

ANTONIO SOGGIA

antoniosoggia5@gmail.com • antoniosoggia.it • github.com/antoniosoggia • FR: +33 7 49 91 79 66

Profilo

Dottore di ricerca in Machine Learning ed Elaborazione del Segnale con solida competenza in ingegneria del software. Specializzato nello sviluppo di CNN leggere per ambienti rumorosi e nella costruzione di infrastrutture AI self-hosted. Completamente autonomo lungo l'intero ciclo ML: dalla modellazione matematica e generazione di dati sintetici al deployment su backend e ottimizzazione multi-GPU. Appassionato nel creare un ponte tra ricerca teorica e sistemi scalabili pronti per la produzione.

Competenze Tecniche

- **Linguaggi di Programmazione:** Python, C/C++, SQL, Bash, Scala
- **Machine Learning e Matematica:** PyTorch, TensorFlow, CNN, LLM (Ollama, llama.cpp), Elaborazione del Segnale (DSP), Regressione Simbolica, Algebra Lineare, Ottimizzazione
- **MLOps e Ingegneria:** Docker, Linux/SSH, Flask, API REST, Spark, Git, Ottimizzazione multi-GPU, JAX, MoE

Esperienza Professionale

AI Scientist (già R&D Apprentice) - Thales DMS

Sophia Antipolis (FR) | 2021 - 2026

- **Sviluppo ML End-to-End:** Guidato la transizione di una pipeline di elaborazione del segnale sonar dalla fase di prototipazione R&D a un sistema brevettato pronto per la produzione.
- **Simulazione e Data Engineering:** Progettato un simulatore sonar basato sulla fisica per generare dataset sintetici su larga scala, risolvendo una grave scarsità di dati reali per l'addestramento ML.
- **Progettazione Architetture:** Sviluppato *AntoNet*, una CNN leggera personalizzata per il denoising 3D del segnale. Raggiunto un guadagno di 11 dB (vicino al limite teorico) con soli 1.2k parametri, ottimizzata per l'inferenza in tempo reale su GPU consumer.
- **Analisi Matematica:** Condotta un'analisi del flusso del gradiente tramite regressione simbolica per caratterizzare la varietà dei minimizzatori globali, dimostrando la robustezza delle architetture lineari contro il rumore gaussiano.

Progetti di Ingegneria Selezionati

Server AI Privato e Infrastruttura LLM

Docker, Ollama, CUDA, Tailscale

- Progettato un server multi-GPU (RTX 4070+3060) per il deployment di modelli MoE quantizzati tramite Ollama e llama.cpp; implementata l'allocazione automatica della VRAM e l'accesso remoto via VPN Zero-Trust Tailscale.
- Sviluppato un motore di ricerca web self-hosted e dockerizzato per la Web-Augmented Generation (WAG) senza API esterne.

Orchestra - Piattaforma di Traduzione AI Locale

Python, Flask, Ollama, System Design

- Progettato una pipeline NLP ibrida con architettura lookup-first deterministica; il sistema risolve la terminologia confermata da una memoria semantica prima dell'inferenza LLM, garantendo il 100% di coerenza terminologica.
- Implementata l'Analisi della Topologia Documentale per la deduplicazione dei paragrafi, ottenendo una riduzione dell'85% delle chiamate LLM; rafforzata l'affidabilità tramite rilevamento della deriva linguistica e controlli di allineamento dei paragrafi.
- Costruito uno Store Semantico Persistente che funziona come Memoria di Traduzione locale, riducendo la dipendenza dall'inferenza man mano che la copertura del dominio dell'utente cresce nel tempo.

Pubblicazioni e Brevetti

- **Brevetto Depositato:** Pipeline di Denoising e Tracking 3D;
- **Articolo su Rivista:** *AntoNet: A Small Linear Contrast-Enhancing CNN for Pixel-Wise Detection*. (In revisione, Signal Processing: Image Communication). Invitato dopo il ranking Top-5 a IPTA 2024;
- **Articoli a Conferenza:** *AntoNet* (IPTA 2024); *Lightweight CNN for Sonar Detection* (EUSIPCO 2025).

Formazione

Dottorato in Matematica Applicata e Machine Learning, Université Côte d'Azur & Thales DMS

05/2026

Laurea Magistrale in Ingegneria Matematica, Université Côte d'Azur

2022, con lode

Laurea Triennale in Matematica Applicata (MASS), Université Côte d'Azur

2020, con lode

Insegnamento e Lingue

Insegnamento: Valorizzazione dei Dati (Apprendimento supervisionato, ROC/AUC) e Big Data (SQL/Spark) - Esercitatore presso Polytech UCA.

Lingue: Francese (Madrelingua), Italiano (Madrelingua), Inglese (Fluente).