+39 351 836 5385
Seregno, Italy
federico.raspanti2000@gmail.com
github.com/federaspa

Federico Raspanti

Data Scientist

Summary

Data Scientist specialised in **Small Language Models**: knowledge distillation, on-premise deployment of compact models as auditable alternatives to frontier APIs. Second specialisation in **Neuro-Symbolic AI**, making AI models more deterministic and reliable for high-stake domains. Two peer-reviewed papers (ACL 2025 Industry Track; LREC 2026).

Education

MSc in Data Science & AI, *EIT Digital Master School:*
·*Polytech Nice Sophia, Nice* **09 2022 – 08 2023**
·*Eindhoven University of Technology, Eindhoven* **09 2023 – 08 2024**
Thesis topic: Improving reliability of Symbolic Reasoning with Large Language Models

BSc in Economics, Management and Computer Science, *Bocconi University, Milan* **09 2019 – 08 2022**

Experience

Analyst (Data Science) **09 2025 — Present**
Prometeia *Milan, Italy*

- Built an on-premise document analysis pipeline on Gemma 4 (2B and 4B) with dots.ocr, keeping client data inside client infrastructure, achieving 90% KPI extraction accuracy.
- Contributed to Greench-v1, a Qwen3-4B ESG screener distilled from a Qwen3-14B teacher; GRPO post-training raised weighted F1 by 49.0%, beating Qwen3-32B and Claude Sonnet 4.5 (LREC 2026).
- Co-developed a RAG chatbot that explains the business and mathematical meaning of transactional credit-model indicators to commercial teams and clients; indicators are translated to English deterministically and tagged with metadata to retrieve matching explanations from past projects.

Doctoral Researcher **09 2024 — 08 2025**
Eindhoven University of Technology *Eindhoven, Netherlands*

- Benchmarked 13 open-source models from 0.5B to 27B (Gemma 2, Llama 3, Mistral, Qwen 2.5) as logical parsers; grammar-constrained decoding raised solver-executable output from 0.00 to 0.98 on 1B models.
- Showed constrained zero-shot generation matches unconstrained five-shot prompting, with gains largest for small models; published at ACL 2025 Industry Track.

Automated Document Analysis Internship **02 2024 — 08 2024**
Philips International B.V. *Eindhoven, Netherlands*

- Built an LLM pipeline that ingested 3GPP specifications and generated improvement reports; on a dozen specs, 15% of its suggestions were accepted by standards reviewers.

Freelance Data Engineer **04 2023 — 01 2024**
Monash University *Melbourne, Australia*

Computer Vision Internship **05 2023 — 08 2023**
INRIA Centre *Nice, France*

Skills

Small Language Models	Knowledge Distillation, GRPO, Supervised Fine-Tuning, Evaluation Across Scales
Constrained Generation	Grammar-Constrained Decoding (GBNF), Semantic Control, Syntactic Control
Knowledge Engineering	Ontology Integration, Symbolic Reasoning, Logical Parsing, RAG
LLM Frameworks	HuggingFace, PyTorch, llama.cpp, LangChain, OpenAI, Azure, vllm
Tools and Programming	Python, Git, Docker, Prover9, SymPy, NLTK, Clingo
Languages	English, Italian, French, Spanish

Publications

[1] F. Raspanti, T. Özçelebi, and M. Holenderski, "Grammar-constrained decoding makes large language models better logical parsers," in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pp. 485–499, Association for Computational Linguistics, 2025.

[2] S. Scala, F. Raspanti, İ. Demirtaş, A. P. Bardelli, M. Di Bari, and M. Filannino, "Greench-v1: distilling slms on greenwashing detection," in *Proceedings of the 2nd Workshop on Ecology, Environment, and Natural Language Processing (NLP4Ecology) @ LREC 2026*, pp. 79–86, European Language Resources Association (ELRA), May 2026.