Hayk Grigorian

haykgrigorian07@gmail.com | LinkedIn | GitHub | Hugging Face

Education

Muhlenberg College **Aug. 2022 – May 2026**
B.S. in Computer Science *Allentown, PA*

Experience

Equity Research Analyst **May 2026 – Present**
Equity Risk Sciences *Providence, RI*

- Build and maintain data pipelines for quantitative research across equities, fixed income, ETFs, commodities, and alternative assets.
- Apply NLP methods to earnings transcripts to measure sentiment and narrative shifts, supporting feature-generation and risk-analytics workflows used in active research.

Data Engineering Intern **Feb. 2026 – Apr. 2026**
Two Worlds Capital LLC *Remote (NYC)*

- Built data aggregation, cleaning, preprocessing, and feature-generation pipelines for financial and macroeconomic datasets across multiple asset classes.
- Automated ingestion and validation of large time-series data sources to improve research throughput.

Undergraduate Researcher, Machine Learning **Sep. 2025 – Feb. 2026**
Muhlenberg College *Allentown, PA*

- Proposed Selective Temporal Training (STT) to pretrain language models on temporally bounded corpora and reduce modern bias.
- Built a 90GB historical text corpus and trained transformer language models from scratch up to 1.2B parameters using PyTorch, Hugging Face, and A100/H100 GPUs.
- Developed tokenization, training, and evaluation pipelines for large-scale historical language modeling.

Software Engineer Intern **Jun. 2024 – Aug. 2024**
Quadcode *Remote (Cyprus)*

- Developed a Telegram trading-signal bot for stocks, crypto, forex, and commodities using Python and financial APIs.
- Implemented SMA, EMA, and reversal-based trading logic with scheduling and background execution using Asyncio, Threading, and APScheduler.

Projects

TimeCapsuleLLM | *Python, PyTorch, Hugging Face, Transformers*

- Built an open-source historical language model project spanning dataset creation, tokenizer training, model pretraining, and evaluation.
- Trained billion-parameter transformer models on A100/H100 GPUs and grew the project to 1,800+ GitHub stars.
- Earned coverage from Ars Technica and Popular Science, leading to a research collaboration with Georgia State University.

Publications

TimeCapsule: Generative Hallucination as a Method for Historical Sensemaking **Jul. 2026**
ACM Creativity & Cognition (20% acceptance) *London, UK*

- Accepted to ACM Creativity & Cognition 2026; presenting in London, July 2026.

Technical skills

Languages: Python, Java, JavaScript **Machine Learning**: PyTorch, Hugging Face Transformers, tokenization, model pretraining, evaluation pipelines **Data & Systems**: pandas, NumPy, REST APIs, ETL, async systems, Git, Linux, cloud GPU infrastructure