

TIAN CHENG "RICCARDO", XIA

✉ riccardo.xia@outlook.com |  linkedin.com/in/tian-cheng-xia
 notxia.github.io |  github.com/NotXia |  orcid.org/0009-0005-5274-7469

WORK EXPERIENCE

ML Research Intern National Institute of Informatics, Tokyo, Japan Jul 2025 - Jan 2026

- Built a multi-agent LLM + RAG system using OpenAI and open-source models, served via vLLM, capable of generating fully grounded scientific documents with 100% citation consistency.
- Deployed and managed an LLM evaluation framework on an 8x NVIDIA A100 GPU cluster, testing robustness across quantization settings and adversarial linguistic inputs.
- Co-designed and annotated a multimodal dataset of 1664 samples across 180 papers for scientific claim verification, released publicly as part of a shared evaluation task.

ML Research Intern SmartData Research Group, Bologna, Italy Apr 2023 - Sep 2023

- Built a biomedical text summarization pipeline using fine-tuned transformer models, improving ROUGE and BERTScore by 5% over prior state-of-the-art baselines.
- Surveyed the LLM landscape for biomedical NLP applications, producing a technical report that shaped the team's future roadmap.

EDUCATION

M.Sc. in Artificial Intelligence University of Bologna, Italy Mar 2026

Thesis title: A Self-Supervised Attribution Method for Explaining Neural Networks

Graduation grade: 110/110 with honours GPA: 29.6/30 (top 5%)

B.Sc. in Computer Science University of Bologna, Italy 2023

Thesis title: Subtopic-Oriented Biomedical Summarization using Pretrained Language Models

Graduation grade: 110/110 with honours GPA: 29.52/30 (top 5%)

SKILLS

ML Focus	LLM fine-tuning, Multi-agent systems, RAG, Multimodal models, Explainable AI
ML Frameworks	PyTorch, JAX, TensorFlow, HuggingFace, Scikit-learn, Pandas, OpenCV, vLLM, Ollama
RAG Frameworks	LangChain, LangGraph, Pinecone
Infrastructure	AWS, GCP, Docker, Kubernetes, Ansible, GitHub Actions, CI/CD, Linux, Git
Web/Data	FastAPI, Flask, NodeJS, React.js, Vue.js, SQL, MongoDB, Apache Spark
Languages and Tools	Python, C, C++, Java, JavaScript, TypeScript, Bash

ACHIEVEMENTS

1st Place at LauzHack 2024, BMS Challenge EPFL, Lausanne, Switzerland 

Developed an explainable time-series forecasting model using Gaussian Processes and Shapley values to predict pharmaceutical demand, ranking 1st out of 12 teams in the Bristol Myers Squibb 24 hours hackathon challenge.

1st Place at Tablut AI Agent Challenge University of Bologna, Italy 

Designed and developed an AI player for the asymmetric board game Tablut using adversarial search and heuristic evaluation. The player ranked 1st out of 10 agents.

SELECTED PROJECTS

Real-Time Open-Set Person Identification from Eye-Tracking Behavior Python, PyTorch

Designed an open-set framework for identifying people using contrastive learning and time-series data from a Tobii Fusion eye-tracker. With few seconds of reference gaze data, the method achieves 94% F1-macro and an inference time of 2-3 milliseconds on CPU.

Street Anomaly Segmentation Python, PyTorch, HuggingFace 

Designed a computer vision framework to detect out-of-distribution objects in urban driving scenes by leveraging vision transformers without needing supervision, achieving a +17% improvement in AUPR over prior baselines.

Optimisation Programming for Multiple Couriers Planning MiniZinc, OR-Tools, Z3, Gurobi, Docker 

Modelled the Capacitated Vehicle Routing Problem using CP, SAT, SMT, and MILP and built a GitHub Actions CI/CD pipeline for automated correctness verification and performance benchmarking.

Homelab Proxmox Cluster Proxmox, Docker, Kubernetes, Ansible

Personal Proxmox cluster consisting of two mini computers and a NAS. Used to experiment with distributed systems and to self-host daily services (DHCP, DNS, VPN, password manager, cloud storage, Git server, ...).