Jake Ward

github.com/jnward | jakenicholasward@gmail.com

Research Experience

MATS   June 2025–Present
Scholar [Adam Shai & Paul Riechers]
- Novel methods for interpreting reasoning models using minimal parameter adapters.
- Eliciting reasoning behavior from base models.
- Currently studying how LLMs represent distributions over their beliefs about the world.

MATS   April 2025–May 2025
Training Phase Scholar [Neel Nanda]
- Identified representations in base models that are repurposed by reasoning models.

AISC   February 2025–April 2025
Mentee [Kola Ayonrinde]
- Explored whether SAEs represent entity-attribute binding vectors.
- Examined dynamics of SAE feature splitting using semi-synthetic data.

University of Rhode Island   March 2019–July 2019
Research Assistant [Marco Alvarez]
- Developed and optimized ML pipelines for subject age estimation.

Publications

Jake Ward, Paul Riechers, Adam Shai. "Rank-1 LoRAs Encode Interpretable Reasoning Signals." *NeurIPS Workshop on Mechanistic Interpretability*, 2025.

Jake Ward. "Reasoning Elicitation is Scale-Dependent." *NeurIPS Workshop on Efficient Reasoning*, 2025.

Jake Ward, Chuqiao Lin, Constantin Venhoff, Neel Nanda. "Reasoning-Finetuning Repurposes Latent Representations in Base Models." *ICML Workshop on Actionable Interpretability*, 2025.

Industry Experience

Melange   February 2021–September 2023
Co-founder & Lead Engineer
- We raised a $10M seed round to build an auction-based prediction markets platform.
- Later pivoted to building LLM agents for patent prosecution.
- Led engineering team of ∼4 engineers.

Spark Insights   July 2019–May 2021
ML Engineer
- CV models to identify and predict damage to properties after natural disasters.

Education

University of Rhode Island   2015–2019
B.S. Computer Science