# Riccardo Campanella

401 Jonkvrouw Sanderijndreef, Utrecht, Netherlands
+39 3498783643 riccardocampanella.ing@gmail.com riccardo-campanella Code Portfolio

## Summary

Pioneering AI researcher with a software engineering background, driven by a strong belief in AI as a force for positive change. Experienced in advancing LLM chain-of-thought reasoning through distillation and test-time compute, with a research focus on belief updates and interpretability.

## Skills

**LLMs:** PyTorch TensorFlow Keras vLLM TRL Unsloth Llama-Factory THF Bitsandbytes

**GenAI (Agentic Workflows & Diffusion Models):** OpenAI Gymnasium N8N LangChain ComfyUI

**Computer Vision:** OpenCV MMCV

**MLOps:** ZenML Weights & Biases TensorBoard Azure Cloud Docker CI/CD Pipelines

**Backend & Databases:** Python (Django) REST APIs (Swagger) PostgreSQL

**Frontend:** JavaScript Angular Flutter

**Blockchain:** R3 Corda Java Kotlin Scala

**Languages:** English (C1) Italian (Native)

## Experience

**Chain-of-Thought Data Curator** **US, Remote**
*Micro1* *September 2025 - Ongoing*

- Develop and curate gold-standard Chain-of-Thought (CoT) datasets across diverse reasoning-heavy tasks.
- Design clear, scalable rubrics and instructions to evaluate and annotate multi-step reasoning processes.
- Write precise, well-structured CoT responses that demonstrate high-level generalist reasoning.
- Critically assess logical flow, correctness, and justification within reasoning chains, ensuring rigor and fidelity.
- Identify and document common model failure types, such as hallucination, shortcut reasoning, and unsupported leaps.
- Collaborate with AI trainers, model evaluators, and RLHF annotators to refine CoT benchmarks.
- Stress-test the depth and reliability of LLM reasoning across varied benchmarks.

**AI Research Intern** **Delft, Netherlands**
*Netherlands Organisation for Applied Scientific Research (TNO)* *November 2024 – June 2025*

- Evaluated SoTA MLLMs on Visual Question Answering benchmarks and Dutch house facades images to detect architectural elements using Azure and L4 GPU servers.
- Crafted multimodal dataset using Cyclomedia and Google Street view APIs to ensure high training data quality.
- Fine-tuning LLMs using LoRA on GPT-4o generated synthetic captions via OpenAI-API and 3D scene graphs.
- Distillating and Scaling reasoning rationales in student models via Monte Carlo Test-time Search.
- Grounding Chain-of-Thought Reasoning by reducing hallucinations and enforcing LLM's reliance on spatial cues.

**Software Engineer** **Dubai, United Arab Emirates**
*Hotdesk* *January 2023 – July 2023*

- Worked with Python using Django framework, REST APIs and Swagger to enhance existing systems.
- Worked with Docker and Pipelines to deploy company app.
- Tested features by providing automated test cases.
- Produced Design Documents for features.

**Software Engineer - Innovation Advanced Technology** **Cosenza, Italy**
*NTT DATA* *September 2022 – March 2023*

- Worked with Python using OpenCV library to find the best algorithm to perform an Image Recognition task.

- Worked with Blockchain DLT for Enterprise to update a CorDapp using Java for the European Banking System.
- Updated a Spring MVC App working as Java backend developer.
- Fundamentals of Agile & DevOps, Microservices: Architectures & Frameworks and JavaEE programming.

**Software Engineering Intern** **Cosenza, Italy**

*Info Edge Technology* *July 2022 - August 2022*

- Worked with Scala to convert a C++ application used in Distributed Systems.
- Fundamentals of Scala.

## Research & Initiatives

**Director** **Utrecht, Netherlands**

*AI Safety Utrecht* *November 2025 - Ongoing*

- Lead the operational implementation of the initiative and oversaw partnership scouting and development.

**Independent AI Safety Research** **Remote**

*AI Safety Camp* *January 2025 - June 2025*

- Co-authored paper about Bayesian Reasoning Elicitation in GPT-2 and SoTA LLM using prompt engineering. (link)
- Proposed Mechanistic interpretability framework to monitor the Chain-of-Thoughs in intelligible and unfaithful CoT models to prevent deceptive alignment and scheming.

**Masterclass: Methods for Spatially-Extended Neurobiological Networks** **Utrecht, Netherlands**

*Utrecht University* *May 2024*

- Implemented tutorials on numerical and analytical methods for spatially-extended neurobiological networks under Dr. Daniele Avitabile at the Centre for Complex Systems Studies.

**AI Safety Fundamentals Course** **Utrecht, Netherlands**

*Effective Altruism Utrecht* *September 2023 - December 2023*

- Demonstrated goal misgeneralization using a Deep Q-Learning Network (DQN) in a reinforcement learning task.

## Education

**MSc. Artificial Intelligence** **Utrecht, Netherlands**

*Utrecht University* *September 2023 – August 2025*

**Intelligent Agents**

- Developed a Goal-Based Agent integrating LLMs and Ontology for Fake News Detection using llama-3.2 via groq API.
- Implemented a custom Ontology for fake news detection using python and Protege.

**Natural Language Processing**

- Estimated N-gram models of different orders from the Treebank corpus to calculate sentence probabilities and evaluate perplexity. Implemented the Viterbi algorithm for sequence modeling and the CKY algorithm for parsing.
- Developed two LSTM-based UPOS taggers trained on nltk data—one using randomly initialized embeddings and another leveraging pre-trained GloVe embeddings. Utilized BERT's contextualized word embeddings for word sense disambiguation.
- Designed and implemented probing techniques for masked language modeling in BERT to analyze language model outputs and assess alignment with human linguistic knowledge.

**Reinforcement Learning**

- Implemented two reinforcement learning agents using semi-gradient SARSA and Q-learning algorithms with a linear approximation function to solve an episodic MDP task.
- Developed an epsilon-greedy reinforcement learning algorithm to solve a k-armed bandit problem in a custom environment (based on the OpenAI Gymnasium framework), evaluated on a 10-armed testbed.
- Explored fundamental reinforcement learning algorithms such as Q-learning, Deep Q-Networks (DQN), and Policy Gradient methods to solve decision-making problems.

**Computer Vision**

- Implemented real-time Geometric camera calibration using OpenCV, voxel-based 3D scene reconstruction using KNNs, color-based voxel labeling using GMMs.

- Trained and tested LeNet-5 and its variants (CNN).
- Developed a two-stream CNN for action recognition focusing on transfer learning—integrating optical flow and combining CNN outputs.

**Machine Learning and Deep Learning**

- Classified magnetoencephalography (MEG) data to infer brain states using CNNs, RNNs, and Transformers.
- Digit classification on the MNIST dataset using logistic regression and support vector machines.
- Implemented SVM, Decision Trees, Random Forests and deep learning architectures (CNNs, RNNs, Transformers).

**Machine Learning for Computer Vision & Natural Language**

- Trained embeddings using GloVe for word representation and built a recurrent neural network–based sequence model.
- Implemented deep convolutional neural networks (DNNs) to detect handwritten numbers and objects using Keras.
- Implemented a recommendation dialog system for text classification using supervised machine learning algorithms.

**BSc. Computer Engineering** **Cosenza, Italy**
*University of Calabria* *September 2020 - June 2022*

- Computational Logic: Created custom circuits (Adder, Multiplier) using Xilinx Vivado and VHDL.
- SWE: Developed a web-app using Java (IntelliJ), PostgreSQL for the backend, and Flutter for the frontend. Developed using OOP, Agile Framework and Design Patterns.
- IoT: OS, Telecom Networks
- Math & Physics: Linear Algebra, Calculus (I, II), Physics I, Control Systems (MATLAB)

## Hackathons

**AI Agent for Financial Trend Forecasting (AI & Finance)** **Amsterdam**
*Bunq* *2-3 May 2025*

- Popular Voted project on implementing Agentic LLM for trend prediction of salary, rent and subscriptions using logistic regression and LLamA-3.2 via Nvidia's API.

**Dutch AI Safety Retreat (AI Alignment)** **Arnhem, Netherlands**
*Effective Altruism Netherland* *13-15 December 2024*

- Participated in technical AI Safety workshops on Mechanistic interpretability, Deep fakes and field building
- Proposed Adversarial attack based on LLC loss for the final hackathon workshop.

**AI 4 Life Sciences (AI & Hydrology)** **Remote**
*University of Vienna* *July 2024 – September 2024*

- Identified and ranked exogenous variables for forecasting the GRACE time series (groundwater data) by evaluating variable predictive power.

**InnovAId (AI & Healthcare)** **Utrecht, Netherlands**
*Utrecht Medical Center* *November 2023*

- Proposed NLP-based approaches to solve a predictive troubleshooting problem.
- Implemented BoW, Google Word2Vec (SONAR-combined dataset), and Google Transformer BERT to recommend preemptive actions for infusion pumps.