

# William Glazer-Cavanagh

## APPLIED ML ENGINEER | BRIDGING RESEARCH & PRODUCTION

---

I'm a Software Engineer / ML Scientist bridging the gap between research and production systems. I can deconstruct State-of-the-Art papers, debug GPU segfaults, iterate on experimental hypotheses, and translate technical complexity into plain English.

## CONTACT INFORMATION

---

**Phone:** 438-389-8650 | **Email:** willcavanagh@hotmail.com | **Location:** Montreal, QC / Vancouver, BC  
**LinkedIn:** linkedin.com/in/williamglazercavanagh | **GitHub:** github.com/williamGlazer | **Website:** williamglazer.github.io

## SKILLS

---

**Modeling:** PyTorch | JAX | Transformers (BERT/ESM) | Object Detection | RAG | Fine-tuning  
**Infrastructure:** Docker | Modal (Serverless) | Dask | Parquet/Zarr | Distributed Systems  
**Core Stack:** Python | SQL | Bash | ML System Design | A/B Testing | GPU Profiling

## WORK EXPERIENCE

---

### AbCellera, Montreal/Vancouver | Machine Learning Scientist II | 05/2023 - Current

#### Project: ML for Protein Engineering (Risk Assessment & Inference)

- **Translated complex science into working software:** Learned the biology domain to bridge Machine Learning, engineering, and research. Converted open-ended goals from Principal Scientists into clear engineering plans.
- **Shipped risk assessment product:** Led the full product lifecycle from prototype to production. Enabled users to filter high-risk sequences, reducing downstream failure rates by 30%.
- **Engineered risk calibration system:** Validated custom data normalization strategies on noisy biological data with ablation studies. Applied this pipeline to calibrate Foundation Model probabilities, ensuring confidence scores aligned with empirical failure rates.
- **Accelerated Transformer inference by 30x:** Re-engineered pipelines to remove serial bottlenecks. Vectorized pre-processing for multiple Foundation Models, ensuring CPU data loading did not starve GPUs.
- **Optimized distributed infrastructure:** Combined serverless fan-out with volume caching to slash cold starts by 90%. Eliminated I/O bottlenecks to achieve linear speedups, meaning processing time decreased proportionally to horizontal scaling.

#### Project: High-Throughput Computer Vision (100k+ images/day)

- **Engineered efficient deep learning architecture:** Designed a custom architecture (MobileNet backbone + detection head) to replace a heavy legacy model. Increased speed from 400 to 1200 images/s, hitting the hardware's physical limits.
- **Doubled performance on edge cases (0.4 to 0.8 mAP):** Built a tool for experts to flag model failures. Used this feedback to fine-tune the model on hard examples, fixing errors on unseen data.
- **Automated data labeling:** Built a semi-supervised pipeline to label complex multi-class assays. Validated quality with blind A/B tests, achieving an 80% preference rate from human experts over the previous baseline.
- **Eliminated Training Bottlenecks:** Engineered a sharded data format (separating metadata in Parquet from tensors in Zarr) for 1TB-scale microscopy data. Leveraged Dask for parallel processing to fully saturate GPU utilization during training..
- **Saved weeks of engineering time:** Assessed the feasibility of Deep Learning on noisy data. Redirected stakeholders toward simpler, viable methods, preventing weeks of wasted engineering.

### Croesus, Montreal | Research Intern | 05/2021 - 08/2021

- **Evaluated Retrieval-Augmented Generation (RAG) feasibility:** Assessed early methods for injecting Knowledge Graphs into BERT-based architectures to define the research group's roadmap for grounding NLP models in factual data.

## EDUCATION

---

### MSc. Machine Learning, Professional Masters | Graduation Year 2023 | University of Montreal (MILA)

- **Benchmarked RL frameworks:** Conducted comparative performance analysis of JAX vs. PyTorch for Deep Reinforcement Learning. [GitHub/jax-4-deeprl]

### BEng. Software Engineering, Artificial Intelligence Profile | Graduation Year 2022 | Polytechnique Montreal

- **Published Research:** Co-authored *Change Taxonomy for ML Pipelines* (Empirical Software Engineering, 2023).
- **Community Leadership:** Co-founded PolyAI, a student initiative for Applied AI.