Sebastian Algharaballi-Yanow

Machine Learning Engineer & Data Scientist

Irvine, CA, 92614 | salghara@uci.edu | GitHub | Linkedin | Website

Technical Skills

Programming Languages/Libraries:

- Python, R, SQL. Extensive experience with Pandas, Numpy, Seaborn, Sci-kit Learn, MatPlotLib, PyTorch, Tensorflow, NLTK, Spacy, OpenCV, SciPy, Transformers, and the Tidyverse.

Machine Learning:

- Supervised/Unsupervised Learning, Deep Learning, Predictive Modeling, Emotional Intelligence, Cloud Services (AWS and Azure), Computer Vision, Human-Centric AI, Large Language Models.

Data Science:

- Data Wrangling, Exploratory Data Analysis (EDA), Data Visualization, Statistical Analysis, Databases.

Tools/Technologies:

- Git/Github, Jupyter Notebook, R Studio, Anaconda, VSCode, mySQL, Power BI, Tableau.

Education:

UC Irvine – *Master of Data Science*
September 2023 - December 2024

UC San Diego – *Bachelor of Science in Machine Learning & Neural Computation*
September 2020 - June 2023

University of California, San Diego Extension – *Specialized Certificate, Machine Learning*
June 2022 - June 2023

Professional Experience:

Scale AI - *Generative Artificial Intelligence Prompt Engineer*
May 2024 - Present

- Refined the responses from a pre-release version of ChatGPT O-1 to enhance emotional intelligence and human-like interactions in real-time conversations, significantly improving beta-specific satisfaction scores from **3.6/5 to 4.67/5.**
- Developed regression models to predict user emotional satisfaction, reducing AI responses that were deemed as hallucinations by an average of **30%** across various models (Claude 3.5, LLaMA 3, and Gemini Ultra).
- Implemented reinforcement learning techniques using human feedback on nearly 500,000 interactions, increasing the percentage of AI responses rated as "relevant and empathetic" by human evaluators from **61% to 83%.**
- Analyzed extensive user-AI interaction datasets (over 3 million data points) to identify patterns of positive engagement, leading to an increase in average user session duration from **8 minutes to 13.5 minutes** across testing platforms.

MoodMe - *Machine Learning Engineer & Co-Founder*
December 2023 - September 2024

- Expanded our emotion detection training dataset from 3 to 7 major ethnic groups, which enabled the development of more inclusive machine learning models. This improvement increased overall accuracy across all demographics from **68% to 87%** and **reduced** misclassification rates in underrepresented groups by **62%**.
- Enhanced our emotion detection model through transfer learning, boosting, and noise injection, improving its accuracy in identifying emotions from **75% to 91.5%** across 8 emotion categories.
- Created EmotionTrack, a well-being diary powered by a customized BERT text emotion detection model. Achieved an F1 score of **89%** in detecting emotions across 8 categories, resulting in a **41% increase** in user-reported emotional self-awareness after 2 months of consistent use in beta testing.

Donald Bren School of ICS - *Master of Data Science Ambassador*
October 2023 - December 2024

- Conducted time series analysis using R's "forecast" library and Python's "statsmodels" to track and predict trends in data science master's programs across the U.S. Confirmed UCI's MDS program as a **top program in the country** and identified additional growth areas, informing several curriculum and operational updates that further increased program relevance in national rankings such as **Forbes.**

Sportradar US - *Sports Data Operator*
September 2022 - Present

- Attended and recorded game-related statistics for **over 250** NCAA and professional sporting events across 4 different sports (basketball, baseball, volleyball, and soccer).
- Improved on-site data collection efficiency by suggesting a UI modification within the basketball play-by-play logging workflow, **reducing** average input time per play from **8 seconds to 3 seconds**.
- Consistently achieved 5-star performance ratings, placing in the **top 10%** of all data operators in the United States and earning "Professional Operator" status after just 6 months, faster than the typical 12-month timeline.

Projects:

Natural Language Financial Analytics on CEO Communication: (Presentation)

- **Problem:** Explored the human element in corporate communication by investigating the relationship between CEO language choices in earnings calls and future financial performance.
- **Action:** Developed an analysis pipeline including text preprocessing, text-to-numeric conversion techniques (TF-IDF and SVD), and correlation analysis with financial metrics. Applied sentiment analysis and statistical tests to uncover nuanced patterns in communication styles.
- **Result:** Uncovered industry-specific relationships between earnings call sentiments and financial metrics for Apple and Lululemon. Found no strong evidence of short-term causality, but identified correlation patterns in sentiment and financial performance between the two companies.
- **Impact:** Emphasized the importance of developing context-aware AI for financial analysis by showing how the impact of human communication on financial metrics varies significantly between companies.

Synthetic Recipe Review Generation: (Research Paper)

- **Problem:** Addressed inconsistencies in online recipe reviews that often confuse users seeking reliable cooking advice, aiming to improve the human experience of finding and using recipes.
- **Action:** Developed and compared two AI models - a traditional Transformer Decoder and a novel Transformer to Hidden Markov Model (HMM) Decoder - to generate human-like recipe reviews.
- **Result:** Both models produced coherent reviews, with the traditional model showing more logical consistency and the HMM model demonstrating a broader vocabulary range.
- **Impact:** Opened avenues for more empathetic AI applications across various fields, showing how sentiment and numerical ratings can be aligned to provide more useful feedback systems in areas such as product reviews and customer service.

Autoencoder Anomaly Analysis for NBA Players: (Research Paper)

- **Problem:** Explored the human element behind game-to-game performance variations in NBA players, with a focus on Kobe Bryant's 2008-2009 season.
- **Action:** Developed an autoencoder machine learning model to uncover patterns in player statistics that might reflect physical and emotional factors affecting performance.
- **Result:** Achieved high accuracy in identifying above or below average games, providing insights into how factors like rest, travel, and game importance influence a player's output.
- **Impact:** Demonstrated how AI can be used to understand the human aspects of athletic performance, potentially leading to more personalized player support and game strategies.

Advanced NBA Referee Analysis: (Research Paper)

- **Problem:** Explored decision-making patterns of NBA referees to enhance fairness and integrity of the sport. Aimed to identify areas where officiating can be optimized.
- **Action:** Developed a comprehensive dataset (over 30,000 data points) and created four neural network models (three Feed Forward and one Convolutional) to analyze referee behavior patterns.
- **Result:** Achieved 92% test accuracy with the Convolutional Neural Network model, confirming the effectiveness of the approach in identifying patterns and areas of improvement in referee standards.
- **Impact:** Highlighted the potential for AI to support, rather than replace, human judgment in complex, real-time decision-making scenarios.

Predicting Civilian Complaint Review Board Outcome on Civilian Complaints: (Presentation)

- **Problem:** Addressed predicting officer remediation amid the complexity of police misconduct complaints.
- **Action:** Extracted data features using the “Eli5” Python library, directing focus on variables affecting disciplinary outcomes. Developed and evaluated F1-scores from logistic regression, K-NN, Random Forests, and Gradient Boosting models.
- **Result:** Unveiled logistic regression as the optimum predictive algorithm, enabling accurate forecasts of officer remediation outcomes.
- **Impact:** Proved how AI can support more fair and consistent decision-making in public sector oversight, potentially improving trust between institutions and communities.