Julian Baldwin

julianbaldwin@u.northwestern.edu | 206-280-9727 | linkedin.com/in/julian-baldwin/

EDUCATION

Northwestern University Evanston, IL
Combined BS/MS in Computer Science, GPA: 4.0/4.0 *anticipated June 2024*

- Relevant Coursework:
  - **Machine Learning:** Applied Linear Algebra, Machine Learning, Deep Learning, AI Perspectives
  - **Software Engineering:** Web Development, Agile Software Development, Game Design and Development, Game Design Studio, Human Computer Interaction
  - **Other Computer Science:** Data Structures & Algorithms, Design & Analysis of Algorithms, Programming Languages, Operating Systems, Distributed Systems

WORK EXPERIENCE

Northwestern Security & AI Lab, *Research Assistant* May 2022 - Present

- Worked under Prof. V.S. Subrahmanian as part of a multi-university team to create game-playing AI capable of complex strategic planning, negotiation, and deception, focusing on the simulated gaming environment of Diplomacy
- Refactored and debugged existing Python code base to run in parallel across nodes on TACC supercomputer cluster
- Built and deployed React UI to facilitate collection of data labels from expert-level Diplomacy players
- Containerized project code using both Docker and Singularity to run across different university environments

UChicago Existential Risks Lab, *Research Fellow* Jun 2023 - Aug 2023

- Designed and executed independent 10-week research project under mentorship of Prof. Victor Veitch and David Reber, expanding on earlier work by Li et al and Neel Nanda that used the toy problem of predicting legal moves in the board game Othello to study world representations in transformer models
- Developed a novel "global" intervention—replacing the residual stream with activations generated from linear probes—that enables precise manipulation of model behavior, successfully steering outputs to reflect arbitrarily edited board states.
- Demonstrated transformers can learn disentangled, linear representations of distinct features like board state and next turn

PROJECTS

Image Colorization GAN

- Implemented a conditional generative adversarial network (GAN) with U-net architecture in PyTorch to colorize black and white images. Trained on a dataset of 25k images from the MIRFLICKR dataset. Used a patch discriminator (inspired by pix2pix image translation) to improve performance

Redex Model of WebAssembly

- Developed a model of WebAssembly in PLT Redex, specifying syntax and operational semantics to cover core aspects like memory, control flow, and the call stack. The final model is tested extensively, with over 50 WebAssembly programs that reduce according to formal specifications

Advent of Code

- Competed in yearly series of unique coding challenges, efficiently implementing many search and dynamic programming algorithms in Python and Java
- Placed top 100 multiple times out of roughly 100,000 yearly participants

SKILLS

Programming Languages: Python (NumPy, PyTorch, Pandas), Java, Go, Javascript (d3.js, React), C# (Unity Engine)
Tools: Git, Docker, Singularity, Slurm

AWARDS

- Grand Prize and Community Favorite, Wildhacks Hackathon 2022
- Best Overall, Badhacks Hackathon 2023

LEADERSHIP

Northwestern Effective Altruism, *Co-President* Feb 2022 - Present

- Plan and oversee all events and activities for the club - includes sending weekly emails that reach hundreds of students, developing a strategic vision, running board meetings and delegating tasks, leading an 8-week fellowship seminar program with >20 students, and building a strong social community