DANI ROYTBURG

droytbur@andrew.cmu.edu | djroytburg.github.io | LinkedIn: in/danijroytburg

Education

Carnegie Mellon University, School of Computer Science **Pittsburgh, PA**
M.S., Machine Learning expected December 2026
Relevant Coursework: Deep Learning Systems, Advanced Machine Learning, Deep Reinforcement Learning, Probability and Statistics

Emory University, College of Arts and Science **Atlanta, GA**
B.S. summa cum laude: Computer Science, Quantitative and Social Sciences. August 2021-May 2025
Cumulative GPA: 3.85/4.0 | **Dean's List** Spring 2024 | Highest Honor's Computer Science (adv. Joyce Ho)
Relevant Coursework: Deep Learning, Machine Learning, Privacy-Enhanced Machine Learning, Systems Programming, Computer Architectures, Linear Algebra, Adv. Calculus, Probability and Statistics (I, II), Discrete Math, Algorithmic Analysis.

Work Experience

DI Group, Carnegie Mellon University **Pittsburgh, PA**
Research Assistant, advised by Daphne Ippolito September 2025 – present
- Develop multi-agent adversarial reinforcement learning environments to improve legibility of reasoning traces.

Martian Research **Remote**
Research Fellow, advised by Jacob Haimes, Phillip Quirke, Narmeen Oozeer May 2025 – present
- Engineer activation profiles of language model evaluators to suppress self-preferential bias, applied to LLM routers.
- Winner of Mechanistic Router Interpretability Hackathon (40 submissions).

Digital Humanities Lab, Emory University **Atlanta, GA**
Research Assistant, advised by Lauren Klein and Sandeep Soni November 2021 – May 2025
- Designed and trained terabyte-scale graph-text models of linguistic network propagation (ICWSM '25).
- Evaluated adversarial robustness of large language models' commonsense reasoning to distributional shift.
- Built historical correspondence networks and analytic dashboards in (D3, Next.js), supported by PBS.

ZS Associates **Evanston, IL**
Intern – Business Technology Solutions Associate June 2024 – August 2024
- Automated ETL pipelines from diverse data sources for consumption forecasts of specialty neurological treatments.
- Implemented lossless restoration of compromised client data using synthetic data imputation techniques.

Cloverpop **Remote**
Intern – Machine Learning January 2024 – May 2024
- Deployed a decision management database with structure prediction on 65k meeting transcripts.
- Forecasted demand growth with traditional machine learning methods (decision trees, regression models).

McCormick and Company **Hunt Valley, MD**
Intern – Global Marketing and Sales Data Science June 2023 – August 2023
- Built self-sustaining Amazon sales data visualization, oversight, and reporting tools in Microsoft Power BI.
- Automated customer review database analytics with GCP BigQuery, augmented with BERT-based embedding search.

Select Publications

Breaking the Mirror: Breaking the Mirror: Activation-Based Mitigation of Self-Preference in LLM Evaluators (2025) Dani Roytburg, Narmeen Oozeer et. al, *Mechanistic Interpretability, Eval-LLM Workshops @* ***NeurIPS*** (preprint)

Words and Action: Modeling Linguistic Leadership in #BlackLivesMatter (2025) **(**paper**)** **(**slides**)**
Dani Roytburg, Debbie Olorunisola, Sandeep Soni, Lauren Klein, *International AAAI Conference on Web and Social Media*

Technical Skills

Languages: Python, Java, R, JavaScript, Typescript, SQL
Human Languages: French (C1 proficiency), Russian (C1), Chinese (B1)
Packages: PyTorch, JAX, HuggingFace Transformers, scikit-learn, multiprocessing, spaCy, networkx, D3.js, React.js
Databases and Deployment: Firebase, Vercel, Amazon EMR, Google Cloud Platform (GCP), Docker, MySQL
Dev Tools: Git, Linux, Power BI, Tableau, Figma