Ashvin Govil

ashvion@gmail.com | 512.964.7834 | Los Angeles, CA | ashvio.github.io

UNIVERSITY OF TEXAS AT AUSTIN

Fall 2015- May 2019 | BS Computer Science (Turing Scholars Honors); BA Government | GPA 3.62

REDDIT ELECTION PREDICTOR | Turing Honors Undergraduate Thesis *2016-2019*

- Successfully orally defended Honors thesis to a cross-disciplinary panel of professors. Advisor: Dr Bruce Porter.
- The research uses Reddit data collected into a Solr database to accurately predict public polling and election outcomes in the 2016 Presidential Election between Donald Trump and Hillary Clinton.
- Began as an independent exploration of how the ideas of data science and NLP could be applied on social media, and grew into a sponsored research project.
- Received funding independently from a prominent AI company.

ARABIC RADIO SENTIMENT ANALYSIS | Sponsored Undergraduate Research *Fall 2018 - Spring 2019*

- Collaborated with Dr Bruce Porter of UT CS and a prominent AI company to lead a small team of undergraduate students on a specially funded research project.
- Researched applying a sentiment model on Arabic song lyrics to create a dataset that could be used to examine how major world events influence the songs that are played on Arabic radio stations.
- Summarized and presented findings to the sponsoring company and professor.

DIRECTED READING PROGRAM (DIRP) | Mentor *Spring 2020, Fall 2022*

- As an alum, volunteered to mentor undergraduate CS students on NLP to help get them involved in research.

CS 378 PRACTICAL APPLICATIONS OF NLP BY DR BRUCE PORTER | Teacher's Assistant *Spring 2019*

- Helped Dr Porter develop and lead a new class for teaching students modern NLP for industrial applications.

UT ONRAMPS | Computer Science Course Grader *Fall 2017 - Spring 2018 | 10-20 hrs/week*

- Course grader for UT's high school dual enrollment program designed to provide college-level computer science courses to high schools with high numbers of underrepresented minorities.

THE DAILY TEXAN | Columnist *Fall 2015; Spring 2019 | One article every two weeks | Unpaid*

- Wrote opinion articles for the campus newspaper on a variety of important higher education issues such as diversity programs in the UT Computer Science Department and UT's landmark affirmative action case in the Supreme Court.

INDEPENDENT RESEARCH

CELESTE REINFORCEMENT LEARNING AI AGENT *Fall 2023*

- Developing the world's first AI to play the renowned platformer video game Celeste using PPO Reinforcement Learning with Ray RLLib and a PyTorch policy model. Code available on Github.
- Capable of completing the first 1.5 chapters of the game. The model behind the policy is a convolutional network with attention layers, and is trained on random levels with a complex vectorized reward function.

INDUSTRY EXPERIENCE

SINGULARITY 6 | Senior Machine Learning Engineer *Jun 2022 - Sep 2023 | Los Angeles*

- Using publically available research and datasets on hate speech, developed a multilingual 190M parameter LLM using Huggingface/PyTorch capable of classifying text chats. Achieved over 95% on relevant metrics of accuracy, recall, and precision for English, 80% for other languages. The model enables real-time evaluation and filtering of hateful, harassing, or offensive chats for Palia, the company's flagship MMO. The deployed model had only a 0.001% false positive rate per message, while capable of preventing up to 30% of player reports for chat abuse.

RIOT GAMES *Jul 2019 - Jun 2022 | Los Angeles*

Data Scientist *Feb 2021 - Jun 2022*

- Researched and developed a custom LLM trained on hundreds of millions of usernames using Huggingface/PyTorch. This base model was further fine-tuned into a model for detecting racist or abusive usernames. Achieved over 96% precision and 90% recall across all languages, compared to 50% precision of the previous model. Currently evaluating millions of names each year for abuse.
- Led effort to minimize racial bias in model outputs by diversifying the dataset to include all human names.
- Developed a zero-shot classification pipeline to quantify the sentiment and topics of free-response surveys. Used to analyze trends in customer sentiment over time for various business KPIs.

Software Engineer *May - Aug 2018 (Internship), Jul 2019 - Feb 2021*

- Created and maintained a variety of game and business data pipelines on Airflow and PySpark across petabytes of game data for analysts and data scientists for every Riot game.

UBER | Software Engineer Intern *May - Aug 2017 | Palo Alto, California*

- Expanded backend services in Java, Python, and Go for Uber's query monitoring pipelines. Pipelines generated performance statistics and resource usage for query engines such as Hive and Spark.

SPARKCOGNITION | Software Engineer/Data Science Intern *Mar 2016 – Jan 2017 | Austin, Texas*

- Worked closely with data scientists to develop data pipelines that processed terabytes of wind turbine sensor data into valuable predictions of turbine failure using machine learning.

VOLUNTEERING

FOOD NOT BOMBS *2020-2022 | Los Angeles, California*

- Prepared and served healthy plant-based food for unhoused individuals in Skid Row.
- Organized and carried out a special Food Not Bombs from my apartment to feed every unhoused individual living in Westwood Park, while also supplying them with essential sanitary supplies. Worked with the individuals to understand their unheard political needs, and provided them an opportunity to send public comments on a LA City Council resolution that was meant to force them out of the park.

RESEARCH INTERESTS

- How can the "pretrain/fine-tune" paradigm of LLMs be applied to RL tasks, such as a platformer model that can be fine-tuned on specific games or a climate model that can be fine-tuned on specific weather predictions?
- Can multilingual LLMs be trained to decode unknown languages, such as the Voynich Manuscript, by correlating the semantic and syntactic structures of the target language to those of other languages understood by the encoder?
- Can LLMs be taught how to create discoveries not within their dataset? For example, a LLM modulated with RL capable of discovering calculus using only first principles of mathematics known by Isaac Newton.

SKILLS

PROGRAMMING LANGUAGES

Proficient: Python •SQL • Java • C++ • C#
Experienced: C • R • Rust • Go • Assembly • Bash • Javascript

DATA SCIENCE FRAMEWORKS

GPT4 • LLMs (eg, LLaMA, RoBERTa) • Ray.io: RLLib, Ray Tune • HuggingFace (NLP) • PyTorch • Tensorflow/Keras • NumPy/Pandas • Scikit-Learn

DATA PLATFORMS

Airflow • PySpark • Databricks • Kafka • Hive • SQL/NoSQL • Presto • Hadoop • Solr/Lucene

SOFT SKILLS

Cross-Team Collaboration and Leadership • Technical Writing • Policy Writing • Consensus Decision-making • Community Management