

Benjamin Wu

Chicago, IL • benwu408@gmail.com • (224) 501-8155 • github.com/benwu408 • linkedin.com/in/benwu408

Education

University of Illinois Urbana–Champaign (UIUC) Champaign, IL
B.S. Computer Engineering Expected May 2028 • GPA: 3.7
Relevant Coursework: Computer Systems, Data Structures, Linear Algebra, Statistics, Differential Equations
Upcoming: Probability with Eng. Applications, Parallel Programming, Analog Signal Processing, Discrete Math

Experience

Mechanize, Inc. — Software Engineering Intern May 2026 – Present
Environments and evaluations for frontier models — backed by Nat Friedman, Patrick Collision, Dwarkesh Patel

- Built **simulated user tool calling** into the agent evaluation harness, allowing persona LLMs to **modify the task environment mid-conversation** and enabling realistic user behavior for evals
- Designed **behavioral alignment evaluations** for LLMs, helping frontier labs train out failure modes
- Developed **self-recognition tasks** measuring whether models can identify their own behavioral failures, supporting frontier work on model self-correction and trustworthiness
- Authored and validated grading rubrics spanning 20+ criteria across task families for LLM-based scoring, **closing reward-hacking loopholes** and achieving **<2% grader variance** across runs

Projects

Onyx Canvas Connector Mar 2026 – June 2026
Merged connector for Onyx, an **open-source (30k+ stars)** enterprise RAG search platform

- Built Canvas connector for Onyx enabling low-latency incremental ingestion and retrieval over user documents
- Designed checkpointed ingestion pipeline with resumable pagination and early-exit filtering for large-scale sync
- Implemented permission-synched indexing (enrollments → ACLs) with caching and fail-safe privacy defaults

Bio Fury — Autonomous Drug Discovery Agent Feb 2026 – Mar 2026
HackIllinois Winner (Social Impact) | YC Bio AI Hackathon — Autonomous target-to-lead drug discovery platform

- Engineered multi-agent orchestration framework with **11 tool integrations** and interrupt-driven control
- Designed ensemble validation pipeline with consensus across **9 ML models** and ChEMBL-grounded scoring
- Architected FastAPI/PostgreSQL backend with real-time streaming and interactive execution interface

NBA Win Probability Engine Jan 2026 – Mar 2026
End-to-end ML system modeling real-time NBA win probability

- Built full data ingestion and replay pipeline processing **2.04M+** events with leakage-safe train/validation splits
- Designed evaluation framework; trained logistic regression achieving **76.6% accuracy** (Brier 0.1587)
- Implemented event-driven simulation engine serving live win probabilities at sub-play granularity via Flask

Skills

Languages: Python, C++, Java, C, SQL, TypeScript
Backend & Systems: FastAPI, Flask, Node.js, REST APIs, WebSocket, PostgreSQL, Docker, Git, Linux
ML & Data: NumPy, pandas, scikit-learn, FAISS
Frontend: React, Next.js, Streamlit

Awards & Honors

HackIllinois 2026 — Best Social Impact Feb 2026
Built Ligands.tech (precursor to Bio Fury), an autonomous drug discovery agent

National Merit Scholar Mar 2025
Perfect PSAT Score; selected among top ~0.5% of 1.3M+ test takers nationwide

Eagle Scout — Boy Scouts of America April 2023
Highest rank in Scouting, awarded for leadership, service, and long-term project execution