RICCARDO CAMPANELLA

401 Jonkvrouw Sanderijndreef, Utrecht, Netherlands

📞 +39 3498783643 ✉ riccardocampanella.ing@gmail.com 🔗 riccardo-campanella ⭘ Code ⭘ Portfolio

Experience

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

- 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   March 2023 – August 2023
Hotdesk   *Dubai, United Arab Emirates*

- 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   September 2022 – March 2023
NTT DATA   *Cosenza, Italy*

- 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   July 2022 - August 2022
Info Edge Technology   *Cosenza, Italy*

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

Education

Master's degree in Artificial Intelligence   September 2023 – August 2025
Utrecht University   *Utrecht, Netherlands*

Fairness and Explainability

- Trained a Logistic Regression classifier on the COMPAS dataset to measure (Intersectional) Fairness using pre-processing methods such as Reweighing and post-processing techniques like Equalized Odds.
- Applied Logistic Regression on the ChaLearn LAP-FI dataset to analyze Feature Importance and Feature Effect. Trained a Multi-layer Perceptron and interpreted its decisions using model-agnostic methods such as PDP, ICE, and PFI.
- Trained Bayesian Networks, RuleFit, EBM, and Neural Networks on the US-130 Diabetes dataset to identify key factors contributing to patient readmission and assess feature influence on readmission risk.

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.

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.

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.

Philosophy of AI

- Theoretical LLMs' internal testing using Mechanistic Reasoning to assess intelligence and commonsense reasoning.
- Analysed Commonsense Reasoning through Piaget and System1-system2 theories of mind.
- Ethical Analysis of using consciousness theories, strong vs. weak AI and AGI emplyoment.

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.

Causal Inference

- Estimated the causal effect of patient drug dosage on recovery outcomes using observational and experimental data within a structural causal model using .
- Implemented the PC algorithm to uncover causal relationships in the sachs2005_combined dataset.

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 and 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.

Data Mining

- Implemented Bagging and Random Forest to classif bugs on the Eclipse-bug dataset.
- Implemened Multinomial naive Bayes and Logistic regression with Lasso penalty for Detection of Opinion Spam from on the Fake review dataset from Mechanical Turk.

Bachelor's degree in Computer Engineering   Graduation date - June 2022
University of Calabria   *Cosenza, Italy*

- Created custom circuits using Xilinx Vivado and VHDL.
- Developed a web-app using Java (IntelliJ), PostgreSQL for the backend, and Flutter for the frontend.
- Graduate Thesis: *"Deep Neural"* – Focused on deep neural networks training and evaluation strategies.

Extracurricular Activities

Independent AI Safety Research   January 2025 - June 2025
AI Safety Camp   *Remote*
- Proposed Mechanistic interpretability framework to monitor the Chain-of-Thoughs in intelligible and unfaithful CoT models to prevent deceptive alignment and scheming.
- Co-authored paper about Bayesian Reasoning Elicitation in GPT-2 and SoTA LLM using prompt engineering. (link)

Masterclass: Methods for Spatially-Extended Neurobiological Networks   May 2024
Utrecht University   *Utrecht, Netherlands*
- 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   September 2023 - December 2023
Effective Altruism Utrecht   *Utrecht, Netherlands*
- Demonstrated goal misgeneralization using a Deep Q-Learning Network (DQN) in a reinforcement learning task.

Mathematics for Machine Learning Specialization   2023
Imperial College London   *Coursera*

Challenges

Bunq Hackathon (AI and Finance)   **2-3 May 2025**
Bunq   *Amsterdam*
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)   **13-15 December 2024**
Effective Altruism Netherland   *Arnhem, Netherlands*
- 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 for Life Sciences Hackathon (AI and Hydrology)   **July 2024 – September 2024**
University of Vienna   *Remote*
- Identified and ranked exogenous variables for forecasting the GRACE time series (groundwater data) by evaluating variable predictive power.

InnovAId Hackathon (AI and Healthcare)   **November 2023**
Utrecht Medical Center   *Utrecht, Netherlands*
- 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.

Technical Skills

AI/ML Specializations

- **Multimodal AI:** Multimodal Large Language Models (MLLMs), Vision-Language Models, Chain-of-Thought (CoT) Reasoning, Spatial Reasoning, Visual Question Answering.
- **LLM Technologies:** Large Language Model Fine-tuning, LoRA, Prompt Engineering, OpenAI API Integration, Hallucination Reduction, Zero-Shot Detection.
- **Advanced ML:** Few-Shot Learning, Domain Adaptation, Reinforcement Learning, Causal Inference, Synthetic Data Generation, Test-Time Computation.
- **AI Safety & Research:** Mechanistic Interpretability, AI Safety & Alignment, Model Evaluation Frameworks, Scaling Laws.

Frameworks & Libraries

- **Deep Learning:** PyTorch, TensorFlow, Transformers, HuggingFace, OpenAI Gymnasium, Scikit-learn.
- **Computer Vision:** OpenCV, 3D Scene Understanding, geometric calibration, voxel-based reconstruction.
- **MLOps:** Weights & Biases, TensorBoard, Azure ML, experiment tracking, model monitoring.

Development & Infrastructure

- **Cloud & DevOps:** Azure Cloud, Docker, CI/CD pipelines, containerization, scalable ML deployment.
- **Backend:** Python (Django), Java (Spring), REST APIs, microservices architecture.
- **Frontend:** Angular, Flutter, responsive web applications.
- **Blockchain:** R3 Corda, Java, Kotlin, Scala, distributed ledger technology.

Languages

- **English:** C1 (Professional proficiency)
- **Italian:** Native speaker