate multi-GPU and multi-node workflows on a single NVIDIA GPU.
- Conducted a structured survey of distributed training systems and GPU libraries, comparing parallelization strategies, communication patterns, and implementation trade-offs.

Git Inspector: Querying GitHub Repositories with Local LLMs 2025
GitHub · Demo · Report · Poster · Website

- Built Git Inspector, a Scala-based RAG system for querying GitHub repositories using semantic code retrieval with Qdrant and local open-source LLMs via Ollama.
- Implemented repository cloning, indexing, semantic chunking, and a REST API with Gradio and Scala.js interfaces for code and documentation queries.
- Validated the system with acceptance tests, traceability documentation, and a usability study that achieved an average SUS score of 85% across five users.

Visual QA: Using Generative Models on Classification Tasks 2024
GitHub · Report · Poster · W&B

- Adapted BLIP-2 from visual question answering to classification by adding a lightweight classification head to the multimodal Q-Former component.
- Fine-tuned the model with LoRA (PEFT), achieving 91% accuracy on Easy-VQA and 78% on DAQUAR while outperforming generative baselines.

QuestLlama: An Autonomous Agent in Minecraft 2024
GitHub · Presentation · Demo

- Extended QuestLlama, a Voyager-based autonomous Minecraft agent that uses open-source LLMs, retrieval, and code generation to complete in-game tasks.
- Integrated local-model backends via Ollama and OpenAI-compatible APIs to support experimentation without reliance on closed-source providers.

Bachelor's Thesis: Sudoku Solver in C++ 2017 – 2018
GitHub · Presentation · Report

- Implemented a CDCL SAT solver in C++ with backjumping, conflict-driven clause learning, and decision heuristics, encoding Sudoku puzzles into SAT.
- Developed an interactive visualizer to inspect implication graphs, unit propagations, conflicts, and clause learning during the solving process.
- Evaluated difficult Sudoku and SAT instances through VSIDS/LIS comparisons, flamegraph profiling, and minimal/extended encoding analysis.

Librarian Assistant 2024
GitHub · Video · Report · Poster

- Built Librarian, a local question-answering assistant for personal PDF libraries using open-source LLMs and retrieval-augmented generation over book contents.
- Implemented a Thought–Action–Observation agent with vector retrieval, citation-grounded answers, and both local/serverless inference via Ollama and HuggingFace.

Docker UI 2024
GitHub · Report · Figma · Docker Hub

- Built a web interface for managing Docker Compose environments from the browser by interacting with the Docker daemon through an HTTP API.
- Implemented Docker Compose orchestration and log inspection in the browser, and refined the interface via Figma prototypes and questionnaire-based usability feedback.

Technical Skills

Languages: Python, C++, MATLAB, JavaScript/TypeScript.

Experienced with: Distillation, SFT, RLVR/GSPO, Tool-Using Agents, RAG, Embeddings & Vector Search, LoRA/PEFT, Multimodal Learning, Distributed Training, GPU Programming.

Mathematical background: Linear Algebra, Matrix Analysis, Differential Equations, Convex Optimization, Real Analysis.

Interests

Scientific computing, optimization, reproducible research, open-source software, Emacs.