

# SONIA SHARAPOVA

Brooklyn, NY · sonia.sharapova@yahoo.com · (408) 797-8495  
linkedin.com/in/sonia-sharapova · github.com/sonia-sharapova · www.soniapolis.com

## OBJECTIVE

Software Engineer – Distributed training with focus on testing, validation, and verification of complex systems. Hands on experience with Machine Learning (ML), Artificial Intelligence (AI), and High Performance Computing (HPC) hardware. Seeking opportunities to further develop performance testing/automation skills.

## TECHNICAL SKILLS AND COURSES

<b>Programming Languages:</b>	Python, C, Java, Go, R, SQL, Bash, JavaScript, HTML/CSS
<b>Machine Learning &amp; AI:</b>	PyTorch, TensorFlow, Scikit-Learn, Hugging Face Transformers, Keras
<b>HPC &amp; Data:</b>	CUDA, OpenMP, MPI, GPU Computing, SLURM
<b>Data Science &amp; Analytics:</b>	NumPy, Pandas, Matplotlib, Scanpy, Jupyter, MATLAB, OpenCV
<b>Cloud, DevOps &amp; Web:</b>	AWS, Docker, Git, Linux Servers, Command-Line Tools, Flask, Streamlit

**Advanced Coursework:** HPC, Parallel Programming, Unix Systems, Algorithms, Statistics, Mathematical Modeling, Applied Data Analysis, Generative AI, Software Security and Privacy, Cloud Computing

## EDUCATION

<b>University of Chicago</b>	Chicago, IL
MS Computer Science   <i>High-Performance Computing</i>	Sept. 2023 - Mar. 2025
<b>McGill University</b>	Montreal, QC, CA
BS Joint Computer Science & Biology	Aug. 2019 - June 2023

*MUHC Studentship Winner:* Awarded maximum research stipend.

## PROFESSIONAL EXPERIENCE

<b>Elevance Health (Formerly Anthem Inc.)</b>	Chicago, IL
<i>Data Analyst Intern - Machine Learning Projects</i>	May 2024 - Aug. 2024

- Model Migration & Optimization:** Led migration of custom PyTorch GPT-2 model (MediClaimGPT) to Hugging Face ecosystem, standardizing deployments via HF Hub. Maintained 100% text-generation accuracy while improving deployment scalability (via state-dict weight transfer) and automating reproducibility (with save\_pretrained).
- ML Pipeline Development:** Implemented Hugging Face training workflows using Trainer API, refactoring data preprocessing (datasets/tokenizers) and optimizing training loops via TrainingArguments (dynamic batching, LR scheduling).
- Leadership & Communication:** Created and presented comprehensive technical documentation and demos to team members.

<b>McGill University Health Center - Ding Lab</b>	Montreal, QC, CA
<i>Computational Biology Researcher - Machine Learning</i>	May 2021 - May 2024

- Deep Learning Architecture:** Designed and trained a novel hybrid deep learning model (autoencoder + multilayer perceptron) for cancer subtype classification, outperforming baseline models by 10% accuracy through architectural optimization and custom loss functions. Iteratively improved training performance through sensitivity analysis and hyperparameter tuning.
- Feature Engineering:** Discovered novel gene markers by developing a custom loss function and ranking genes by their predictive significance and feature importance, improving interpretability of model outputs.
- Data Processing & Visualization:** Analyzed multi-million record genomic datasets (from cBioPortal, the National Cancer Institute, and the TCGA Database). Evaluated performance through clustering, accuracy metrics, and gene expression analysis.

## COMPLETED PROJECTS ON GITHUB

<b>Parallel Medical Image Processing</b>	Dec. 2024
<b>Code repo:</b> <a href="https://github.com/sonia-sharapova/OpticalFlow">github.com/sonia-sharapova/OpticalFlow</a> - Created high-performance system for cardiac motion analysis of DICOM medical image sequences.	
- Implemented pipeline parallelism, <b>achieving 3.22x speedup</b> with 6 workers (against sequential baseline).	
- <b>Issues Addressed:</b> OpenCV memory management, load balancing, and DICOM frame-order dependencies.	

<b>GPU-Accelerated Ray Tracing System</b>	Mar. 2024
<b>Code repo:</b> <a href="https://github.com/sonia-sharapova/RayTracing_HPC">github.com/sonia-sharapova/RayTracing_HPC</a>	
- Developed a ray tracer for rendering a 3D sphere (processing 1B rays across $1000 \times 1000$ pixel grid).	
- Applied systematic performance analysis (precision tuning, GPU kernel design, and low-level optimizations).	

- **758x speedup** with CUDA (NVIDIA V100) and **13.16x speedup** with OpenMP against serial baseline.