



# FEDERICO BRANCASI

Machine Learning Researcher & Certified AWS Cloud Architect

 federicobrancasi  
 federicobrancasi@gmail.com  
 (+41) 76 623 52 77

## EXPERIENCE

Machine Learning Researcher

📅 August 2025 – February 2026

CERN (European Organization for Nuclear Research)

- Investigating **novel approaches to model compression**, integrating quantization, pruning, and decomposition techniques for optimal performance-efficiency trade-offs.

Machine Learning Researcher

📅 January 2025 – July 2025

ETH Zurich (Federal Institute of Technology Zurich)

- Developed **novel quantization techniques** for deploying large transformer models on resource-constrained embedded systems, optimizing performance on devices without floating-point units (FPUs).
- Built a custom backend for AMD's Brevitas (PyTorch Quantization Framework) to enable **automated neural network quantization and deployment on low-power embedded processors**, bridging the gap between simulated and hardware-deployed models.
- Conducted comprehensive benchmarking of different quantization schemes, with particular attention to non-linear operator quantization and automatic network optimization for embedded deployments.

NLP (Natural Language Processing) Researcher

📅 September 2024 – June 2025

University of Trento & European Commission

- Designed and implemented a Docker-based evaluation framework processing **250,000+ scientific papers to benchmark 20 keyword extraction algorithms**, focusing on scalability and reproducibility.
- Leading the development of a systematic methodology that will directly support the HUMAINT team in implementing the EU's Digital Services Act (DSA), impacting content moderation across European digital platforms.

Research Assistant

📅 February 2023 – July 2024

University of Trento

- Modernized Computational Logic course materials by integrating Knowledge Graph methodologies and practical examples, achieving a **30% increase in student pass rates** and improved course engagement.
- Spearheaded 70% of the development of a **new computational logic textbook**, managing technical content creation and producing advanced visualizations and diagrams using LaTeX and TikZ.

## EDUCATION

Double M.Sc. in Computer Science

📅 September 2023 – December 2025

University of Trento & Eotvos Lorand University of Budapest

- Selected as one of the **200 Students in Europe** to be admitted into this program co-funded by the European Union.
- Key role as a **Student Ambassador**: Promoted program benefits and organised events to engage prospective students.
- Current GPA 29/30, Expected Graduation Grade 110/110. Master Thesis on ML Quantization developed at **ETH Zurich**.
- Main Courses**: Deep Learning, Advanced Deep Learning, Machine Learning, Statistical Learning, Statistical Methods, Cyber Security Risk Assessment, Distributed Systems, Business Intelligence, Cloud Systems, DevOps Engineering.

Summer School on Smart Cities

📅 August 2024

Aalto University of Helsinki

- Explored the **intersection of entrepreneurship and smart cities**, studying platform economies and digital innovation.

Postgraduate Diploma

📅 September 2023 – June 2024

Trento School of Innovation

- Evening program** for select students, focused on soft skills led by industry professionals and seasoned researchers.
- Main Courses**: Data Governance, Challenge Action Plan, Emotional Intelligence, Collective Creativity in Organizations.















B.Sc. in Computer Science

📅 September 2020 – September 2023

University of Trento

- Main Courses**: Machine Learning, Operating Systems, Data Structures and Algorithms Design, Object-Oriented Design, Networking, Software Engineering, Databases Systems, Formal Languages and Compilers, Computer Architectures.

# EXTRACURRICULAR

- **Anthropic's Claude 4 Hackathon London Winner**   
Selected as winner at Anthropic's inaugural student hackathon at the London School of Economics and Political Science (LSE), building an intelligent security system with natural language control. Developed a full-stack application combining real-time video monitoring with Claude's conversational AI capabilities in just an hour.
- **AI Hackathon Munich Winner (Organized by AWS, OpenAI, Lovable, n8n, ElevenLabs, TUM.ai)**   
Won the prototyping sprint at CDTM Munich, creating a complete application that transforms any data source (YouTube videos, websites, PDFs, CSVs, etc.) into professional business reports and presentations with one-click AI automation in just 6 hours, demonstrating end-to-end product development from concept to working prototype.
- **AWS MLOps (Machine Learning Operations) Engineer Path**   
Obtained **AWS Certified Cloud Practitioner**, **AWS Certified AI Practitioner** and **AWS Certified Solutions Architect**, currently pursuing additional AWS certifications related to Machine Learning and Cloud Computing.
- **2nd Place at Startup Lab Competition**   
Secured 30K euros pre-incubation funding for an innovative agritech solution, developing an autonomous harvesting prototype that combines YOLO object detection, ROS, and SLAM for precision agriculture.
- **Industrial AI Challenge Winner**   
Led the development of an award-winning AI solution for BLM Group's inventory management, implementing machine learning models that reduced waste by 60 percent, improving overall supply chain efficiency.
- **Building RAG Agents with LLMs Certification from NVIDIA**   
Mastered advanced LLM optimization techniques, focusing on RAG (Retrieval-Augmented Generation) architectures, vector databases, and efficient information retrieval methods for enhanced model performance.
- **Mathematics for Machine Learning Imperial College of London**   
Acquired essential mathematical foundations for machine learning including linear algebra, calculus, probability and PCA (Principal Component Analysis) strengthening the theoretical basis for algorithm development and optimization.

# SKILLS

Strategic Thinking

Active Listening

Motivator

Time Management

Adaptability

English (IELTS 8)

Docker

Git

AWS

Cloud Computing

Linux

Complexity Analysis

Python

Scikit-learn

SQL

Pandas

PyTorch

NumPy

R

Jupyter Notebook

C/C++

RAG (Retrieval-Augmented Generation)

Generative AI

Quantization

AI Agents

Foundational Models

LLMs (Large Language Models)

NLP (Natural Language Processing)

Fine-tuning

Prompt Engineering

MLOps

CUDA

Containerization


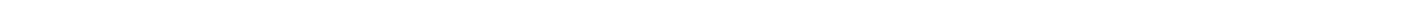

LaTeX

Tikz

Technical Documentation

Googling

# SELECTED PROJECTS

- Deep Learning Study Book** 
  - Authored a comprehensive deep learning textbook using LaTeX and TikZ, creating detailed technical illustrations of neural architectures and advanced ML concepts, publicly available in the University of Trento's shared resources.
  - Developed extensive documentation of core deep learning principles, including custom visualizations of attention mechanisms, CNN architectures, and optimization techniques, serving as a key learning resource for future students.
- DeepQuant Neural Network Quantization Library** 
  - Led the development a comprehensive Python library for converting quantized models with AMD's Brevitas library to true integer-only representations, enabling efficient deployment on embedded systems without floating-point units.
  - Implemented advanced graph transformation algorithms including quantization node splitting, dequantization unification, and custom forward pass injection for neural network layers, supporting complex architectures like CNNs such as ResNet and Transformer-based models including Vision Transformers.
- Enhanced File Explorer for Chrome** 
  - Developed a browser extension to enhance local file browsing experience across Chromium browsers.
  - Gathered more than 900 daily users and surpassing 5000 installations on Chrome Web Store.