Adnan Aidan Taha

aidantaha.com | linkedin.com/in/adntaha | github.com/adntaha

Education

McGill University **Jun. 2029**

Joint Honours B.Sc. in Computer Science & Statistics *Montreal, Canada*

- AI Safety Fundamentals Fellow (MAIA, MIT AI Alignment).
- Co-director of AI Alignment McGill, facilitating AGI Strategy reading groups and running topical events.
- **Award:** University of Waterloo President's Scholarship of Distinction (Software Engineering admit).

Experience

Harvard University **Sep. 2026 – present**

Undergraduate Researcher *Remote*

- Evaluating sovereign LLMs for various biases under Prof. Juncheng Yang.
- Reproduced an LLM censorship benchmark from its paper and extended it to **13** models, showing judge choice moves results **5–10**× more than model choice, via an audited LLM-as-judge pipeline.

Horizon Omega **Fall 2025 – present**

Core Member, AI Safety Montreal *Montreal, Canada*

- Authored the starter codebase used by **15+ teams** at our inaugural hackathon, introducing participants to the libraries that they will use (Control Tower and inspect_ai).
- Organized reading groups and workshops held on AI safety topics that vary from papers to policy drafts.

University of Basel **Jul. 2024 – Apr. 2025**

Research Assistant *Remote*

- Trained and validated **U-Net** models for biomedical image segmentation, running grid search over learning rate and epoch count to optimize performance.

Holopin **Oct. 2022 – Nov. 2023**

Software Developer & Integrations Lead *Contract, Remote*

- Developed key integrations with GitHub, Slack, and Discord using **TypeScript** and **Node.js** from scratch for major potential clients such as **Cloudflare** or **Hacktoberfest**, and Salesforce, serving **50,000 users** in their first month.
- Authored technical documentation and articles consulted by **500+ organizations** using Holopin's platform.
- Contributed to open source projects, notably LLaMA.cpp with **120,000 GitHub stars** by resolving GitHub issues.

BrébeufHx **Aug. 2024 – May 2026**

Co-President

- Led a **150+ participant** hackathon, synchronizing **20+ organizers**, and secured **over $12,000** in sponsorship value.

Projects

METR Reward-Hacking Challenge **Mar. 2025**

Ranked top 4 globally, responsibly disclosed a platform security issue

Telos: a minimal self-improving harness [GitHub] **Aug 2026 – Sep 2026**

OpenRouter, Harness Engineering, Python

- Built a minimal **agent harness** with Python as its only tool, allowing the model to execute code and spawn copies of itself.
- Enabled the agent to modify its own harness, edit its system prompt and restart to load changes, supporting **(self-)experiments in harness engineering**.

McGill Course Assistant & Graph [GitHub] [Live] **May 2026 – present**

HTML/CSS, JavaScript, Cohere API, Cloudflare Workers, Vectorize & D1, GitHub Pages

- Created a **RAG pipeline** on Cohere Embed → Vector DB → Rerank → Command A+ (high-recall retrieval, query-specific reranking, grounded generation), run through Cloudflare's **AI Gateway** for caching & rate-limiting.
- Paired it with an interactive force-directed prerequisite graph of **5,800 nodes** using a Barnes-Hut quadtree for a **109x speedup**.

Skills

Base skills: Python, TypeScript/JavaScript (← advanced), HTML/CSS (← intermediate), Java, SQL, GitHub, Cloudflare, Firebase

Frameworks: Next.js, React (← intermediate), TensorFlow, PyTorch, Linux, CI/CD (← intermediate), Control Tower, Inspect AI (← beginner)