

Guillaume Atencia

Pittsburgh, PA · gatencia@andrew.cmu.edu
linkedin.com/in/guillaume-atencia-2786b8252 · github.com/gatencia

EDUCATION

Carnegie Mellon University

B.S. Computer Science, Concentration in Language Technologies

Pittsburgh, PA

Aug. 2022 – Expected May 2027

Relevant coursework: Human Language for AI, Deep Learning, Machine Learning in Practice, Computer Vision, Probability for Computing

Languages: English, French, Spanish (native); German (fluent)

RESEARCH EXPERIENCE

WavLab — Carnegie Mellon University

Undergraduate Researcher; Advisor: Shinji Watanabe

Pittsburgh, PA

Sep. 2026 – Present

- Investigating whether higher-resource related languages measurably improve translation quality for low-resource languages, running controlled cross-lingual transfer experiments over Common Voice and FLEURS.
- Reproducing multilingual speech-translation baselines in ESPnet and PyTorch with pretrained multilingual speech models as the comparison point.

LLM-Based Constraint Modeling — Carnegie Mellon University

Undergraduate Researcher; Advisor: Ruben Martins

Pittsburgh, PA

Jan. 2026 – May 2026

- Built a multi-agent system that translates natural-language problems into verified SAT encodings, deliberately separating probabilistic LLM generation from deterministic compilation and verification across 8 problem families.
- Designed verifier-gated candidate ranking over clause statistics, solver feedback and fuzzing pass rates, lifting solve rate from 40% to 76%, and ran ablations isolating each gate's contribution to that gain.

EXPERIENCE

Dynamiks.AI

Software Engineer Intern

San Francisco, CA

May 2025 – Aug. 2025

- Built the synthetic-data engine behind the company's RL pipeline, pairing adaptive Markov transition models with Gaussian-process feature drift to generate 1M+ labelled CRM trajectories from a small real dataset.
- Vectorized the sampling path in NumPy to replace nested Python loops, and validated 15-dimensional trajectory fingerprints for redundancy and fidelity against held-out data before they reached the policy.

Jane Goodall Institute

Software Engineer

Pointe-Noire, Republic of the Congo

May 2024 – Aug. 2025

- Built the geospatial ML pipeline behind forest-elephant conservation planning in the Congo Basin, joining 27K+ GPS telemetry records to 16 environmental features to serve an XGBoost habitat model at 0.81 ROC-AUC.

SELECTED PROJECTS

ProspectisAI — NLP Research Discovery Platform

May 2025 – Present

- Built a platform that matches open technical problems to existing solutions across 4 ingestion sources (Reddit, arXiv, PatentsView and job APIs), backed by PostgreSQL 16 with pgvector and HNSW indexing for semantic retrieval.
- Led 2 remote engineers in France and the UK and partnered with an ESILV Paris professor to ship the system and produce research output; owned planning, review and execution.

TrueAI — LLM Hallucination Detection

Feb. 2025 – May 2025

- Built an evaluation harness that scores whether an LLM answer is grounded in its source, reaching 0.85 AUC against 0.62 for SelfCheckGPT and 0.79 for SAPLMA on annotated benchmarks.
- Ran 150+ I-Corps customer interviews whose latency findings redirected the design from LLM-as-judge re-evaluation to sub-100 ms logit and entropy scoring.

SoloMuse — Controllable Neural Audio Generation

Jul. 2024 – Sep. 2024

- Built a system generating instrumental solos over a backing track, ending in an autoregressive Transformer over 4-codebook EnCodec tokens at 75 Hz — the discrete-token sequence modelling used in modern speech systems — trained on 2,100 multitrack songs.

TECHNICAL SKILLS

ML/NLP: PyTorch, Hugging Face, ESPnet, Transformers, LLMs, machine translation, speech recognition, sequence modeling, model evaluation

Programming: Python, C, C++, Java, Rust, Swift, SQL, Bash

Systems & Data: Linux, FastAPI, PostgreSQL, pgvector, Docker, Git, async programming, NumPy, scikit-learn

Activities: CMU AI Student Initiative (CASI); technical writing on Medium; flamenco guitar