

# Arnav JAISWAL

+1(469) 604-7909 | [arnavjaiswal149@gmail.com](mailto:arnavjaiswal149@gmail.com) | [www.arnavjaiswal.com](http://www.arnavjaiswal.com) | [GitHub](#) | [LinkedIn](#)

## EDUCATION

**Northeastern University, The Roux Institute, Portland, Maine**

Sept 2024 – May 2026

Master of Science in Artificial Intelligence

**Nanyang Technological University, Singapore**

Aug 2020 – Jun 2024

Bachelor of Engineering in Computer Science with a Second Major in Business and Dual Specialisation in AI and Data Science

**Relevant Coursework:** Computer Vision, Natural Language Processing, Deep Learning, Advanced Software Engineering

## SKILLS

**Programming:** Python | Go | TypeScript/JavaScript | Java | C/C++ | SQL | Flutter

**AI & Machine Learning:** PyTorch | TensorFlow | Ollama | CUDA | OpenCV | LangChain | Groq | E2B | Cohere | Pinecone

**Web Development:** Next.js | Tailwind CSS | FastAPI | Vercel | Firebase

**Cloud & Infrastructure:** PostgreSQL | MongoDB | Supabase | GCP | AWS | Terraform | Docker

## WORK EXPERIENCE

**AI PRODUCT ENGINEER FELLOW** | BURNES CENTER (AI4IMPACT) with CITY OF BOSTON

Sept 2025 – Dec 2025

- Designed and deployed OpenFeedback, a mayor-commissioned AI platform integrating two-tier LLM pipelines with tool-calling and enriching entries with **8** new attributes, reducing manual feedback processing by **70%** in **2** months.
- Architected AWS infrastructure (App Runner, RDS, Amplify) and implemented PostgreSQL & Neo4j to support graph-based visualization and code-free administrative updates, ensuring scalability and modifiability for city-wide deployment.
- Collaborated with Boston's CIO and UX Director to align executive vision with technical delivery; led stakeholder testing and product discovery, ensuring full-stack ownership.

**COMPOUND AI ENGINEER INTERN** | GROQ

May 2025 – Aug 2025

- Developed a **Go SDK** (E2B-based) for computer-use in **Compound**, evaluated alternative providers (Daytona, Browser Use, BrowserBase), and implemented backend support for **Wolfram Alpha** tool and file uploads in Compound.
- Implemented benchmarks including **CRUXEval**, designed custom benchmark for Python library import correctness, and evaluated Compound against **SimpleQA**, **MMLU** and **RealtimeQA** to test its reasoning and web-search ability.
- Developed **interactive charting**, integrated **LaTeX rendering**, and introduced targeted UI/UX improvements, resulting in more reliable functionality and consistent user experience across web and desktop applications.
- Engineered automation and **multimodal** applications including Compound-powered **PR review bot**, a **Claude-based GitHub cross-repo CI/CD pipeline**, **GroqsApp** (WhatsApp bot **4x faster** than Perplexity), and **Groquette** (**real-time** AI meeting moderator).

**ADVANCED AI ALGORITHM DEVELOPER INTERN** | PANASONIC

Jan 2023 – Jun 2023

- Analysed video segmentation techniques proposed in Background Matting V2 and Robust Video Matting (RVM) papers.
- Automated video mask creation process using **python** and HSV values, resulting in a **95%** reduction in processing time.
- Leveraged custom data augmentation methods like motion jitter, resulting in **5%** accuracy improvement.
- Evaluated alternative models such as **Meta's SAM**, **YOLOv8**, **resnet50/101**, semantic guided human matting and replaced RVM model's backbone with YOLOv8's and resnet101's to achieve a **10%** improvement in segmentation performance.

**ROBOTICS SOFTWARE INTERN** | OTSAW

May 2022 – Jul 2022

- Conducted in-depth analysis of proprietary industrial robot fleet management simulation system to understand architecture, functionality, and implementation details
- Rectified a **300+** lines robot job-generation script to handle identical job source problem and faulty job overriding.
- Built **Django** WebSocket server bridge for **C++** clients to modernize software while following agile methodology streamlining tasks and tracking daily progress through Slack and Jira.

## PROJECTS

**Trending AI Papers** [[Next.js](#) | [Groq](#) | [OpenAI](#) | [Pinecone](#) | [Ollama](#) | [Python](#) | [FastAPI](#) | [Cohere](#) | [MongoDB](#) | [Vercel](#)] **Jan 2025 – Mar 2025**

- Built **Vercel**-hosted website that fetches and generates structured summary of arXiv research papers using **Python/FastAPI** backend, **Ollama** with Llama 3.2-3B and DeepSeek-R1 for offline summarization, implemented **RAG**-based chat powered by **Groq (Langchain)** using **OpenAI** for chunked embeddings, **Pinecone** for vector storage and retrieval and **Cohere** for reranking.

**Tennis Pose Estimation, Detection and Classification** [[Python](#) | [PyTorch](#)]

Oct 2024 – Dec 2024

- Researched tennis player detection, pose estimation, and shot classification using **YOLOv11-Pose** (87% IoU, 62 MSE, 99% accuracy) and custom models (65% IoU, 94 MSE, 99% accuracy).