Gatlen Culp

📍 Cambridge, MA ✉ gculp@mit.edu 🔗 gatlen.me ⬡ GatlenCulp 🎓 Google Scholar

Education

BS **Massachusetts Institute of Technology** **Sept 2021 – Dec 2025**
Computer Science with a concentration in AI & Decision Making Cambridge, MA

Experience

🏛 **AI Governance Research Fellow** **June 2025 – Aug 2025**
ERA · AI State Entrenchment (WIP) · Analysis of LLMs in Gov · Math for Power Analysis (Draft) Cambridge, UK
Conducted independent research on the intersection between AI, government, and power. Core work examined how AI systems can entrench existing state power and resist institutional reform. Additional projects analyzed current LLM deployment in government and developed mathematical frameworks for power modeling.

📊 **AI Evaluations & Software Engineering Contractor** **Sept 2024 – present**
Model Evaluation and Threat Research (METR) · Vivaria · Homebrew · Eval Templates Remote
Developed CLI tools, evaluation templates, and installers for Vivaria, a platform used by METR to conduct AI capability and risk evaluations in partnership with OpenAI, Anthropic, and US/UK AI safety institutes. Enhanced documentation for rapid adoption. Investigated integrating SWE-Agent and UK AISI's Inspect framework

🐛 **DevOps & Full Stack Engineer** **June 2024 – Aug 2024**
MIT Computational Biology Lab · MantisAI Website Cambridge, MA
Engineered NetworkX-based graph visualization system, DevOps infrastructure, and testing suite for MantisAI, a Django web platform for interactive text and data analysis. Architected CI/CD pipeline with GitHub Actions, Docker, and AWS deployment. Led PostgreSQL migration, identified critical security vulnerabilities, and lead a project-wide codebase and documentation refactor.

📄 **Director of Operations** **Feb 2023 – Nov 2024**
MIT AI Alignment (+ Harvard AI Safety Team) · MopMan · Public Resources · GatPack Cambridge, MA
Led operations for student organization across 300+ people and 36+ programs. Created MopMan and GatPack automation tooling saving 7 hours/week across 3+ AI safety groups. Halved expenses via office relocation. Managed facility logistics, interior design, digital infrastructure, communications, and program operations.

🖲 **Machine Learning Researcher** **June 2022 – Oct 2022**
Stanford Existential Risks Initiative & MIT Algorithmic Alignment Group · CommonClaim Berkeley, CA
Developed RL fine-tuning technique for GPT-2 to autonomously identify diverse prompts resulting in harmful model outputs (violence, disinformation, etc.). Built evaluation pipelines and created survey resulting in the CommonClaim dataset of 20,000 boolean statements. Paper - "Explore, Establish, Exploit- Red Teaming Language Models from Scratch".

🌐 **Full Stack Web Developer** **May 2022 – June 2022**
MIT's Civic Data Design Lab (CDDL) Cambridge, MA
Finalized UN Heritage Website and WhatsApp interface in under three weeks. Implemented contribution web-form using React, SASS, REST API, and Node. Created WhatsApp informational messaging interface with Twilio. Developed image compression script to improve MongoDB load times.

Misc

Professional Certifications: AWS Certified Cloud Practitioner (Dec 2023 - Dec 2026), Autodesk Certified User - Inventor, OSHA 10-Hour General Industry (Feb 2020)

Notable Projects: Linear Embedding Alignment – 6-day research sprint measuring similarities between text embedding model representations · OpenGesture – OpenCV app for custom keyboard/mouse inputs · Transformative AI – Technical article on AI trajectory and impacts · Personal Blog – Public thought pieces · GOTem – Dynamic python project template for researchers, engineers, and students