Bruno Guzzo

Machine Learning Engineer — Cloud Specialist — Consultant

Work Experience

Mid-Senior Consultant
January 2025 – Present

Go Reply, Turin

- Coordinated a multi-company effort to gather customer requirements and design a cloud-native **research-oriented healthcare platform**, enabling cloud-centralized and rapid clinical and **genomic data retrieval**.
- Directed a cross-functional team of developers in delivering a **B2C cloud-native mobile application** enabling after-sales support, digital documentation access via Gen-AI chat-bot and seamless multi-cloud customer side **CRM integration**.
- Designed and implemented a scalable **legal treaties search engine and RAG solution** using **Vertex AI** with a highly efficient and concurrent data ingestion pipeline, enabling insight generation and optimization of core business processes. Improved data processing speed by **94%**.
- Advised clients on **AI and agent-based system** integration, facilitating the adoption of Google Cloud enterprise ecosystem via **custom-tailored RFP solutions** to improve critical business processes. Increased client adoption rate of Vertex AI by **30%**.
- Managed **strategic client engagement** within the global development industry, coordinating **multi-project cloud advisory** and **resource allocation** that mobilized **20%** of the BU workforce and drove **30%** of unit profitability.
- Mentored junior engineers, **promoted agile best practices** through knowledge-sharing events, and **coordinated sprint reviews** and retrospectives to accelerate team performance and continuous improvement. Successfully decreased code-review time by **60%**.

Consultant
June 2022 – December 2024

Go Reply, Turin

- Contributed to the development of a digital platform for the **agriculture and food sector**, enhancing **data collection and analytics** capabilities to support evidence-based decision-making on sustainable farming and antimicrobial resistance, while leading its **migration to Google Cloud** for scalability and performance. Successfully reduced infrastructural costs by **30%**.
- Participated in the implementation and deployment of a **digital learning solution** in the **education industry**, delivering scalable and cloud-based features aligned with client innovation goals.
- Modernized a large-scale **telecommunications cybersecurity platform**, improving scalability and enhancing user experience through **architectural refactoring**, cloud integration, and automation to strengthen **service reliability** across mobile and fixed networks.

Personal & Academic Projects

MSc Thesis: Anomaly Detection with Deep Autoencoders

- Investigated a semi-supervised anomaly detection pipeline using a novel Autoencoder (AE-SAD) architecture, specifically addressing the challenge of severe class imbalance during training.
- Demonstrated that applying advanced under-sampling techniques can significantly boost detection accuracy (AUC) in complex many-vs-many scenarios while reducing training time by up to two orders of magnitude.

Transformer-GNN for Link Prediction

- Built a scalable data pipeline to create a large-scale graph dataset from Wikipedia using web crawling and NLP.
- Trained and evaluated multiple Graph Attention Network (GAT) models on the custom dataset, analyzing graph metrics and model performance.

Anomaly Transformer Analysis

- Implemented a novel time-series anomaly detection model using the Anomaly-Transformer architecture, leveraged self-attention mechanisms to identify anomalies within the Mars Science Laboratory (MSL) dataset.
- Engineered a robust machine learning pipeline encompassing model training and rigorous evaluation. Optimized model performance through extensive hyper-parameter tuning via grid search.

Italian LLaMA 3 Fine-Tuning and Evaluation

- Enhanced the "LLaMAntino-3" model by fine-tuning it on the Italian Wikipedia, resulting in significantly improved language modeling capabilities, as evidenced by reduced perplexity.
- Developed a RAG-based chatbot using the fine-tuned model to answer questions about Italian literature, leveraging a custom corpus of documents.

Education

MSc, AI & ML Engineering
Oct 2022 – Jun 2025

UNICAL, Rende (CS), Italy

BSc, Computer Engineering
Oct 2018 – Jun 2022

UNICAL, Rende (CS), Italy

+39 339 830 5458
brunoguzzo18@gmail.com
github.com/bGuzzo
linkedin.com/in/guzzobruno

Technical Skills

Google Cloud (*Advanced*)

- **Compute**: Cloud Run, App Engine, Cloud Functions, Compute Engine
- **AI/ML**: Vertex AI, Agent Builder
- **Data**: BigQuery, Cloud SQL, Firestore, and Firebase Realtime Database, Cloud Storage
- **Infra**: VPC, Load Balancer, IAM, Cloud Armor, Cloud DNS, Cloud CDN, Cloud Task, Cloud Pub/Sub

AI & Machine Learning (*Academic*)

- **Frameworks**: PyTorch, PyTorch Geometric, Langchain, Hugging Face Transformers, PEFT (Parameter-Efficient Fine-Tuning), TRL (Transformer Reinforcement Learning), Unsloth
- **Libraries**: Pandas, NumPy, Scikit-learn, SciPy, NLTK, Sentence-Transformers, FAISS (Vector Database), BitsAndBytes (for Quantization)
- **Concepts**: LLM, RAG, NLP, Fine-Tuning (QLoRA), Vector Databases, Embeddings, Anomaly Detection, Time-Series Analysis, Reinforcement Learning, GNNs, CNNs, Autoencoder, Transformer, GPT, BERT, Hyper-parameter Tuning, Model Evaluation

DevOps (*Advanced*)

- **IaC**: Terraform
- **CI/CD**: Cloud Build, GitLab CI, GitHub Actions, Artifact Registry
- **Containers**: Docker
- **Others**: Linux, Git, A/B Testing

Software Development (*Advanced*)

- **Languages**: Python, Java, C, JavaScript, Typescript, Dart
- **Frameworks**: Spring Boot, Hibernate, JPA, Flask, Django, SQLAlchemy Angular, React, Flutter
- **Databases**: PostgreSQL, MySQL, NoSQL
- **Others**: Agile, Scrum, Jira, Sprint Planning, Cloud-native

Languages

- **Italian**: Native
- **English**: B2, Cambridge certified