Jarod Bloch

Pittsburgh, PA • 484-550-5079 • jarodb@andrew.cmu.edu • linkedin.com/in/jarodbloch • jarodblo.ch

Education

Carnegie Mellon University, School of Computer Science May 2027
Bachelor of Science, Artificial Intelligence Pittsburgh, PA
GPA: 3.3, Dean's List: Fall 2024 (High Honors), Fall 2025 (High Honors)
Awards: 3rd Place, CASI × Gray Swan AI Jailbreaking Hackathon (Fall 2025); 6th Place, top solo vs. teams of 2–3 (Spring 2026)
Member, **Plaid Parliament of Pwning (PPP)**, multi-time DEF CON CTF champion team, consistently world top-ranked
Coursework: ML for Cybersecurity (grad), Generative AI, Autonomous Agents, Natural Language Processing (grad)

Publications

Cheng, Y. F.*, **Bloch, J.***, et al. "Auditorily Embodied Conversational Agents: Effects of Spatialization and Situated Audio Cues on Presence and Social Perception." *Published in Proceedings of ACM CHI 2026*. https://doi.org/10.1145/3772318.3791794
***(Co-first author; co-presented oral talk at ACM CHI '26)**

Experience

Connected Vehicle Software Engineering Intern | Python, LangGraph, RAG, BigQuery, Pandas May 2026 - August 2026
Ford Motor Company Dearborn, MI

• Building an agentic framework (Python, LangGraph) that orchestrates Ford's internal LLM platform (FordLLM) to predict severity classifications for vehicle diagnostic trouble codes (DTCs), helping the triage team prioritize EV issues.

• Designing the RAG pipeline (diagnostic documentation and live EV telemetry via BigQuery, Pandas) so the agent reasons from objective evidence rather than prior labels, establishing a consistent ground-truth for a previously subjective process.

Research Assistant | Python, Docker, vLLM May 2025 - May 2026
Carnegie Mellon University Pittsburgh, PA

• Evaluated LLM robustness for vulnerability-detection tasks under Prof. Lujo Bauer (CyLab), measuring how 100+ adversarial input transformations degraded model recall and surfaced systematic failure modes.

• Optimized a containerized vulnerability benchmarking framework, increasing self-hosted LLM inference throughput by 8.6x.

Teaching Assistant | C, Google Apps Script May 2025 - August 2025
Carnegie Mellon University Pittsburgh, PA

• Served as head of code reviews for 220+ primarily graduate-level students in Computer Systems (15-213/15-503), designing rubrics and coordinating sessions to evaluate C-based memory allocators, proxy servers, and shell programs.

• Built an automated TA management pipeline in Google Apps Script (roster imports, tracking, statistics) whose auto-generated mailing lists replaced manual lookup, saving the 13-person staff ~3 hrs/week and adopted as infrastructure for future semesters.

Projects

Repo-Level Type Inference | Python, CodeQL, NetworkX, vLLM, Slurm February 2026 - May 2026

• Built a language-agnostic neuro-symbolic type-inference framework (Qwen3 + CodeQL) annotating repository-scale codebases, achieving 0.84 median type-match score (0-1) across 50 repos while matching 96% of a model 13× larger.

• Reduced cross-file type annotation errors by 37% by mapping dependency DAGs to inject explicit type signals into context.

• Optimized annotation quality via large-scale Slurm GPU ablation studies on 50 real-world repositories across Qwen3 model sizes (0.6B–30B), evaluating seeded root nodes, model types (Coder vs. Instruct), and context augmentation strategies.

Indirect Prompt Injection Portfolio Site | HTML, CSS, JavaScript February 2026

• Designed and deployed a multi-vector indirect prompt injection demonstration exploiting DOM text extraction, image alt-text, and metadata pipelines using invisible Unicode obfuscation to test tokenizer resilience (OWASP LLM01).

• Deployed an injection payload that caused 4 leading LLMs (Claude, ChatGPT, Gemini, Copilot) to execute benign hidden instructions in the page to educate end-users when asked to summarize or analyze it, highlighting systemic weaknesses.

Terrabot | Python, ROS, PyTorch, ONNX, OR-Tools August 2025 - December 2025

• Built a layered autonomous control agent (4-person team, Raspberry Pi) enforcing hard operational constraints via an OR-Tools CP-SAT planner with rules to prevent conflicting actuator commands across 7 concurrent behaviors, regenerated daily.

• Added runtime self-monitoring through 13-channel redundant-sensor cross-validation and a SAT-based fault-diagnosis layer; deployed a color-calibrated UNet vision pipeline (~97% pixel accuracy) chosen over a higher-accuracy ViT that exceeded RAM.

Skills

AI Security & Red Teaming: Adversarial ML, Prompt Injection, RAG Poisoning, MITRE ATLAS, OWASP LLM Top 10, garak, PyRIT
Offensive Security & AppSec: Reverse Engineering, Binary Exploitation, Penetration Testing, SAST/CodeQL, Threat Modeling
ML & Evaluation: PyTorch, Hugging Face, vLLM, LoRA/PEFT, DPO, RAG, Agentic Systems, Capability Elicitation, Classification
Programming Languages: Python, C, C++, Java, JavaScript/TypeScript, C#, SML, SQL
Infrastructure & MLOps: Docker, Slurm/HPC, AWS, Oracle Cloud, W&B, FastAPI, Linux/Bash, Git