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, the Tidyverse, and LLM APIs.

Machine Learning:

- Supervised/Unsupervised Learning, Deep Learning, Predictive Modeling, Emotional Intelligence, Prompt Engineering, Computer Vision, Human-Centric AI, Large Language Models (LLMs).

Data Science & Analytics:

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

Tools/Technologies:

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

Education:

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

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

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

Professional Experience:

Stealth Talent Solutions - *Machine Learning/Artificial Intelligence Engineer*
November 2024 - Present

- Engineered and deployed a production LLM chatbot that transforms basic job requirements into comprehensive job descriptions (JDs), enabling recruiters to create **200+ tailored JDs per day.**
- Built a custom transformer-based model for parsing unstructured job descriptions and resumes, identifying **150+ successful** placements that traditional keyword matching would have overlooked.

Plink.bio - *Software Engineer Intern - GenAI*
October 2024 - Present

- Architected a multi-modal content analysis pipeline combining video frame analysis, Optical Character Recognition (OCR), and speech-to-text to extract comprehensive metadata from creator content, processing videos in **under 3 seconds.**
- Developed an **end-to-end LLM system** that analyzes creator content across multiple languages, providing personalized content strategy recommendations by processing visual elements, transcripts, and audience engagement patterns
- Built and integrated computer vision models for real-time object and brand detection in videos, with initial testing across **1000+ frames** showing **90% accuracy** in identifying monetizable product placement opportunities.

Scale AI - *Generative Artificial Intelligence Prompt Engineer*
April 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, etc).
- Implemented reinforcement learning techniques using human feedback on nearly 500,000 interactions, increasing the percentage of AI responses rated as "relevant and humanistic" by human evaluators from **61% to 83%.**
- Engineered several prompt frameworks utilizing chain-of-thought and few-shot learning techniques, reducing response latency across multiple production-scale LLMs from an average of **2.3 seconds to 1.1 seconds** while improving task completion accuracy from an average of **88% to 95%**.

MoodMe - *Machine Learning Engineer & Co-Founder*
December 2023 - October 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 United States. Confirmed UCI's MDS program as a **top program in the country** and identified additional growth areas.

Sportradar US - *Sports Data Operator*
September 2022 - October 2024

- 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**.
- Achieved 5-star performance ratings, placing in the **top 10%** of data operators in the United States.

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 context-aware 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 machine learning applications across nuanced 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 machine learning and data analysis can be used to understand human performance, potentially enabling more personalized enhancement strategies.