Mahatav Arora

250-681-6810 | mahatavarora@gmail.com | linkedin.com/in/mahatav-arora | github.com/Mahatav

Summary

Computer Science student with a Data Science minor at UBC Okanagan, experienced in AI/ML, full-stack development, data pipelines, and applied research. Skilled in building Django/Docker systems, LLM workflows, GitHub analytics tools, and web solutions with practical experience across university research, startups, and industry projects.

Experience

AI/ML Intern Jun. 2024 – Jul. 2024

Ayeyarvaddy Farmers Development Bank *Yangon, Myanmar*

- Prototyped facial-recognition and eKYC workflows using Python, OpenCV, dlib facial landmarks, YOLOv8 face detection, and head-pose estimation for banking identity-verification research.
- Built data-capture and labelling scripts to collect face images, generate bounding-box labels, split train/test/validation datasets, and train custom YOLOv8 models for face-verification experiments.
- Prepared an implementation proposal covering open-source model options, integration risks, privacy/security considerations, and deployment steps for a Myanmar banking context.

Web Developer & Digital Content Lead Apr. 2025 – Apr. 2026

Green Construction Research & Training Center, UBC *Kelowna, BC*

- Redesigned and maintained a multi-page Astro website for a UBC-affiliated research centre, improving access to research, events, funding, metrics, news, team, and contact content.
- Built a custom site search feature using Astro page indexing and client-side filtering to improve navigation across research pages, events, faculty profiles, and centre updates.
- Fixed the Contact Us page and integrated supporting PHP functionality to improve form handling, submission reliability, and coordination between front-end pages and server-side scripts.
- Maintained the Wilden Living Lab website for a UBC Okanagan net-zero housing and sustainable construction research initiative.

Web Developer & Digital Systems Consultant Apr. - Sep. 2025 & 2026

OKGN Boat Rentals *Kelowna, BC*

- Rehired across two summer contracts to modernize a local boat rental website whose booking flow, mobile layout, and search visibility needed improvement before peak season.
- Rebuilt customer-facing pages in Astro, HTML, CSS, and JavaScript, including landing, fleet, reviews, FAQ, crew, and booking sections to improve navigation and service clarity.
- Developed a multi-step booking request system with date/time selection, vessel choice, passenger details, validation, spam protection, rate limiting, and PHP email handling to streamline booking inquiries.
- Implemented local SEO improvements including metadata, canonical URLs, Open Graph tags, JSON-LD schema, robots.txt, and dynamic sitemap generation for better search engine readability.

Freelance Web Developer Apr. - Sep. 2026

Channeling Skinny Bob *Kelowna, BC*

- Built a custom alien-themed website for a YouTube creator, turning supporting videos, documents, audio, and written lore into an interactive web archive for viewers.
- Developed a retro terminal interface in HTML, CSS, and JavaScript with command parsing, autocomplete, access levels, animated output, hidden commands, and file-opening workflows.
- Integrated WebGL assets, Rive animations, video, custom fonts, and responsive styling to create a polished sci-fi landing experience aligned with the creator's brand.
- Configured Python static serving, Docker, and Nginx routing to support deployment and ensure media, animation, and 3D assets loaded correctly in hosted environments.

Projects

GPT-2 From Scratch | *Python, PyTorch, NumPy, Transformers, BPE Tokenization* 2026

- Implemented a GPT-2-style language model from scratch, including token and positional embeddings, causal multi-head attention, transformer blocks, MLP layers, weight tying, and AdamW optimization.
- Built tokenizer, training, generation, checkpointing, and evaluation utilities with support for character-level and BPE tokenization, gradient accumulation, cosine warmup scheduling, and top-k/top-p sampling.
- Added 100+ tests covering tokenizer round trips, causal masking, tensor shapes, gradient flow, scheduler behaviour, checkpoint save/load, and generation utilities.

TA Application System (Capstone) | *Django, MySQL, FastAPI, Ollama, nginx* May 2025 – Aug. 2025

- Integrated an Ollama-backed FastAPI resume extraction service with the Django backend, enabling PDF resume uploads to return validated technical and soft skills for TA applications.
- Connected AI-extracted skills into the TA application flow, updating frontend experience and confirmation pages so extracted skills were displayed consistently and stored with submitted applications.
- Built and refactored TA filtering functionality for coordinators, including course-taken, past TA experience, skills, program type, and year-of-study filters with backend serializer fixes and unit test coverage.
- Configured Docker and nginx infrastructure for local development/testing across React, Django, MySQL, and the AI resume API, including container health checks, service routing, and environment-based Ollama connection setup.

Autonomous Drone Person Tracking (Hackathon) | *Python, YOLOv8, OpenCV, djitellopy* Jan. 2024

- Built a real-time drone tracking prototype using YOLOv8 and OpenCV to detect a person from the camera feed and map their location to a 3x3 movement grid.
- Implemented autonomous Tello-style drone control with `djitellopy`, converting detection quadrants into rotation, altitude, and directional follow commands.
- Added supporting prototypes for live camera streaming, CUDA/GPU checks, and MediaPipe hand gesture detection for alternate control input.

Computer Vision Spider Bot | *Python, YOLOv7, OpenCV, PyTorch, ESP32-CAM, Arduino* Mar. 2024

- Built a vision-based tracking prototype using YOLOv7, OpenCV, and BYTETracker to detect and follow a selected person from a live camera feed.
- Connected detection output to ESP32-style movement control, sending HTTP/socket commands based on bounding-box size and target position.
- Configured ESP32-CAM video streaming and CUDA/PyTorch setup for real-time object detection experiments.

Autonomous Landing Pad Detection | *Python, OpenCV, DroneKit, NumPy* Apr. 2023

- Built a UAV landing-pad detection prototype using OpenCV, HSV masking, image moments, and Hough circle detection to identify pad position from live camera footage.
- Added OCR-based character detection with Tesseract and drafted DroneKit navigation logic to translate camera offsets into landing-position corrections.
- Created HSV calibration tools to tune colour thresholds and improve landing-pad detection reliability across lighting conditions.

Research

Directed Studies: Corpus-Conditioned LLM Pre-Training Jan. 2026 – Apr. 2026

DATA 448 — Supervisor: Dr. Bowen Hui, UBC *Kelowna, BC*

- Trained two GPT-2-style causal language models from scratch on Eastern and Western philosophical corpora to measure how pre-training data shapes model behaviour and bias.
- Built an end-to-end PyTorch and Hugging Face Transformers pipeline for tokenization, fixed-length block generation, custom language-model datasets, padding collators, attention masks, and next-token labels.
- Designed a progressive training workflow across 21 chronological corpus periods, producing reusable checkpoints, training logs, and evaluation artifacts for each model.
- Evaluated model divergence across 284 prompts and 14 categories using cross-perplexity, concept-marker frequency, type-token ratio, repetition rate, and distributional-overlap metrics.

Directed Studies: Human vs. AI Code Analytics Jan. 2026 – Apr. 2026

DATA 448 — Supervisor: Dr. Bowen Hui, UBC *Kelowna, BC*

- Developed an AI-code detection pipeline to analyze student capstone repositories and estimate whether source files were likely human-written, AI-generated, or uncertain.
- Fine-tuned CodeBERT-based classifiers using GPTSniffer and CoDET-M4 methodology, producing binary human-vs-AI code classification models for repository-level analysis.
- Built repository scanning scripts that chunked long files, filtered generated and third-party code, skipped boilerplate, and exported per-file and team-level CSV summaries.
- Automated GitHub organization cloning and analysis workflows using shell scripting, Python, Hugging Face Transformers, PyTorch, and GitHub access tokens.

Directed Studies: Software Process Analysis Sep. 2025 – Apr. 2026

COSC 447 — Supervisor: Dr. Bowen Hui, UBC *Kelowna, BC*

- Built a Python analytics pipeline to extract and study collaboration patterns from GitHub Classroom repositories, including pull requests, commits, reviews, branches, and file changes.
- Implemented GitHub REST API extraction with pagination, retry logic, rate-limit handling, request tracking, and authenticated data collection across multi-team repositories.
- Created preprocessing and labelling utilities for PR communication, code structure, branching behaviour, bot filtering, anonymization, and cleaned CSV generation.
- Generated process-model visualizations using pandas, NetworkX, Graphviz, clustering, and statistical summaries to compare team development workflows.

Education

University of British Columbia **Jan. 2023 – Apr. 2026**

Bachelor of Science in Computer Science, Minor in Data Science

Technical Skills

Languages: Java, Python, C/C++, JavaScript, HTML, CSS, R, Assembly
Frameworks: Astro, WordPress, Material-UI, FastAPI, Spotify API, Ollama
Developer Tools: Git, VS Code, PyCharm, IntelliJ, CUDA, Docker, GitHub API
Libraries: Pandas, NumPy, Matplotlib, OpenCV, Djitellopy, PyTorch, Requests, TQDM