Pranav Gujarathi

Dallas, TX 75206 | pranavdg1997@gmail.com | (331) 248-7381

LinkedIn: linkedin.com/in/pranav-gujarathi/ | **GitHub:** github.com/pranavdg1997 | website

Education

Master of Science in Data Science | Indiana University, Bloomington |GPA: 3.54 **August 2019** - **May 2021**

Bachelor of Technology in Electrical Engineering | Indian Institute of Technology, Roorkee **July 2014** - **May 2018**

Work Experience

Generative AI Engineer | Walmart Global Tech (Dallas, TX) **November 2023 – Present**

• Developing Generative AI based automation solution that carries out exploratory Data analysis on complex data types simply based on voice/text commands, removing the learning curve for non-programmers to use analytical tools, in some cases removing the need to build dashboards. Saved 130,000 associate work hours per week.

• Delivered end-to-end automation solution for extracting item attributes and competitor using a RAG system (Retrieval Augmented Generation) to generate structured usable data – in contrast to a legacy system based on manual annotation – achieved 80% in cost savings and over 24 times improvement in turnaround time.

• Took charge of ML Engineering - containerization (Docker, configuring Kubernetes and deploying API endpoints in tandem with frontend solutions - enforcing high standards for quality, reliability, and security in deployed machine learning solutions.

Data Scientist | Walmart Global Tech (Dallas, TX) **January 2022 – October 2023**

• Deployed and managed large-scale anomaly detection engine with into production with real-time user feedback, providing upwards of 70% capture rate.

• Achieved > 80% explained variance and less than 5% global error by contributing to a novel causal-inference forecast model, helping make around $1.6 Billion worth of sales, more explainable and interpretable. Awarded internal MAD (Making a difference) award for achievements in explainable AI.

Data Science Associate | ZS Associates (Los Angeles, CA) **July 2018 – December 2021**

• Deployed a product with favorable client feedback and improved performance in the form of a cross platform application.

• As part of the project, utilized Python libraries, Deep Learning frameworks and transformer models to implement a Natural Language Inference pipeline, i.e., extracting domain-relevant inferences from textual data (news articles, publications, etc.).

• Deployed a novel ML based solution for marketing strategy planning with 60% more projected efficiency on target reach and market penetration ROI, using multivariate time series models and Linear Optimization.

Research Engineer | **Indiana University (Bloomington, IN)** **January 2020 – May 2021**

• Building a Mind Lab: Designed and implemented pipelines to successfully conduct experiments as part of NSF funded project under the guidance of Professor Justin Wood. The project involved working across topics in Computer Vision and Deep Reinforcement Learning.

• IUPUI Data Lab**:** Conducted research and experiments in Natural Language Processing models and architectures towards a successful end to end process from ideation to eventual publication under the guidance of Prof Sunandan Chakraborty.

• Kelley School of Business: Successfully deployed an MLOps pipeline starting from a PoC formulation to a GUI dashboard using Big Data libraries and cloud-based parallel computation.

Technical Skills

Programming Tools– Python (Llama Index, Langchain, Open AI API, PyTorch, Tensorflow, NumPy, Pandas, GPU use, OpenCV, Scikit learn), Open-source LLMs (LLAMA2, Pythia, Falcon, Mistral etc.), Big Data (SQL, Spark), Version Control (Git, bash, SVN), Cloud computing (AWS, Databricks, GCP)

Conceptual skills – Generative AI and LLM: RAG systems, RLHF, Prompt Engineering | ML Engineering : CI/CD, Docker, Kubernetes | Machine Learning : Statistical Experience and Quantitative analysis, Data Mining, Predictive modelling**,** Computer Vision, Natural Language Processing (NLP), Reinforcement Learning, Information Retrieval, Optimization algorithms, Regression, Classification, Clustering, Time series forecasting

Projects

Chick AI: Understanding Animal behavior using Computer Vision and Reinforcement Learning **January - April 2021**

- Experimented with various computer vision architectures to simulate biological object detection process in baby chickens' brain to a digital environment.
- Achieved improved accuracy (76% to 93%) as well as computation cost (-34%) for object detection using models such as SIMCLR, Autoencoders and as well as improvement in RL tasks using state-of-art models (A2C and PPO models) as part of experiments to reverse engineer visual understanding in animal brains.

Cause-Effect Entity Recognition using Natural Language Processing & LLMs **May 2020 – August 2021**

- Achieved state-of-art performance (91% Recall) for the NLP task of predicting causal-inference based entity recognition from text, by implementing modified Transformer models in PyTorch.
- Refined and improved the model and scope for further publications.

Publications, Achievements and extra-curricular

- Published **"Controlled-rearing studies of newborn chicks and deep neural networks"** at **Shared Visual Representations in Human & Machine Intelligence workshop, NeuRIPS 2021** winning best paper award at the event.
- Published **"Using Causality to Mine Sjögren's Syndrome related Factors from Medical Literature"** at ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)
- Awarded ***Luddy Outstanding Research Award*** for research contributions during MS degree. (May 2021)