Alexey Gronskiy

ML Software Engineer / Engineering Manager / TLM

+41798454046   alexey@gronskiy.com   Zurich, Switzerland

LinkedIn: agronskiy ⌞   Github: agronskiy ⌞   Homepage: gronskiy.com ⌞

Profile

As a mathematician, computer scientist, and engineering leader, I have dedicated my career to the advancement of artificial intelligence through both research and practical application. My experience ranges from developing state-of-the-art machine learning systems for code analysis and aviation safety to leading teams in the full cycle of AI product development.

Areas of experise: LLM evaluation – ML for automatic vulnerability fixing (Snyk DeepCode AI Fix ⌞) – Low-memory and low-latency model optimization (quantization, pruning, distillation) – Computer vision for autonomous flight (Daedalean's Spotter ⌞) – Industrial ML research ("Fixing security vulnerabilities with LLMs", co-author ⌞) – Data engineering and orchestration – NN verification and certification ("Concepts of design assurance for NNs", co-author ⌞) – Engineering management, team growth, mentoring.

Experience: 10 years industry (Software Engineering & Management) and 6 years academia (PhD from ETH Zurich, Computer Science)

Skills, Frameworks and Stacks

- **Main languages:** C++, Python, Go
- **Build & CI/CD:** Bazel, Pantsbuild, CircleCI, Github Actions
- **ML Training:** PyTorch, Tensorflow, DeepSpeed; Determined AI, Weights & Biases, Vertex AI
- **ML Serving:** HuggingFace, ONNX, TensorRT, MCore
- **ML Sub-areas:** Transformers, U-Nets, ResNets, VAEs; Quantization/Pruning

- **Cloud/HPC:** GCP/AWS, GCS/S3, Slurm HPC
- **Data & Formats:** PostgreSQL, Redis, MongoDB, Parquet, Kafka
- **Deployment & Balancing:** k8s, Helm, Terraform, Envoy, Docker
- **Observability:** Datadog, OpenCensus, Prometheus, Grafana
- **Org:** Phabricator, JIRA, UML/C4
- **DevEx:** Git, Mercurial, Bazaar, NeoVim, LATEX

Professional Experience

Senior SWE III *Nvidia (nvidia.com ⌞)*   **Zurich, Switzerland** *02/2025 – current*
- Working on a company-wide scalable LLM evaluation suite, integrating hundreds of benchmarks and accelerating feedback loops for numerous research and production orgs.
- Worked on a toolset for operating both internal and external leaderboards providing research insights to the teams for cutting-edge LLMs.
- Architected and led the development of a unified adapter feature that allows LLM benchmarking logic extensibility.
- Worked on the platform infrastructure for inference across multiple LLM types.
- Partnered with product and multiple orgs to accelerate research team integration.
- **Skills & Stacks:** Python, k8s, Slurm (distributed HPC), Gitlab; NeMo, TensorRT, Megatron-LM Core; technical guidance and mentorship.

Engineering manager, Staff ML SWE *Snyk (snyk.io ⌞)*   **Zurich, Switzerland** *06/2023 – 01/2025*
- Drove the launch of Snyk DeepCode AI Fix ⌞ from Beta to GA (announcement ⌞) via a series of steady model improvements and closing the product-train-eval feedback loop.
- Oversaw development of in-house data labelling service and launched labelling collaborations with external parties, improving the automated fix coverage from 1 to 7 programming languages, and increasing the quality by 20%.
- Oversaw transition of ML train/eval and crawling pipelines to automated KubeFlow cloud, speeding up the release and experimentation from days to hours.
- Developed services for hosting company-wide ML hackathons and label-a-thons. Implemented event-streaming service for large-scale analytics of DeepCode AI Fix.
- Led and grew a team from 3 to 7 engineers; advised MSc theses (jointly with ETH Zurich).
- **Skills & Stacks:** C++17/20 (for serving), Python/PyTorch/ONNX (for deep learning stack), Go (for LSP), Bazel, Kafka, k8s (GKE/EKS), Helm, Envoy, Docker; production cycle, requirements management, mentoring, people management.

Senior ML SWE, Tech lead *Snyk (snyk.io ⌞)*   **Zurich, Switzerland** *04/2021 – 06/2023*
- Core participant of Beta launch of Snyk DeepCode AI Fix ⌞; created the initial PoC of a later productionized Snyk Code Custom Rules ⌞.
- Built the infrastructure for training/evaluation and serving of LLMs for code repair and search (real-time inference of 3B+ models). Designed and built internal libraries for simplified storage of various artifacts over any S3-compliant storage.
- Brought in codebase best practices, test-driven development, scaled the experimentation.
- **Publications:** *"DeepCode AI Fix: Fixing Security Vulnerabilities with Large Language Models"* (arXiv ⌞).
- **Skills:** C++17/20, Bazel; Python/Pytorch/ONNX; Go; k8s (both GKE/Amazon EKS), Helm, Envoy, Docker; Language Server Protocol; Weights & Biases; program analysis.

Senior data engineer *SIX Banking Services (six-group.com ⌞)*   **Zurich, Switzerland** *12/2020 – 02/2021*
- Founding role in a newly established data science unit. Organized technical stack, code review and other processes.
- Performed analytical support of stakeholder business units, provided analysis and computed KPIs for ongoing projects in payment ecosystems.
- Participated in preparatory work for transition to cloud pipelines.
- **Skills:** Python, Geospatial analysis, Mircosoft Azure (Data Factory, DataLake, Databricks), JFrog Artifactory, data security and compliance.

Machine learning & computer vision engineer *Daedalean AI (ddln.ai ⬈)*   **Zurich, Switzerland** *06/2018 – 10/2020*
- Prototyped real-time visual air-to-air object detection and tracking system, currently Daedalean's core product (Daedalean's Spotter ⬈).
- Designed and built and in-house model evaluation framework which allowed to automate the release cycle and ablation experiments.
- Shaped out data acquisition and annotation strategies. Defined evaluation metrics, navigated aviation industry standards on operational performance. Supervised interns/MSc students. Carried out flight tests as a fixed-wing aircraft pilot.
- **Publications:** (with European Aviation Safety Agency) *"Concepts of Design Assurance for Neural Networks"* (Published by EASA ⬈).
- **Skills:** C++14, Python, Tensorflow, Git, object detection and tracking, cloud computing and automation (GCP/KubeFlow/Jenkins), technical ownership, requirements design, product development cycle.

Engineering intern *Google, Research & Machine Intelligence*   **Zurich, Switzerland** *05/2017 – 09/2017*
- Worked on answer ranking module for active question reformulation project. Developed several deep learning models for answer ranking.
- **Blog post:** *"Open Sourcing Active Question Reformulation with Reinforcement Learning"* (Google AI blog post ⬈).
- **Skills:** Python, C++11, Blaze (Bazel), Tensorflow (v1), NLP, experimental design, MapReduce (Flume).

Software engineer *ABBYY Software (abbyy.com ⬈)*   **Moscow, Russia** *04/2010 – 08/2012*
- Worked on a framework (ABBYY Compreno ⬈) for automated native language understanding and translation. Designed, developed SDK wrappers for language morphology analysis and text classification.
- **Skills:** C++03, C#, COM/COM+/WinAPI, SVN, interface design, internal libraries, UML.

Freelance C++ developer   **Moscow, Russia** *2008 – 2011*
- Developed a cross-platform (Mac/Linux/Windows) tool for photo management, sold and maintained it with multiple publishing houses (archive GitHub ⬈).
- **Skills:** C++03, Qt 4, MercurialHg, PostgreSQL.

Education

PhD in Computer Science *ETH Zurich, Dept. of Computer Sience, ML Group*   **Zurich, Switzerland** *2012 – 2018*
- Conducted research in theorectical ML (discrete optimization under uncertainty).
- Launched medical collaborations with University Hospital Zurich.
- Supervised several MSc theses.
- Taught "Introduction to Machine Learning" and "Statistical Learning Theory" courses, created a script for the latter.
- **Thesis:** *"Statistical Mechanics and Information Theory for Approximate Robust Inference"* (ETH Dissertation ⬈)

Specialist in Mathematics *Moscow State University, Dept. of Mathematics and Mechanics*   **Moscow, Russia** *2006 – 2011*
- Specialist (equiv. BSc + MSc) degree in Pure and Applied Mathematics, with Honors.
- **Thesis:** *"On some metrical properties of Boolean functions"* (https://gron.sk/msc-thesis ⬈)

Volunteering

Firefighter, Emergency Medical Technitian *Zurich Fire & Rescue Service*   **Zurich, Switzerland** *2014 – 2021*
- Regularly trained as an ambulance assistant (EMT).
- Participated in ambulance shifts during public events in the city of Zurich.
- Was a first responder to alarms from large-scale emergencies.

Hobbies and Other

- Privat Pilot Licence
- Sports (running, cycling, karate)
- Music (flute & guitar)
- Technical blog (gronskiy.com/posts ⬈)

Languages

- **Russian** [Native]
- **English** [Fluent]
- **German** [Fluent, C2, Goethe]
- **French** [Fluent, C1, DALF]