Jai Dhyani

jai.one | github.com/jaidhyani | jai@jai.one | it's pronounced "J"

MODEL EVALUATION AND THREAT RESEARCH (METR) | Visiting Fellowship
2024 via ML Alignment and Theory Scholars (MATS) 6.0

- Co-Author on "RE-Bench: Evaluating frontier AI R&D capabilities of language model agents against human experts"
- Developing evaluations and tooling to detect AI Research Capabilities as part of Wijk Hjalmar's team

INDEPENDENT PROJECTS
2024: Delphi - Evals Technical Lead

- Led team of software engineers and researchers to develop infrastructure for training and evaluating small, reproducible language models
- Rewrote model training pipeline for significantly improved reliability, speed, and configurability

2023 - Present: Rational Animations Scriptwriter

- Writing, editing, and consulting on scripts about AI Alignment, Effective Altruism, and related topics.
- Over two million views on videos based on my scripts as of January 2024

TEMPLE CAPITAL | Quantitative Investing Strategies
2022: ML Engineering Consultant

- Optimized custom in-house investment analysis engine
- Expanded support for synthetic assets
- Built CI/CD Pipeline from Scratch
- Established coding style and formatting guidelines

BLUE ROSE RESEARCH | Public Opinion Modeling & Research
2021: Senior Machine Learning Engineer

- Optimized TF-Probability Variational Inference pipelines inferring public opinion distributions given noisy data
- Optimized database queries to speed up time-consuming operations

META (FACEBOOK) | ML Engineering for Hate Speech
2017 - 2020: Software Engineer - Machine Learning & Infrastructure

- Developed Facebook's first proactive hate speech classifiers in English, French, Arabic, and Portuguese
- Optimized models to run billions of times per day
- Integrated computer vision and multi-lingual signals to improve performance
- Identified and addressed novel issues in existing training pipeline infra causing unintended classifier behavior
- Trained and deployed model that accidentally removed the Declaration of Independence from Facebook on July 4th, 2018. Whoops.
- Built volume-adjusted thresholding system to optimize experts' time, maximizing per-sample training value while prioritizing high-impact content
- Developed stratified offline datasets to enable rapid testing and iteration.
- Built fully-automated classifier deployment pipeline capable of daily updates across multiple languages with A/B tested pilot models
- Surfaced extremely surprising issues impacting user well-being

LENDINGROBOT | Peer Lending Investments
2016 - 2017: Senior Backend Engineer; Financial Data Analyst

- Delivered machine-learning investment strategy for world's first peer-lending automated portfolio platform
- Automated data collection from multiple sources
- Weekly reports to CEO

AMAZON | AWS Networking
2014 - 2015: Systems Engineer II

- Automated testing pipeline for the custom networking software AWS runs on physical routers, reducing testing cycle from weeks to hours
- Built custom tools to automate management of highly specialized hardware.

DEMOCRATIC NATIONAL COMMITTEE | Systems Engineering
2011 - 2013 : Linux Systems Engineer | 2013 : Lead Systems Engineer

- Built/owned 1000's of metrics to maximize uptime of critical infra 24/7
- Revamped backup system to achieve better fidelity, 2x cheaper and 3x faster
- Saved over $100k by analyzing TCO on physical vs cloud DB solutions
- Built an automated webpage capture system used for research and debate prep