

# Jonathan Liu

✉ [jonathanliu@princeton.edu](mailto:jonathanliu@princeton.edu)  
🌐 [liu-jonathan.com](http://liu-jonathan.com)  
🔗 [liujonathan24](https://github.com/liujonathan24)  
in [28-jonathan-liu](https://www.linkedin.com/in/28-jonathan-liu)

## Education

2024–2028 **A.B. Computer Science, Princeton University**  
**GPA:** 3.93  
**Minors:** Mathematics, Statistics & Machine Learning, and Computational Biology  
**Relevant Coursework:** (\*denotes graduate course) Systems & ML\*, Natural Language Processing, Reinforcement Learning\*, AI for Formal Methods\*, Operating Systems, Data Structures and Algorithms, Probability & Stochastic Systems, Multivariable Analysis, and Linear Algebra.

## Research Interests

Adapting foundation models to specialized scientific, clinical, and game-based domains where domain knowledge and verifiability matter.

- Building synthetic Reinforcement Learning environments for specialized agents and tasks.
- Leveraging language-model knowledge as a source of verifiable data and RL rewards.
- Designing efficient systems for large-scale training and inference.

## Publications

- [1] Sachin Konan, **Jonathan Liu**, Zhuang Liu. “PaperLens: Finding Acceptance Signal in Papers.” *In review*, 2026.
- [2] **Jonathan Liu\***, Kia Ghods\*. “Continuous Diffusion Transformers for Designing Synthetic Regulatory Elements.” *ICLR Generative AI in Genomics Workshop*, 2026.
- [3] **Jonathan Liu**, Haoling Qiu, Jonathan Lasko, Damianos Karakos, Mahsa Yarmohammadi, Mark Dredze. “Demo: Statistically Significant Results on Biases and Errors of LLMs Do Not Guarantee Generalizable Results.” *NeurIPS Generative AI for Health Workshop*, 2025.
- [4] Abhishek Joshi, Beining Han, Jack Nugent, Yiming Zuo, **Jonathan Liu**, Hongyu Wen, Stamatis Alexandropoulos, Tao Sun, Alexander Raistrick, Gaowen Liu, Yi Shao, Jia Deng. “Infinigen-Sim: Procedural Generation of Articulated Simulation Assets.” *CoRL Learning to Simulate Robot Worlds Workshop*, 2025.

## Professional & Research Experience

- 05/2026– **ML Research Intern, Abridge, San Francisco, CA**  
Present
  - Training foundation models for clinical decision-making on a fleet of 256 B200 GPUs.
  - Fine-tuning agents using custom-built medical RL environments and distillation algorithms.
- 05/2026– **RL Residency, Prime Intellect, San Francisco, CA**  
Present
  - Re-architected the 150K-line NetHack C engine from global to per-instance state, raising throughput 33×, and enabling live Monte-Carlo simulations.
  - Designed 17 live generation knobs for curriculum learning algorithms.
- 08/2025– **Research Assistant, Prof. Zhuang Liu Group**  
Present
  - Fine-tuned LLMs for paper-acceptance prediction using LlamaFactory and TRL, with DeepSpeed for tensor parallelism and vLLM for inference [1].
  - Analyzed multimodal LLM behavior against ablated architectures in JAX to isolate the contribution of individual design choices.

- 01/2026– **SPAR Research Fellow**, *SPAR Research*
- 06/2026 ○ Researching Specialized Learning with Interpretable Component Experts (SLICE), a continual-training methodology that incrementally transforms pre-trained dense LLMs into a modular mixture-of-experts architecture.
- 06/2025– **Machine Learning Research Intern**, *BBN Technologies*, Cambridge, MA
- 08/2025 ○ Demonstrated that LLM-based evaluation on question datasets can yield statistically significant but non-generalizable results [3].
- Detected hallucinations in clinical LLMs using agentic RAG and LLM-as-a-judge pipelines built with AutoGen, LangChain, and FAISS.
- 08/2024– **Research Assistant**, *Princeton Computer Vision Lab (Prof. Jia Deng)*
- Present ○ Contributed to the Infinigen-Sim framework for procedural generation of articulated assets, improving zero-shot sim-to-real success rates by 70% [4].
- Reduced instability in complex Genesis and MuJoCo scenes via convex hull decomposition implemented in C.

## Projects

- 2026 **NetHack RL Environment**  
RL training on a custom NetHack fork: fine-tuned a Qwen3-VL agent that surpassed the NetPlay GPT-4 baseline on mean maximum dungeon depth (3.0 vs. 2.6)
- 2025 **RL for the Andrews–Curtis Conjecture**  
Applied RL (A3C, PPO) and contrastive learning to solve long-horizon tasks in group theory.
- 2025 **Discovering Transformer Circuits**  
Replicated edge pruning on GPT-2 to identify sparse computational subgraphs for induction mechanisms.
- 2024–2025 **Sight Support**  
Developed an assistive app using multimodal LLMs (Faster-Whisper, LM Studio) to detect social cues in online meetings via “earcons.”

## Awards & Honors

- 2026 SPAR Research Fellow
- 2025 Best Poster Award Winner (Top 4 of 60+), MASC-SLL
- 2025 Qualifier, ICPC Regional Round
- 2024 1st Place Winner (\$2,500), Amazon/MIT #HackDisability
- 2024 Presenter, Yale Engineering and Science Symposium
- 2020–2025 4x AIME Qualifier

## Service & Teaching

- Reviewer 2026 ICLR Generative AI for Genomics Workshop
- Section Editor Princeton AI Review: Technical Section
- Teaching MAT 216/218: Multivariable Analysis and Linear Algebra I & II

## Talks

- 08/2025 Statistically Significant Results On LLM Biases (RTX BBN)
- 06/2026 Progress in Optimizers (Abridge)
- 06/2026 Statistically Significant Results On LLM Biases (Abridge)

- 07/2026 Tool-Call-Optimized EHR Agent (Abridge)
- 07/2026 LLM Algorithms for NetHack (Prime Intellect)

---

## Skills & Interests

- Technical Python (PyTorch, JAX, Pandas, NumPy, Hydra, SciPy, Gym), C, CUDA, Triton, ASM.
- Leadership Association for Computing Machinery (ACM) Officer; Princeton CV/NLP Reading Groups; Assistant Director, Business Today International Conference.
- Interests Classical Cello (Princeton University Orchestra); Distance Running (Jersey City Marathon '25).