Scott Pitcher

New York, NY, 10025 | +1 (813) 557-1210 | [Email](#) | [Linkedin](#) | [GitHub](#) | [Website](#)

SUMMARY

Driven Data Scientist with 2+ years of experience in research and development, specializing in neuro-symbolic AI and statistical analysis. Skilled in optimizing processes, seeking to expand my skill set through challenging, impactful projects.

SKILLS & LANGUAGES

Tools: PyTorch, TensorFlow, Scikit-Learn, MySQL, NoSQL, AWS, GCP, Apache Spark, Flask, Docker, Kubernetes

Programming Languages: Python, SQL, R, MATLAB, Java, Ruby, HTML, Javascript, CSS, Lua, C++

WORK EXPERIENCE

Moffitt Cancer and Research Institute, *Tampa, FL*   *September 2022 – June 2024*

Research Data Scientist

- Integrated both Causal Inference to identify and refine key variables, and Active Learning to improve model training and accuracy, resulting in mean model efficiency increase by 31%.
- Collaborated with a cross-disciplinary team to present data insights through reports and visualizations (seaborn, Tableau), informing strategic decisions for future projects.
- Developed data pipelines in SQL to clean and process 250GB of patient data, reducing API access time by 36%, and ensuring data integrity for downstream analysis and reporting.
- Collaborated with teams to develop machine learning models (scikit-learn) and operationalize patient outcome predictions, improving model accuracy to 97% and informing strategic decisions.
- Led my team in integrating advanced statistical techniques (e.g., XGBoost and Bayesian Inference) into our predictive models, improving model precision by 12%, and driving actionable insights for research decision-making.

EDUCATION

University of South Florida, *Tampa, FL*   *August 2020 – June 2024*

Bachelors of Science in Computational Mathematics*: Data Analytics and Business Intelligence Concentration*

- Cumulative GPA: 3.86/4.00 (Magna Cum Laude); USF Green & Gold Scholarship *(~$15,000/year),* Dean's List
- **Relevant Coursework:** Advanced Probability, Data Analysis, Machine Learning Development/Deployment, Statistics

PROJECTS

Pokemon Platinum AI - *[(GitHub)](#)*   *July 2024 - September 2024*

Tools: PyTorch, Computer Vision, Reinforcement Learning (with Human Feedback), MLflow

- Developed and optimized an AI gameplay model leveraging reinforcement learning with human feedback (RLHF), increasing task completion rate by 35% and improving decision-making efficiency in complex gameplay environments.
- Implemented action pruning to streamline the model's decision-making process by limiting the action space, resulting in a 25% reduction in time to reach the target in-game objectives.
- Enhanced the AI's ability to retain sequential gameplay data by integrating LSTM-based memory in PyTorch, boosting decision-making accuracy by 31%.
- Integrated YOLOv9s object detection for real-time gameplay frame analysis, achieving a 78% reduction in manual annotation time through automated labeling and model fine-tuning.
- Reduced computational costs by 22% via model architecture optimization and effective GPU utilization, maintaining high model accuracy while lowering resource usage.

Spotify User Analysis And Recommender System – *[(GitHub)](#)*   *July 2024 - August 2024*

Tools: PyTorch, Pandas, Numpy

- Implemented reinforcement and active learning techniques, elevating predictive accuracy by 20% through user-interaction adaptation, addressing diverse user behavior patterns.

- Reduced customer churn by 44% with a scalable playlist recommender in PyTorch using GCNConv layers, enhancing user retention across varied demographics.

Tampa GPT - *(GitHub)*   *May 2024 - July 2024*

- Developed a language model (LLM) specifically tailored to handle local language nuances, utilizing advanced NLP techniques to preprocess and tokenize large text datasets sourced from Tampa's local forums and news articles.
- Implemented the model in PyTorch, applying transformer-based architecture to optimize for contextual understanding and prediction accuracy.
- Fine-tuned the model with domain-specific datasets, improving text generation and semantic understanding in specialized contexts by 18%.
- Optimized model training and inference through effective GPU utilization, reducing computational costs by 20%.

Decoder LLM from Scratch - *(GitHub)*   *April 2024 - May 2024*

- Built a decoder-based large language model (LLM) from scratch using PyTorch to gain a deep understanding of LLM architecture and operations.
- Processed 10GB of OpenWebText data, utilizing natural language processing (NLP) techniques such as tokenization and data preprocessing for efficient input management.