Isadora White

isadora.c.white@gmail.com icwhite.github.io/website

Education

University of California, Berkeley Expected graduation date: May 2024
Honors B.A. in Computer Science and Applied Math ***GPA: 3.95/4.0***

Graduate-Level Coursework: Natural Language Processing, Deep Reinforcement Learning, Deep Learning (cross-listed graduate class), Machine Learning (cross-listed graduate class)
Relevant Coursework: Probability Theory and Random Processes, Data Structures, Machine Structures, Signals and Systems, Abstract Algebra, Complex Analysis, Honors Real Analysis, Abstract Linear Algebra, Discrete Math & Probability Theory, Linguistics, Computational Creativity

Preprints

LMRL Gym: Benchmarks for Multi-Turn Reinforcement Learning with Language Models, *In Submission* Nov. 2023
Marwa Abdulhai, **Isadora White**, Charlie Victor Snell, Charles Sun, Joey Hong, Yuexiang Zhai, Kelvin Xu, Sergey Levine

- Proposed 8 benchmark tasks to test capabilities of training RL algorithms for language
- Evaluated 7 baseline experiments including supervised fine-tuning, offline and online RL and few-shot prompting
- **paper:** https://arxiv.org/abs/2311.18232 **website:** https://lmrl-gym.github.io/

Research Experience

Research Assistant, Berkeley AI Research Feb. 2023 - Present
Advised by Prof. Sergey Levine

- Designed text-based chess and maze benchmark tasks to test complex strategic reasoning in language models and RL algorithms
- Evaluated reinforcement learning with language baselines and tuned hyperparameters for efficiency and stability
- Utilizing implicit feedback from dialogue for improved collaboration with humans through conversation

Research Assistant, DAAD RISE Germany, University of Muenster May 2022 - Sep. 2022
Advised by Prof. Joerg Becker and Kilian Mueller

- Compared efficacy of fine-tuned single language and multi-language BERT and T5 language models for comment moderation
- Experimented with using SHAP values and attention rollouts to explain predictions of the T5 classifier
- Identified and mitigated biases in the predictions of the automated content moderation system

Undergraduate Research Apprentice, Cardiac Vision Lab Feb. 2021 - May 2021
Advised by Prof. Jan Christoph

- Mitigated motion artifacts in videos of hearts using PWC-Net architecture, trained PWC-Net in TensorFlow
- Visualized predicted motion of the trained model, modified architecture for improved precision of predictions
- Generated data for data augmentation and better generalization using 2D simulations of the heart

Talks

Moderat!: Language Models for Fair and Explainable German Comment Moderation Aug. 2022
Isadora White, Kilian Mueller

- Improved accuracy of hate-speech detection by fine-tuning single-language and multi-lingual language models
- Explained predictions using SHAP values and mitigated biases in the predictions
- **presentation:** icwhite.github.io/website/papers/moderat_presentation.pdf
- **code:** https://github.com/icwhite/moderat_transformers_rise

BTA: Business Term Glossary Association with Active Learning Aug. 2021
Bojan Furlan, **Isadora White**, AnHai Doan

Developed an active learning random forest classifier to match business glossary terms using their corresponding acronym. Internal Informatic Report and Presentation.

Research Projects

Movement Classification and Artifact Correction for Reliable Brain-Computer Interfaces May 2023
Deep Learning Course Project

- Explored using Transformers and CNNs for classifying ECoG signals as robotic motions
- Designed Transformer-based autoencoder and data imputation techniques to correct artifacts in ECoG signals
- **report:** icwhite.github.io/website/papers/ecog_bci.pdf

Reward and Exploration Strategies for Wordle with Deep Reinforcement Learning Dec. 2022
Deep Reinforcement Learning Course Project

- Trained Deep Reinforcement Learning agents on Wordle with custom reward functions and exploration algorithms
- Achieved 98% win ratio and an average of 4.3 moves in an elimination win condition
- **report:** icwhite.github.io/website/papers/deep_woRdLe.pdf **code:** https://github.com/icwhite/deep_woRdLe_agent

Work Experience

Software Development Intern, AWS AI: Lookout for Metrics Aug. 2021 - Dec. 2021

- Detected outliers in time series data using the unsupervised learning technique Random Cut Forest (RCF)
- Visualized the detected outliers, analyzed false positives, experimented with new applications
- Developed AWS infrastructure to build a pipeline for outlier detection on new use cases

Machine Learning Intern, Informatica Jun. 2021 - Aug. 2021
With Prof. AnHai Doan

- Implemented querying strategy for selection of unlabeled examples to label for active learning
- Developed stopping strategy and empirically determined stopping point to efficiently utilize crowd-worker labeled examples
- Improved F1 score by over 5% on all datasets for business term glossary association

Leadership & Volunteering

Association of Women in EE&CS at UC Berkeley
Mentor Aug. 2022 - Present

- Guided underrepresented students in EE&CS on choosing courses, managing stress, and getting involved in ML/AI research

DE&I Committee Member and Officer Feb. 2023 - Present

- Organized events to support underrepresented minorities in EE&CS such as socials, study hours, and fundraisers
- Advocated for initiatives that make campus more inclusive such as menstrual products in restrooms, classroom accessibility, and working towards a more inclusive climate within the EE&CS department

Computer Science Mentors Aug. 2020 - Dec. 2020
Junior Mentor

- Built confidence in problem-solving skills for six students in Computer Science
- Explained and communicated concepts such as recursion, linked lists, trees and SQL

Awards and Honors

DAAD RISE Germany Scholarship 2022

- Awarded a scholarship to pursue research in Germany for 3 months by DAAD (German Academic Exchange Institute)

Dean's List Fall 2020, Fall 2022, Spring 2023

- Top 10% of all undergraduates in the College of Letters & Sciences at UC Berkeley by GPA over the semester.

EECS Honors Program at Berkeley Aug. 2022 - Present

- Designed for very talented undergraduate students interested in undergraduate research
- Honors students pursue an academic concentration outside of the department, engage in research, receive a special faculty advisor, and are invited to special events with faculty and EECS Honors alumni.