Joey Hejna

(Donald Joseph Hejna III)

jhejna@stanford.edu
650-404-7775
joeyhejna.com
github.com/jhejna

Education

Stanford University *September 2021 - Present*
PhD in Computer Science, AI **GPA: 4.3/4.0**

- *Funding Awards:* I am graciously supported by a DoD NDSEG Fellowship, roughly 5% selection rate.
- *Research*. I am advised by Dorsa Sadigh and part of the ILIAD lab. My research focuses on deep learning for intelligent decision-making systems and robotics.

University of California, Berkeley *August 2017 – May 2021*
B.S. in Electrical Engineering and Computer Science **GPA: 4.0/4.0**

- *Academic Awards:* Highest Honors, top 3% of graduates; Regents and Chancellors Scholar, top <2% incoming
- *Research:* Advised by Pieter Abbeel and Lerrel Pinto. CRA Undergrad Research Award Honorable mention

Publications

Show, Don't Tell: Aligning Language Models with Demonstrated Feedback *ArXiv Preprint*
O Shaikh*, M Lam*, Joey Hejna*, S Yao, M Bernstein, D Yang https://arxiv.org/abs/2406.00888

- Few-shot adaptation of language models using automatic preference generation from <10 demos.

Scaling Laws for Reward Model Overoptimization in Direct Alignment Algorithms *ArXiv Preprint*
R Rafailov*, Y Chittepu*, R Park*, H Sikchi*, J Hejna, WB Knox, C Finn, S Niekum https://arxiv.org/abs/2406.02900

- Empirical study of over-optimization for new alignment methods like DPO etc.

From r to Q*: Your Language Model is Secretly a Q-Function *ArXiv Preprint*
Rafael Rafailov*, Joey Hejna*, Ryan Park, Chelsea Finn

- Understanding alignment methods for LLMs through a token-level MDP perspective

DROID: A Large Scale In-the-Wild Robot Manipulation Dataset *Published at RSS 2024*
Aleksander Khazatsky, Karl Pertsch, … Joey Hejna, et al. https://droid-dataset.github.io/

- A multi-institutional effort to collect vast quantities of robot learning data.

Octo: An Open Source Generalist Robot Policy *Published at RSS 2024*
Octo team, … Joey Hejna, et al. https://octo-models.github.io/

- An open source robot policy trained on hundreds of thousands of demonstration trajectories.

Contrastive Preference Learning: Learning from Human Feedback without RL *ArXiv Preprint*
Joey Hejna, R Rafailov, H Sikchi, C Finn, S Niekum, WB Knox, D Sadigh https://arxiv.org/abs/2310.13639

- We learn the optimal policy from regret-based preferences without RL, scaling elegantly to high dimensions.

Inverse Preference Learning: Preference-based RL Without a Reward Function *Published at NeurIPS 2023*
Joey Hejna, Dorsa Sadigh. https://arxiv.org/abs/2305.15363

- Algorithm for directly aligning Q-function with user preferences, circumventing reward learning.

Distance Weighted Supervised Learning *Published at ICML 2023*
Joey Hejna, Jensen Gao, Dorsa Sadigh. https://arxiv.org/abs/2304.13774

- Derived method to learn optimal KL-constrained policies in offline goal conditioned RL without TD learning.

Extreme Q-Learning: MaxEnt RL without Entropy *Published at ICLR 2023 (Oral)*
Div Garg*, Joey Hejna*, Mattheiu Gesit, Stefano Ermon. https://openreview.net/pdf?id=SJ0Lde3tRL

- Introduce a Q-learning framework that models the optimal soft-values without needing to sample from a policy.

Few-Shot Preference Learning for Human-in-the-Loop RL *Published at CoRL 2022*
Joey Hejna, Dorsa Sadigh. https://openreview.net/pdf?id=IKC5TfXLuW0

- Leverage pretraining strategies to improve the query-complexity of preference learning by 20X on robotic tasks.

Improving Long-Horizon Imitation Through Instruction Prediction *Published at AAAI 2023*
Donald Joseph Hejna III, Pieter Abbeel, Lerrel Pinto. https://openreview.net/pdf?id=1Z3h4rCLvo-

- We show that modeling language instructions drastically improves generalization in low data regimes.

Task-Agnostic Morphology Evolution *Published at ICLR 2021*
Donald Joseph Hejna III, Pieter Abbeel, Lerrel Pinto. https://openreview.net/pdf?id=CGQ6ENUMX6

- We introduce the first unsupervised algorithm for agent design optimization using unsupervised objectives.

Hierarchically Decoupled Imitation for Morphological Transfer *Published at ICML 2020*
Donald Joseph Hejna III, Pieter Abbeel, Lerrel Pinto. https://arxiv.org/abs/2003.01709

- Leverage imitation techniques to develop approaches for transferring robot policies across embodiments.

Improving Latent Representations via Explicit Disentanglement *Course Project – Unsupervised Learning*
Donald Joseph Hejna III*, Ashwin Vangipuram*, Kara Liu*. http://joeyhejna.com/files/disentanglement.pdf

- Introduce three methods for disentangling latent representations: cycle loss, divergence penalty, factor prediction.

Advising

Hristo Todorov, Stanford CURIS Program *Summer 2023*

- Advised undergraduate student on summer research involving distribution shift in imitation learning.

Chet Bhateja *Autumn 2023 – Present*

- Data Quality in Imitation learning.

Industry

Google Deepmind, Student Researcher *June 2024 – Sept 2024*

- Student researcher on the DeepMind robotics team based in Mountain View

Citadel Global Quantitative Strategies, Intern *June 2019 – August 2019*

- Developed C++ proxy and API to improve job monitoring, KDB testing scripts for multi-server trading systems.
- Created APIs for trade messages, unified with query systems under a central platform for easy use by traders.
- Explored techniques for reducing RAM usage of decision tree training libraries. Achieved 75% load reduction.

Intel AI Products Group, Intern *May 2018 –August 2018*

- Produced demo-products for Intel OpenVino Model Optimizer. Computer vision project featured on intel's blog.
- Developed workflows for AWS model training, explored gradient based explanations for CV and NLP models.

Switchboard, Contracted Android Developer *Jan 2018 – August 2018*

- Programmed a multi-user voice-communication android app for Berkeley Skydeck Startup via TokBox API.
- Routed user events using SocketIO, guaranteed delivery with ack messages. Custom API for feed, notifications.

Fellowships and Awards

DoD NDSEG Fellowship 2021, roughly a 5% selection rate.

Eta Kappa Nu (EECS Honors Society). Top students in EECS.

Highest Honors, UC Berkeley Engineering 2021, top 3% of graduating class.

CRA Undergraduate Research Honorable Mention. Awarded to top undergraduate CS researchers in the US.

Regents and Chancellors Scholar. Awarded to <2% of top entering undergraduate students at UC Berkeley

EECS Honors Program. Program for high achieving students in academics and research.

Dean's List. Awarded for maintaining academic position in top <10% of engineering students at UC Berkeley.

Rambus Innovator of the Future 2017. Scholarship awarded for exceptional academics and research.

Kraft Award for Freshmen. Awarded to ~4% of freshmen UC Berkeley students for academic standing.

Eta Kappa Nu (EECS Honors Society). Top students in EECS.

NeurIPS 2023 Distinguished Reviewer

Teaching and Open Source

CS 221: Artificial Intelligence, Head Course Assistant *Autumn 2023*

- Head course assistant for Stanford CS 221. Lead development of new course assignments, exams, etc.

Research Lightning *https://github.com/jhejna/research-lightning*

- A lightweight open-source framework used for quickly implementing deep learning algorithms in Pytorch
- Reproduce SOTA implementations of SAC, TD3, PPO etc.

OpenX *https://github.com/jhejna/openx*

- An open source framework for training large behavior models using the OpenX Embodiment datasets

CS 189: Machine Learning, Teaching Assistant *Spring 2020, Spring 2021*

- Taught sections on topics from classic ML to modern deep learning. Author of neural nets homework, still used.
- Earned overall rating of 4.61/5.00 from students in comparison to department average of 4.41

EECS 127: Optimization Models, Teaching Assistant *Fall 2020*

- Taught sections on linear alg, duality, convex models. Managed website and internal course logistics.

CS 70: Discrete Math and Probability Theory, Teaching Assistant *Fall 2019*

- Taught two weekly discussions, held office hours. Earned overall 4.68/5.00 rating in comparison to 4.33 average.

Hack:Now – CalHacks, Volunteer Workshop Instructor *April 2020*

- Gave an introductory machine learning tutorial for an online version of Cal Hacks, the largest collegiate hackathon. Prepared all materials and presented. https://github.com/jhejna/mlworkshop