ansh.agrawal.cs@gmail.com
+1 (437) 679 3760

Ansh Agrawal

Machine Learning Engineer | LLMs, Agents & ML Systems

LinkedIn
Website

WORK EXPERIENCE

Software Engineer II, Machine Learning **07/2026 – Present**
Autodesk *Toronto, Canada*
- Building an **MCP**-based agent for tool execution, scheduling, self-correction, & workflow verification.
- Developing & deploying **ML models** to improve product behavior through end-to-end training & evaluation.

Machine Learning Researcher **06/2026 – Present**
University of Toronto *– Prof. Uzma Maroof* *Toronto, Canada*
- Developed a stage-aware TSA-WF method, improving partial-trace closed-world accuracy by **6.6 percentage points** and open-world r-precision by **8.4 points**.
- Built a multi-GPU evaluation pipeline for closed-world, open-world, and traffic-defense experiments.

Machine Learning Systems Researcher **05/2026 – 08/2026**
University of Toronto *– Prof. Bogdan Simion* *Toronto, Canada*
- Built & deployed a **CSC369 CPU scheduling simulator** implementing **FCFS, SJF, STCF, RR, & MLFQ**.
- Analyzed context switches, CPU migrations, and load balancing using **perf, pidstat, trace-cmd, & KernelShark**.

Machine Learning Engineer **05/2025 – 08/2025**
Autodesk *Toronto, Canada*
- Built an **agentic package-management chatbot** with **Rasa**, backend APIs & feature-flagged **JWT auth**, achieving **90%** action accuracy & projected **$6.7M/year** in customer time savings.
- Prototyped **RAG** across **1000+** package configs using **semantic similarity**, improving lookups by **10×**.

Software Engineer **09/2024 – 04/2025**
Thomson Reuters *Toronto, Canada*
- Resolved **20** production issues & migrated **C#/.NET** services, improving platform stability by **15%**.
- Supported **monthly releases** via cherry-picks and blocker triage, helping achieve **98%** on-time delivery.
- Authored **25** SQL scripts for fixes & data corrections across **seven** environments serving **20,000** customers.

LLM Robotics Researcher **08/2024 – 03/2025**
MEDCVR *at UofT* *Toronto, Canada*
- Designed a **multi-agent** robotics application with VLMs & LLMs for autonomous surgical tasks.
- Developed high-level motion planning, low-level action functions & deployed code on **Franka Emika** robot.

PROJECTS

NanoChat LLM Training & Post-Training
- Trained & evaluated a **560M-parameter** NanoChat model on **ClimbMix-400B** using **Modal** & **Weights & Biases**.
- Built a **long-context eval** to test model performance as context length increased from **512** to **2048** tokens.
- Improved **NanoChat** post-training with reward shaping, raising **ChatCORE** from **0.3226** to **0.3681**.

TECHNICAL SKILLS

Languages: Python, Java, C, C#, SQL, JavaScript
Machine Learning: PyTorch, Transformers, Scikit-learn, Fine-Tuning, Post-Training, Reinforcement Learning
AI Agents & Retrieval: MCP, RAG, VLMs, Embeddings, Semantic Search, Vector Search, FAISS, LLM Evaluation
ML Systems & Infrastructure: Linux, KVM, Docker, AWS, Modal, Weights & Biases, FastAPI

EDUCATION

University of Toronto - HBSc, Computer Science (AI focus) **09/2021 – 06/2026**
GPA: 3.85/4.00 *Toronto, Canada*
- ICPC ECNA Regional Competitor & 6x hackathon award winner (PennApps XXV, HackPrinceton, UofTHacks).

Coursework: Operating Systems, Machine Learning, Deep Learning, Computer Vision.

Awards: International Scholar ($180,000), Principal's Entry ($12,000), UofT Scholar ($7,500), 3× Dean's List.

PUBLICATIONS

- A. Agrawal, "From Attention to Application: The Evolution of Transformer Models & Their Impact on Modern Healthcare & Cancer Research," manuscript under review at *J. Mach. Learn. Res.*, 2025.