

Alexander Morgan

Email: Redacted
Tel: Redacted

[LinkedIn](#)
[GitHub](#)

Based in Barcelona, Spain
US/UK Dual Citizen, Spanish Visa

Education

McGill University: PhD Music Information Retrieval
University of Cincinnati: Master of Music Theory
UNC-Chapel Hill: BA Music/Political Science

Languages

Programming: Python, FastAPI, Postgres, TypeScript
Natural: Bilingual English/French, Intermediate Spanish

Summary

I am a senior backend engineer who owns production Python services and computational data pipelines end to end, from architecture through FastAPI/NodeJS backends, Vue/React/REST frontends, Postgres/MongoDB, and cloud deployment. I am strongest in Python performance work and relational database design, and also regularly ship fullstack product features.

I have built services primarily on AWS and Cloudflare orchestrating a variety of workflows and pipelines. Alongside industry work I research efficient ML systems, including a novel BPE tokenization optimization and contributions to LLM training speedruns. I am searching for a role with a high degree of ownership and autonomy.

Open Source & Research

[BatchBPE](#): Author

- Created a novel batched optimization to the BPE algorithm that builds tokenization vocabularies 300× faster than serial BPE while producing equivalent vocabularies; pure-Python toolkit with an [arXiv paper](#)
- V2 well underway with major memory optimizations, bottlenecks outsourced to Rust, and parallelization

[Humlib](#): Major Contributor

- Creation of novel time-series analyses
- Optimization of C++ analysis tools and CI maintenance

[modded-nanogpt](#): Contributor

- Minor contributor to the nanogpt training speedrun repo with more upcoming focused on tokenization

Work Experience

Music Information Retrieval Academic Consultant: (Feb. 2021 – Present)

- [Tasso in Music Project](#): Advisory Board Member - Defining and automating event detection in multi-stream musical data
- [pyAMPACT](#), CUNY/Brooklyn College: Music Information Retrieval Developer/Researcher - Developing pyAMPACT, a signal processing tool built on top of Pandas and Numpy
- [Haverford College CRIM Research Group](#): Lead Programmer- Developing a Pandas-based time-series analysis program

Datamaran: Senior Backend Developer: (Sep. 2024 – July 2026)

- Owned document-intelligence computational pipelines end to end, cutting costs 93% while improving output quality, reliability, and maintainability
- Fixed several Postgres and MongoDB performance issues refined indexing and querying strategies
- Built and hosted fullstack internal QA tools on our cloud stack to improve change confidence
- Broke up a FastAPI/Vue monolith into multiple microservices

[EcoRate](#): Founder (July 2020 – Dec 2023)

- Created a sustainability ratings platform for cafes and refill stores available as an app and website
- Generated a database of sustainability information about various types of businesses

Bank of America via Apex Systems: Python Developer (July 2019 – July 2020)

- Delivered 100× performance improvements in large time-series data risk-assessment tools
- Sharing best practices across teams via recorded presentations and documentation

Google via QualiTest: Natural Language Processing Bilingual Engineer (Aug. 2018 – July 2019)

- Built tools to gather, analyze, fingerprint, and structure messy data spread across Google's stack
- Increased team productivity by automating several steps in our workflow with scripts
- Launched improvements to query understanding for Google in French

[Josquin Research Project](#) Stanford University/Université libre, Brussels: Developer (Oct. 2016 – Sep. 2017)

- Postdoctoral fellowship developing three time-series data analysis tools (C++11, GitHub, Humdrum)
- Developed open source dissonance classifier among other specialized tools

Distributed Digital Music Archives and Libraries McGill University: Research Assistant (Aug. 2012 – Dec. 2016)

- Lead programmer of VIS, a Pandas-based time-series data analysis application
- Added a dozen+ analytics and quality-control checks for a production relational database
- Delivered 100× efficiency improvements in core database/ analytics functions (Pandas, Numpy)