Giovanni Pinna

AI ENGINEER | ML ENGINEER | APPLIED AI SCIENTIST | SOFTWARE ENGINEER

giovannipinna96 | giovanni-pinna | giovanni.pinna.17l96@gmail.com |

Giovanni Pinna | Trieste, Remote, Willing to Relocate

SUMMARY

AI Researcher and Engineer with a **Ph.D. in Applied Data Science & Artificial Intelligence** and 3 years of industry-academic experience in Natural Language Processing (NLP), Large Language Models (LLM) optimization, AI Agents, and genetic improvement. Author of **10+ peer-reviewed publications** in top venues (Scientific Reports, IEEE, Springer). Built production AI systems—including a Retrieval Augmented Generation (RAG) solution that **reduced client support calls by 30%**. Visiting Researcher at **UCL** (London) and **NOVA IMS** (Lisbon), with publications at SSBSE, Scientific Reports, and IEEE.

SKILLS

Programming Languages:	Python, SQL, Java, C++
ML & AI Frameworks:	PyTorch, scikit-learn, HuggingFace Transformers, spaCy, NLTK, BERTopic
LLM & Agents:	LangChain, LlamaIndex, LangGraph, RAG Pipelines, Prompt Engineering
NLP & Methods:	Natural Language Processing (NLP), Text-to-SQL, Sentiment Analysis, Topic Modeling, Genetic Improvement, Grammatical Evolution, NSGA-II
Tools & Infrastructure:	Git, Docker, Linux, Streamlit, Gradio

EXPERIENCE

University College London (UCL) — London, UK

Visiting Researcher (supervised by Prof. Federica Sarro) — Sept. 2025 — Dec. 2025

- Co-engineered **GA4GC**, the first multi-objective framework for optimizing Large Language Model (LLM) coding AI agent configurations using NSGA-II on the SWE-Perf benchmark, **achieving 135× hypervolume improvement, 43.6% runtime reduction** while maintaining correctness, and outperforming random search baselines (83.0% vs 53.1% cumulative hypervolume).
- Led empirical study analyzing **7,156 AI-generated pull requests** across 5 coding AI agents, revealing that task type drives a **16-percentage-point gap in acceptance rates** (documentation: 82.1% vs features: 66.1%); first-authored paper accepted at MSR 2026.
- Co-engineered **HotCat**, a green feature selection framework using NSGA-II for hotfix bug taxonomy, **achieving 0.59 accuracy and 0.58 NMI across 17 categories in 129 seconds** with no additional resource overhead with respect to the default configuration.
- Co-authored study on message-code inconsistency in **23,247 agentic pull requests**, manually annotating 974 PRs and identifying 8 inconsistency types; **high-inconsistency PRs showed 51.7% lower acceptance and 3.5× longer merge times**; published at MSR 2026.

NOVA Information Management School (NOVA IMS) — Lisbon, Portugal

Visiting Researcher (supervised by Prof. Mauro Castelli) — May—Aug. 2024 & May—Aug. 2025

- Created a novel Text-to-SQL evaluation metric (QAS) integrating semantic and structural SQL similarity, benchmarked across **11 models** on the **BIRD dataset**; published in **Scientific Reports (Nature)**, addressing key limitations of binary evaluation approaches.
- Developed an AI Agent for Text-to-SQL tasks using **only small open-source models**, enabling cost-effective natural language database querying without proprietary APIs.

PLUS S.r.l., Area Science Park Trieste, Italy
Applied AI Scientist (Industrial PhD. Fellow) Jan. 2023 — Dec. 2025

- Architected and deployed a production Retrieval Augmented Generation (RAG) system using Python, LangChain, LlamaIndex, and HuggingFace Transformers, enabling natural language querying of corporate documentation and **reducing client call center volume by 30%**.
- Mentored 1 intern for 16 weeks in designing and deploying AI-powered Retrieval Augmented Generation (RAG) solutions for enterprise clients, guiding the full development cycle from architecture design to production deployment.
- **Designed and presented a proof-of-concept AI Agent system** for automated querying of municipal service portals, enabling non-technical users to navigate complex regulatory procedures through natural language.

University of Trieste Trieste, Italy
Teaching Assistant and Department Tutor Jan. 2024 — Dec. 2024

- Designed lab exercises and evaluated projects for **Database Systems** (B.Sc. Computer Engineering), guiding **100+ students** through database design and SQL implementation cycles.
- Restructured the Bachelor's, Master's, and Doctoral websites of the Department of Data Science & AI, improving information accessibility for **200+ students**.

BDF S.p.A. / Fincantieri S.p.A. Trieste, Italy
IT Support Specialist Apr. 2019 — Aug. 2019

U-BLOX Italia S.p.A. Sgonico, Italy
Quality Testing Intern Jul. 2015 — Sept. 2015

EDUCATION

University of Trieste Trieste, Italy
Ph.D. in Applied Data Science and Artificial Intelligence, Doctor Europaeus Jan. 2023 — Dec. 2025

- Thesis: "Application of Large Language Models: Addressing Real-World Challenges" — Grade: **With Honors**.
- Engineered a Genetic Improvement pipeline using Grammatical Evolution to automatically correct LLM-generated code, **improving solution accuracy in 50–75% of cases** across 4 LLMs and 25 benchmark problems; published at EuroGP 2024 (Springer) and SN Computer Science.
- Mentored **3 M.Sc. and 1 B.Sc. thesis students** on Natural Language Processing (NLP) and code generation research; 2 theses contributed directly to peer-reviewed publications at EuroGP 2024 and IEEE WI-IAT 2024.

University of Trieste Trieste, Italy
M.Sc. in Computer Engineering Sept. 2019 — Oct. 2022

- Thesis: "An Automatic Tool for the Recognition of Punches in Late-Medieval Panel Paintings".

Montanuniversität Leoben Leoben, Austria
Erasmus+ Exchange Sept. 2021 — Feb. 2022

- Core Courses: Production Data Analysis & Modeling, Fundamentals of Logistics Systems Engineering, Computational Methods and Tools, Decision-Making and Risk Analysis.

University of Trieste Trieste, Italy
B.Sc. in Electronic and Computer Engineering Sept. 2015 — Mar. 2019

SELECTED PUBLICATIONS

1. **Pinna, G.**, Gong, J., Williams, D., Sarro, F. (2026). "Comparing AI Coding Agents: A Task-Stratified Analysis of Pull Request Acceptance." *MSR* 2026.
2. Gong, J., **Pinna, G.**, Bian, Y., & Zhang, J. M. (2026). "Analyzing Message-Code Inconsistency in AI Coding Agent-Authored Pull Requests." *MSR* 2026 **(MSR 2026 Distinguished Mining Challenge Paper Award)**.

3. **Pinna, G.**, Perezhohin, Y., Manzoni, L., Castelli, M., De Lorenzo, A. (2025). "Redefining Text-to-SQL Metrics by Incorporating Semantic and Structural Similarity." *Scientific Reports* 15.1.
4. **Pinna, G.**, Ravalico, D., Rovito, L., Manzoni, L., De Lorenzo, A. (2025). "Exploring the Effect of Genetic Improvement for LLM-Generated Code." *SN Computer Science* 6.7.
5. **Pinna, G.**, Ravalico, D., Rovito, L., Manzoni, L., De Lorenzo, A. (2024). "Enhancing LLMs-Based Code Generation by Leveraging Genetic Improvement." *EuroGP 2024*, LNCS vol. 14631, Springer.
6. **Pinna, G.**, Tugnoli, D., Manzoni, L., De Lorenzo, A. (2024). "From Courts to Comprehension: Can LLMs Make Judgments More Accessible?" *IEEE/WIC WI-IAT* 2024.
7. Gong, J., et al. incl. **Pinna, G.** (2025). "GA4GC: Greener Agent for Greener Code via Multi-Objective." *SSBSE* 2025.
8. de la Cal, L., et al. incl. **Pinna, G.** (2025). "HotCat: Green and Effective Feature Selection for HotFix Bug Taxonomy." *SSBSE* 2025.
9. Zullich, M., Macovaz, V., **Pinna, G.**, Pellegrino, F.A. (2023). "An AI System for Automatic Recognition of Punches in 14th-Century Panel Painting." *IEEE Access*.
10. **Pinna, G.**, Ravalico, D., Rovito, L., Manzoni, L., & De Lorenzo, A. (2025). Improving LLM-Generated Code via Genetic Improvement: A Summary of Recent Advances. *CEUR Workshop Proceedings*.

PROJECTS

NLP Pipeline for Multilingual Historical Newspaper Analysis. Developed the full software pipeline (Python, BERTopic, spaCy, HuggingFace sentiment models) for a Digital Humanities project comparing Italian and Slovenian newspapers from 1902 Trieste, including LLM-assisted OCR correction, topic modeling, and cross-lingual sentiment analysis on **50 issues**. Paper accepted at **AIUCD 2026**.

Contamination Lab (CLab), University of Trieste, 2020–2022. Reached **top 7** in the university entrepreneurship competition, developing a business idea with faculty mentors and industry experts. (Video Presentation)

TEACHING AND ADVISING

CLASSES

Database Systems, University of Trieste, B.Sc. in Computer Science Engineering and B.Sc. Artificial Intelligence and Data Analytics, 2024.

M.Sc. STUDENTS

Damiano Ravalico, Thesis in Computer Science Engineering, 2023.
Davide Tugnoli, Thesis in Computer Science Engineering, 2024.
Lorenzo Medeot, Thesis in Computer Science Engineering, 2025.

B.Sc. STUDENTS

Andrea Giacomazzi, Thesis in Artificial Intelligence and Data Analytics, 2025.

AWARDS AND GRANTS

Research Grant for Compute Credits (5k $) Google Cloud, 2024.

CERTIFICATIONS

McKinsey Forward Program — 10-week professional development by McKinsey.org.

PROFESSIONAL DEVELOPMENT

Specialization Schools:	AthNLP (Athens, Sept. 2024; selected from 200+ applicants), LxMLS (Lisbon, July 2024), ALPS (French Alps, Apr. 2024; ∼50 selected participants), DeepLearn (Gran Canaria, July 2023), OxML (Oxford, July 2023).
Volunteering:	EACL 2024 (St Julian's, Malta), IEEE CCTA 2022 (Trieste, Italy).
Memberships:	AILC (Italian Association of Computational Linguistics), AI2S, Mentors4U (mentee).

LANGUAGES

Italian (native), English (fluent).