

# Zachary Burton

617-417-1070 | [zburton@mit.edu](mailto:zburton@mit.edu) | [linkedin.com/in/zac-burton](https://www.linkedin.com/in/zac-burton) | [github.com/zacn04](https://github.com/zacn04)

## EDUCATION

### Massachusetts Institute of Technology (MIT)

Cambridge, MA

*B.S. in Mathematics (GPA: 4.6/5.0)*

Feb 2026

**Coursework:** Probability, Statistics, Real Analysis, Abstract Algebra, Reinforcement Learning (Grad), Program Synthesis (Grad)

## PUBLICATIONS

**Evaluation framework for LLM mathematical proof generation** (with Yale) | Benchmark suite measuring correctness, step validity, and generalization across proof strategies | Under review at ICML 2026

**Structured Hints for Sample-Efficient Lean Theorem Proving** | Under review at ICML 2026 (arXiv:2601.16172)

**Empirical study on AI disclosure effects on belief formation** (n=174) | [zacburton.com/polisci.pdf](https://zacburton.com/polisci.pdf)

## EXPERIENCE

### Meta

Feb 2026 – Present

*Machine Learning Engineer*

New York, NY

- Designed offline experiments to measure revenue impact of multimodal ranking signal changes in the Reels late-stage pipeline; instrumented metrics to surface tradeoffs between revenue lift and engagement quality
- Proposed architectural change to the core value model to enable more granular signal-level debugging and backtesting; designed evaluation spec and presented to team

*Software Engineer Intern*

May 2025 – Aug 2025

- Refactored copyright enforcement engine processing 2B+ daily uploads; achieved 97% consistency with legacy outputs, verified via SQL logging and automated regression suite
- Uncovered and fixed 16% false-positive error affecting 3 internal teams; designed data quality checks in Hack/PHP to detect and surface annotation inconsistencies

### MIT Computer Science & Artificial Intelligence Lab

Nov 2024 – Aug 2025

*Research Engineer*

Cambridge, MA

- Implemented reproducible data collection pipelines for human-in-the-loop experiments; built tooling to audit feedback quality, detect annotator inconsistencies, and version evaluation datasets
- Built **PIEFACE**, an end-to-end human feedback platform: annotation web UI, behavioral trace parser, and SQL data pipeline for collecting, storing, and joining labeled human preference data at scale
- Designed grading and evaluation infrastructure to measure alignment between learned reward signals and human judgments; integrated reward model training into distributed PPO rollouts in a custom simulation environment

## SELECTED RESEARCH & ENGINEERING PROJECTS

**arxivstagram** | *Next.js, PostgreSQL, Prisma, Vercel* | [live](#)

2025

- Built a personalized research feed that infers user preferences from implicit behavioral signals (reading time, engagement patterns) without explicit search or input; ranking driven by a learned user model updated continuously from interaction data
- Recommendation engine uses exponential decay, asymmetric feedback weights, implicit engagement boosting, and novelty jitter to prevent filter bubbles; daily RSS ingestion with keyword extraction at write time

**zOS: Adaptive Learning Environment** | *Rust, Tauri, Svelte, Ollama*

2025

- Designed behavioral skill-tracking engine with Bayesian difficulty calibration and automated rubric generation—a personalized evaluation system that adapts to individual learning trajectories over time
- Built multi-strategy structured output pipeline with JSON extraction and repair chains; optimized precomputation against navigation patterns, reducing API calls by 50%

## SKILLS

**Languages/Systems:** Python, C++17, Rust, Lean, SQL, Docker, Linux, Bash

**ML/AI:** PyTorch, PPO, RLHF, human preference annotation systems, behavioral metric design, reproducibility tooling