

Andrea Gurioli

PhD Student in LLMs for Efficient, Trustworthy and Agentic AI

andrea.gurioli5@unibo.it • +39 331 124 0387

Google Scholar • University profile • DBLP



PROFILE

PhD student in Data Science and Computation working at the intersection of Large Language Models (LLMs), representation learning, information retrieval and trustworthy software generation. My research develops **efficient models that adapt their computational footprint, structured learning signals grounded in executable feedback, and scalable retrieval and provenance pipelines**. I now aim to transfer this expertise to onboard agentic AI: resource-aware models that produce constrained, verifiable decisions and coordinate heterogeneous Earth-observation satellites under dynamic operational conditions.

Adaptive inference

LLM fine-tuning & RL

Embeddings & ranking

Graph representation learning

Reproducible ML

RESEARCH EXPERIENCE

Sep 2023 - Present

PhD Student, Data Science and Computation

University of Bologna

Research on code authorship, LLMs for software development, efficient transformer encoders, retrieval and model adaptation. Designed reproducible experiments across multilingual datasets and open model families.

Gen - May 2025 **Visiting Research Intern**

Télécom Paris, Institut Polytechnique de Paris

Research on deep metric learning, code stylometry and large-scale data mining. The collaboration seeded the work leading to *Is This You, LLM?*.

May - Sep 2023 **Research Fellow**

University of Bologna

Developed code-stylometry and generative-code research using the Software Heritage archive; built data-mining and machine-learning pipelines for real-world open-source code.

COLLABORATIONS

APPLE MLR

Efficient model deployment

Co-authored MoSE, a modular multi-exit encoder accepted at AAI-26.

CODE COMMONS

LLM code provenance

Addressed the project's challenge of scalable attribution and licensing.

TÉLÉCOM PARIS / IP PARIS

International research

Joint work on code stylometry, authorship and AI-generated code detection.

EDUCATION

2023 - Present **PhD in Data Science and Computation**

University of Bologna

Thesis topic: *Code authorship and large language models for software development*.

2023

MSc in Computer Science - Software Techniques

University of Bologna

Thesis: *Code stylometry: a metric-learning approach*.

2020

BSc in Computer Science for Management

University of Bologna

Thesis on Ethereum smart contracts for financial-option implementation.

SELECTED PUBLICATIONS

Selected for relevance, citation impact and venue quality. Andrea Gurioli is highlighted in bold.

2026 **MoSE: Hierarchical Self-Distillation Enhances Early Layer Embeddings** **AAAI-26**

A. Gurioli, F. Pennino, J. Monteiro, M. Gabbrielli. *Fortieth AAAI Conference on Artificial Intelligence*, pp. 30897–30906. Co-authored with Apple MLR.

A 1B-parameter multi-exit encoder for adaptive deployment. Reducing computation by about 90% from the deepest to the earliest exit costs only 6.4 points in text-to-code retrieval, directly demonstrating accuracy–efficiency control under resource constraints.

2025 **Is This You, LLM? Recognizing AI-written Programs with Multilingual Code Stylometry** **SANER-25**

A. Gurioli, M. Gabbrielli, S. Zacchiroli. *IEEE SANER 2025*, pp. 394–405. Research developed with Télécom Paris / Institut Polytechnique de Paris.

Introduced one transformer classifier for detecting AI-written code across ten programming languages and released H-AIRosettaMP, an open dataset of 121,247 labelled snippets.

2026 **Efficient and Scalable Provenance Tracking for LLM-Generated Code Snippets** **CODE COMMONS**

A. Gurioli, D. D’Ascenzo, F. Pennino, M. Gabbrielli, S. Zacchiroli. *arXiv:2605.28510*; preprint.

Developed SourceTracker and a hybrid vector-search/fingerprinting pipeline over a 10-million-snippet corpus, combining scalable candidate retrieval with exact re-ranking for transparent, high-precision attribution.

2026 **From Reasoning to Code: GRPO Optimization for Underrepresented Languages** **ICLP-26**

F. Pennino, B. Raimondi, M. Rondelli, **A. Gurioli**, M. Gabbrielli. *International Conference on Logic Programming*; accepted.

Used Group Relative Policy Optimization with interpreter-based rewards for logical correctness and structural formatting, providing practical experience with reinforcement learning from verifiable feedback.

2025 **Trajectory-Embedded Matryoshka Representation Learning for Enhanced Similarity Analysis** **ESANN-25**

F. Pennino, **A. Gurioli**, M. Gabbrielli. Graph neural network with nested embeddings and two-stage ranking; 8-fold retrieval speed-up at comparable precision.

2024 **Stylometry for Real-World Expert Coders: A Zero-Shot Approach** **PEERJ CS**

A. Gurioli, M. Gabbrielli, S. Zacchiroli. *PeerJ Computer Science* 10:e2429. Open dataset of 114,400 real-world code snippets and zero-shot authorship recognition for unseen developers.

TECHNICAL TOOLKIT

Machine learning	Transformers, representation and metric learning, contrastive learning, self-distillation, GRPO, graph neural networks, retrieval, ranking, model evaluation
Engineering	Python, PyTorch, Hugging Face, vector search / HNSW, reproducible pipelines, dataset construction; Java (advanced), JavaScript, Scala and C
Languages	Italian (native), English (C1), French (B1)