

JAEHO LEE

(518) 847-1315 • jaeho_lee@brown.edu • www.linkedin.com/in/jaeho-lee-brown

WORK EXPERIENCE

- Jan 2026 – Current **Constellation**, Talent Fellow
- Reviewing applications for Anthropic's AI Safety/Security Fellows program
- Aug 2025 – Current **Epoch AI**, Associate Data Analyst
- Writing [research blogposts](#) on AI trends, 200,000+ views across platforms
 - Maintaining Epoch's Inspect-based benchmarking repo, running experiments with FrontierMath token limits and enabling web-search tool usage in tasks
- Sep 2025 – Current **Forecasting Research Institute**, Research Assistant
- Analyzing data from forecasting RCTs through statistical programming
 - Leading development of CivBench, a contamination-proof benchmark for LLM forecasting within simulated worlds

PROJECTS

- Dec 2025 – Current **MABEL (Multi-Armed Bandit Evaluation Lattice)**
- Building multi-turn benchmark to measure deceptive and manipulative behavior in multi-agent LLM systems
 - Funded by [Cosmos Institute](#) through a General Track grant
- Apr 2025 – Dec 2025 **AssertBench**
- Accepted at 2025 NeurIPS LLM Eval Workshop with no research advisor
 - Constructing novel sycophancy benchmark for factual verification
- Nov 2024 – May 2025 **Humanity's Last Exam**
- Top 3 (out of 1000+) dataset contributor
 - Wrote 50+ questions, audited 140+ questions for final dataset

Activities

- AI Security Bootcamp (Conjecture)
- Summer Program for Applied Rationality and Cognition (SPARC), Junior Counselor
- Founder, Brown Forecasting Club
- Assistant Tournament Director, Brown University Math Olympiad (BrUMO)

Awards

- Silver Medal, Bridgewater/Metaculus Forecasting Tournament (Top ~1%, 16th/1400)
- Cosmos Institute, General Grant (\$4000)
- Brown AI Safety Team, Research Funding
- FABRIC, European Summer Program in Rationality (ESPR) 2023 invitee

EDUCATION

Brown University, Sc.B, Statistics & Computer Science (2024-2028)

Coursework: Statistical Inference I/II, Principles of Biostatistics and Data Analysis, Applied Regression Analysis; Scientific Computing, Probabilistic Methods in CS, Micro-Organizational Theory, Operations Research, Ethnography/Context Research for Innovation, Forces in AI Governance; [Graduate] Practical Data Analysis; [Graduate] Machine Learning

PUBLICATIONS

Poster Presentations

Lee, J., & Chowdhary, A. (2025, December). *AssertBench: A benchmark for LLM resistance to user-induced factual bias*. **NeurIPS 2025 Workshop** on Evaluating the Evolving LLM Lifecycle: Benchmarks, Emergent Abilities, and Scaling, San Diego, CA.

Lee, J., Zheng, A., Gu, T. E., Kazanis, W., & Tamang, S. (2025, November). *Clustering County-level Overdose Mortality Trajectories in Ohio: A Latent Profile Analysis Approach*. **AMIA 2025** Annual Symposium, Atlanta, GA.

Zheng, A., Gu, T. E., **Lee, J.**, Capaldi, E., Mittal, A., Kazanis, W., & Tamang, S. (2024, November). *Disentangling social determinants of overdose rates in Ohio using temporal clustering methods*. **AMIA 2024** Symposium Workshop: Data Science for Injury Prevention, San Francisco, CA.

Reports

Buka, B., Khan, S., Riebe, N., **Lee, J.**, Maher, C., Wang, D., *et al.* (2025, February). *Rhode Island voters and electoral reform: Insights from Every Vote Counts' 2024 Exit Poll*. Report requested by the Rhode Island Secretary of State's Office.

Lee, J., & Edelman, Y. (2025, September). *AI developers accurately report GPQA Diamond scores for recent models*. Published online at epoch.ai. Retrieved from <https://epoch.ai/data-insights/self-reported-gpqa> [Online resource].

Lee, J., & Emberson, L. (2025, November). *Epoch's Capabilities Index stitches together benchmarks across a wide range of difficulties*. Published online at epoch.ai. Retrieved from <https://epoch.ai/data-insights/interpreting-eci> [Online resource].

Lee, J. (2025, November). *Microsoft's Fairwater datacenter will use more power than Los Angeles*. Published online at epoch.ai. Retrieved from <https://epoch.ai/data-insights/fairwater-power-usage> [Online resource].

Lee, J. (2025, December). *Today's largest data center can do more than 20 GPT-4-scale training runs each month*. Published online at epoch.ai. Retrieved from <https://epoch.ai/data-insights/gpt-4s-trainable> [Online resource].

Edelman, Y., & **Lee, J.** (2025, December). *AI capabilities progress has sped up*. Published online at epoch.ai. Retrieved from <https://epoch.ai/data-insights/ai-capabilities-progress-has-sped-up> [Online resource].

Open Contributions

Phan, L., Gatti, A., Han, Z., Li, N., Hu, J., Zhang, H., Zhang, C. B., Shaaban, M., Ling, J., Shi, S., Choi, M., Agrawal, A., Chopra, A., Khoja, A., Kim, R., Ren, R., Hausenloy, J., Zhang, O., Mazeika, M., Dodonov, D., Nguyen, T., **Lee, J.**, *et al.* (2025). *Humanity's Last Exam*. Preprint. [Over 1000 total authors].

Hartmann, H., Azevedo, F., Röseler, L., Wallrich, L., Aldoh, A., Elsherif, M. M., Liu, M., *et al.* (2025). *Tracking and mainstreaming replications in the social, cognitive, and behavioral sciences*. [Under Review]. [108 total authors, including **Lee, J.**]