Isadora White

isadora.c.white@gmail.com | isadoracw@berkeley.edu | +1 (612) 845-9120
github.com/icwhite | linkedin.com/in/isadora-c-white

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 Courses: Natural Language Processing, Deep Reinforcement Learning, Deep Learning, Machine Learning
Computer Science Courses: Probability Theory, Algorithms and Data Structures, Computer Architecture, Signal Processing, Structure of Computer Programs
Math Courses: Abstract Algebra, Complex Analysis, Real Analysis, Linear Algebra, Discrete Math, Numerical Analysis
Other Relevant Courses: Introduction to Linguistics, Computational Creativity

Research

LMRL Gym: Benchmarks for Multi-Turn Reinforcement Learning with Language Models Nov. 2023

Marwa Abdulhai, **Isadora White**, Charlie Victor Snell, Charles Sun, Joey Hong, Yuexiang Zhai, Kelvin Xu, Sergey Levine

We propose 8 benchmark tasks to test RL with language models and a suite of 7 baselines to further research into using language models with RL. paper, website, code

Movement Classification and Artifact Correction for Reliable Brain-Computer Interfaces May 2023

Annie Taylor*, **Isadora White***, Zheyuan Hu, Shiangyi Lin

We explore new deep learning architectures involving transformers and CNNs for classifying ECoG signals as robotic actions. We propose data imputation and a transformer encoder/decoder architecture to correct artifacts. paper, code

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

Isadora White*, Bennett Cohen*, Sean Tsung

We explore an application of Deep RL to Wordle. We propose and compare three different reward functions and several different exploration algorithms. We achieve 98% win ratio and an average of 4.3 moves in an elimination win condition. paper, code

Moderat!: Language Models for Fair and Explainable German Comment Moderation Aug. 2022

Isadora White, Kilian Mueller, Joerg Becker

I independently worked on a project on detecting hate speech using language models. I increased accuracy on RP-Mod hate speech detection benchmark, explained predictions using mechanistic interpretability such as SHAP values, and mitigated biases against marginalized groups in the predictions. presentation, abstract, code

BTA: Business Term Glossary Association with Active Learning Aug. 2021

Bojan Furlan, **Isadora White**, AnHai Doan

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

Work Experience

Research Assistant, Berkeley AI Research **Feb. 2023 - present**

Explored applications of multi-turn reinforcement learning to text-games and interactive dialogue with Sergey Levine, Marwa Abdulhai, and Charlie Snell.

Research Assistant, DAAD RISE Germany, University of Muenster **May 2022 - Sep. 2022**

Investigated fine-tuning language models for hate speech detection, explainability techniques, and biases of the model and datasets.

Undergraduate Research Apprentice, Cardiac Vision Lab **Feb. 2021 - May 2021**

Detected motion artifacts in videos of hearts using PWC-Net architecture for motion detection.
Visualized training curves, generated data for data augmentation, trained PWC-Net in tensorflow

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

Detected outliers in time series data using the unsupervised technique Random Cut Forest (RCF)
Visualized the detected outliers, analyzed false positives, experimented with new applications
Deployed AWS infrastructure to assist users in loading data

Machine Learning Intern, Informatica **Jun. 2021 - Aug. 2021**

Implemented querying strategy for selection of examples to label in active learning.
Improved F1 score by over 5% on all datasets for business term glossary association.

Leadership and Volunteering

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

Provided guidance to 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 and inspire underrepresented minorities in EE&CS.
Met with the Chair of EE&CS to advocate for DE&I initiatives such as SVSH prevention

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

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