# DAVIDE GIARDINI

Data Scientist & AI Developer

🌐 giardinidavide.com    ✉ davide.giardini00@gmail.com

📞 +39 327 376 3390    DavideGiardini

📍 Milan, Italy    in Davide Giardini

**LANGUAGES**
**English** - C1; **Italian** - native

AI Developer & Data Scientist with strong theoretical foundations in statistics and machine learning. I transform unstructured data into structured Knowledge Bases through advanced NLP techniques and build Retrieval-Augmented Generation (RAG) systems on top of them. Experienced in LLM fine-tuning and deploying production-grade AI services. Published author on knowledge-graph–driven NLP for forensic analysis.

## EXPERIENCE

**3/2024 – Present**    **AI Developer**        **PricewaterhouseCoopers Italy**

- Built and deployed an ML-assisted text **summarization service** integrated into the company's main platform, handling **650+ monthly** user requests through the production UI.
- Developed an AI-powered **data extraction system** from documents and multimedia to create and populate **structured configurations**. The service receives **6000+ monthly** requests through the prod. UI.
- Designed and maintained a backend **NER** + entity disambiguation pipeline for enabling real-time association of orphan documents to archived requests. Processes **2400+ daily** requests during busy season.
- Development of Master Degree Thesis Project.

```
LLMs / VLMs / GPT / Claude / Gemini / GCP / FastAPI / Azure / GIT / NER / Summarization
```

**8/2023 – 1/2024**    **Student Researcher**        **CINI - INSID&S Lab**

- **Co-authored a scientific publication** on the integration of KGs and NLP for forensic analysis: *Combining Knowledge Graphs and NLP to Analyze Instant Messaging Data in Criminal Investigations*.
- Applied advanced **NLP** techniques to **extract named entities** from unstructured data originating from real judicial investigations.
- Structured and stored WhatsApp chat data into a Neo4j-based **Knowledge Graph** (379k+ nodes), enriched with extracted entities and metadata.
- Implemented locally running **Automatic Speech Recognition** (ASR) pipelines to transcribe evidentiary audio content into searchable text.

```
Research / KGs / Neo4j / NLP / Spacy / NER / NEL / Bert / ASR / Whisper
```

## EDUCATION

**9/2022 - 03/2025**    **Master Degree: Data Science**        **University of Milano Bicocca; Milan, Italy**

- **110 with honours**. GPA: 29.53
- **Thesis Project**: *Unveiling Hidden Information in Unstructured Documents: Organization and Hybrid Retrieval with Knowledge Graphs*.

```
NLP / Fine Tuning / ML / Deep Learning / NNs / RAG / Data Management / Decision Models
```

**9/2019 - 7/2022**    **Bachelor Degree: Statistical Sciences of Economy and Business**        **Alma Mater Studiorum; Bologna, Italy**

- **110 with honours**. GPA: 28.87
- **Thesis Project**: *Evaluation of the Effectiveness of Twitter Sentiment Analysis as a tool for Stock Market Forecasting*

```
Statistics / Probability Theory / Linear Algebra / Big Data / Data Analysis / Forecasting
```

## PROJECTS & INTERESTS

### Projects        giardinidavide.com

A comprehensive demonstration of my skills can be found in the selected projects available in my portfolio.

```
Python / R / LoRA / Mistral / TensorFlow / PyTorch / transformers / Word2Vec / CNNs / MongoDB / Scraping
```

### Interests

Deep personal interest in neuroscience and its connection to AI, with ongoing independent reading and research into biologically inspired models and cognitive architectures.