

PRAVEEN BUSHIPAKA

bushipakapraveen@gmail.com | linkedin.com/in/praveen-bushipaka | github.com/praveensonu | bushipakapraveen.com

RESEARCH & ENGINEERING PROFILE

PhD researcher in Artificial Intelligence, working on LLM unlearning, data selection, and trustworthy AI under the supervision of Prof. Tommaso Cucinotta and Dr. Lucia Passaro. My research focuses on efficient data selection, reliable evaluation of LLM unlearning, and privacy-preserving and secure AI systems. My broader experience includes distributed training, parameter-efficient fine-tuning (PEFT), retrieval-augmented generation (RAG), reinforcement learning (RL), LLM evaluation, and end-to-end AI tooling. Prior to my PhD, I gained industry experience developing and deploying traditional machine learning models in real-world settings. My research interests span LLMs, multimodal and vision-language models (MLLMs/VLMs), reinforcement learning, and the privacy, security, and trustworthiness of modern AI systems.

CORE TECHNICAL SKILLS

LLM Research & Alignment	Machine Unlearning, data selection / coresets, SFT, DPO, RLHF, GRPO, LoRA/QLoRA, PEFT, LLM-as-a-Judge
LLM Systems & Evaluation	PyTorch, Hugging Face Transformers, TRL, Accelerate, distributed/multi-GPU training, custom losses & datasets, benchmarking, error analysis
Retrieval & Applied NLP	Hybrid / Agentic RAG, embedding fine-tuning, reranking, GraphRAG, Milvus, ChromaDB, LlamaIndex, Neo4j, LangChain
Engineering & Deployment	Python, FastAPI, Docker, vLLM, Ollama, ONNX Runtime, TypeScript, SQL, experiment tracking, reproducible ML pipelines

RESEARCH EXPERIENCE

PhD in AI for Society Scuola Superiore Sant’Anna & University of Pisa	Nov 2023 – Present
- Researching data-centric LLM unlearning forget/retain set construction, gradient-guided selection, sampling strategies, privacy, and utility-preserving evaluation across models and unlearning algorithms. - Developed GRACE, a gradient-guided coreset selection method for constructing compact forget and retain sets; accepted to Findings of EMNLP 2026. - Designed an end-to-end Bank of Italy regulatory RAG system (Human in Loop setting) with embedding fine-tuning, hybrid retrieval, reranking, and automated evaluation using GPT-4o, Gemini Flash, and Llama-3.3-70B. Reduced months of response time to hours. - Designed a Hybrid Agentic AI system for Vodafone that updates network changes to the network operator. Uses both open and closed models, a graphRAG with Neo4j, classifier for efficient tool calling/decision making (similar to Jev). - Building ForgetLLM, a full-stack LLM unlearning and data-selection platform with modular algorithms, distributed training, experiment evaluation, and interactive model/metric workflows; research from the platform is being prepared for AAAI submission.	
Researcher Scuola Superiore Sant’Anna	Sep 2022 – Aug 2023
Built legal-NLP pipelines for extractive summarization and rhetorical-role classification using Italian-Legal-BERT and Hugging Face; contributed to three peer-reviewed publications (ICAIL, ASAIL, CEUR).	
Freelance ML Engineer Rubidium Data, Germany	Sep 2023 – Oct 2023
Built an automated ad-placement mapping pipeline for Ecosia using SERP data; fine-tuned BERT and benchmarked against GPT-3.5, reaching 95% accuracy.	
Junior Data Scientist Wunderman Thompson, Milan	Dec 2019 – Feb 2021
Developed customer segmentation, recommendation, and analytics pipelines with clustering, market-basket analysis, Python, R, SQL, and Bash.	

SELECTED PUBLICATIONS

GRACE: Gradient-guided Coreset Selection for LLM Unlearning <i>Praveen Bushipaka, Andrea D’Angelo, Lucia Passaro, Tommaso Cucinotta</i> — Findings of EMNLP 2026 — Accepted
OARS: Objective-Aware Retain-Set Selection for LLM Unlearning <i>Praveen Bushipaka, Lucia Passaro, Tommaso Cucinotta</i> — Under review at NAACL’2027
ForgetLLM: A Practical Implementation of LLM Unlearning <i>Praveen Bushipaka et al.</i> — Under review at AAAI’2026 Demo
Standard vs. Modular Sampling: Best Practices for Reliable LLM Unlearning <i>Praveen Bushipaka, Lucia Passaro, Tommaso Cucinotta</i> — ECML-PKDD Workshop / Springer CCIS, 2025
Embracing Diversity: A Multi-Perspective Approach with Soft Labels <i>Praveen Bushipaka, Benedetta Muscato, Gizem Gezici, Luica Passaro et al.</i> — HHAI’ 2025
A Novel Multi-Step-Prompt Approach for LLM-based Q&As on Banking Supervisory Regulations <i>Daniele Licari, Canio Benedetto, Daniele Bovi, Praveen Bushipaka, et al.</i> — Clic-IT’2024

TALKS & RESEARCH COMMUNICATION

- ECML-PKDD 2026 — WIPE OUT Workshop: presented research on data selection and reliable machine unlearning for LLMs.
- Learning from Large, Complex and Structured Data Workshop, Scuola Superiore Sant’Anna: invited talk on machine unlearning methods for LLMs.

EDUCATION

PhD in AI for Society Scuola Superiore Sant’Anna & University of Pisa	2023 – Present
MSc in Computer Science University of Milan	2017 – 2020