Sunit Singh

Columbus, OH, USA 43201
614-812-9608
sunitsingh.bitsg@gmail.com

OBJECTIVE

Data Scientist with 6 years of full-time work experience in building and deploying reliable machine learning systems. Proficient in data analytics, cloud-based ML systems, big data frameworks like Apache Spark and MLOps. Experienced in training / fine-tuning LLMs for generative AI tasks.

Seeking full-time opportunity from May 2024. Open to relocation.

EXPERIENCE

Amazon – Alexa Taskbot Challenge 2023, Columbus, OH
Research Associate, Apr 2023 - Sep 2023

- Led the development of Tacobot, a task-oriented dialog system for the Amazon Alexa Taskbot Challenge - 2023.
- Built comprehensive agent dialogue management strategy by leveraging PEFT techniques(LoRA) to instruction-finetune LLMs on Amazon's Wizard-of Tasks customer interaction datasets.
- Deployed end-to-end system on AWS and observed 37% increase in customer engagement rate with new dialogue strategy, and an overall 24% increase in task success rate.

Compass Real Estate, Bangalore, India
Senior Machine Learning Engineer, Jul 2021 - Jul 2022

- Built learning-to-rank based DNN scorer for 'Similar Homes', a content-based recommendation system for users prospecting real-estate properties.
- Achieved over a 20% improvement in customer clickthrough rate on recommended properties, compared to the legacy system.
- Contributed to the creation of automated ETL pipeline using Databricks for data cleaning, preprocessing using Spark SQL, transformations and loading to Databricks feature store.
- Built a scalable MLflow pipeline for reproducible model training and experimentation, with Apache Kafka for real time generation of property recommendations.
- Created document intelligence system using OCR and BM25 ranker, to create a textual semantic based document ranker for property seller-buyer contracts.

Salesken, Bangalore, India
Senior Data Scientist, May 2020 – Jun 2021

- Developed a conversational autopilot AI tool for sales conversations by fine tuning GPT2 model and ensemble learners to track user intents and sales pitch score during sales conversations.
- Generated an average ROI of 0.8 by providing real-time cues and lead prioritization recommendations to sales agents.
- Assigned sales agent to customers based on agent's pitch score and customer's intent to buy. Agent re-assignment contributed to prevention of annual losses of $14 million for EdTech vendor.

IPsoft, Bangalore, India
Research and Development Engineer, Mar 2018 – Apr 2020

- Developed an automated customer service response system for IPsoft's digital employee 'Amelia' by training hierarchical bidirectional LSTM with customer urgency index modeling.
- Raised customer satisfaction rating to 4.3 from 3.2, and improved issue resolution rate by over 25%.
- Deployed the model using on-premise NVIDIA Triton Inference Servers, ensuring fault tolerant failovers and integrated with 'Amelia' using GRPC framework.
- Developed cascaded-CNN based face detection and face recognition customer authentication system.

Tonbo Imaging, Bangalore, India
Software Development Engineer, Jul 2015 – Feb 2018

- Developed a vehicle monitoring system using motion segmentation, Kalman tracking and CNNs applied to IR images from toll booth cameras.
- Built a scalable and modular network communication library based on UDP and TCP protocols for optimized and secure communication of tracker annotated video feed.
- Developed infrared image processing pipeline comprising of image smoothing, non-uniformity correction(NUC) and histogram equalization using CUDA C++ programming.

QUALIFICATIONS & SKILLS

- **Interests:** Data Science, Large Language Models, Conversational AI, Task Oriented Dialog Agents, Recommender Systems, MLOps, Distributed system design, Software Development
- **Software and frameworks:** AWS, PyTorch, Tensorflow, Trax, Scikit-learn, Numpy, Pandas, JAX, Apache Spark, Apache Kafka, Apache Airflow, Databricks, MS Excel, MS Access, Airflow, RDBMS, SQL, DynamoDb, MongoDb, Apache Cassandra, Docker, GIT, Atlassian JIRA, ElasticSearch
- **Operating Systems:** Windows, UNIX, Mac
- **Programming Languages:** Python, Java, C, C++
- **Personal:** Presentation and public speaking skills, collaborative research, highly organized, project leadership and team management

PUBLICATIONS

"Roll Up Your Sleeves: Working with a Collaborative and Engaging Task Oriented Dialogue System" - ACL/ISCA Special Interest Group on Discourse and Dialogue [SIGDIAL 2023]

"Optimal Route Searching in Networks with Dynamic Weights Using Flow Algorithms" - 2015 International Conference on Computational Intelligence and Communication Networks (CICN)

EDUCATION

The Ohio State University, Columbus, OH GPA: 3.85 (4.0 scale)
M.S., Computer Science and Engineering, Apr 2024
- **Thesis:** "Task Oriented Dialogue Systems"
Divided research into two phases. **Phase 1:** Exploring in-context learning and parameter efficient instruction fine-tuning of LLMs to develop domain specific and open-ended Question and Answering module for task-oriented dialogue agent. **Phase 2:** Utilizing PEFT techniques, knowledge distillation and model sharding with DeepSpeed's Zero Redundancy Optimizer (ZeRO Stage 3 Offload) for training LLMs.

Birla Institute of Technology and Science, (BITS), Pilani, India
B.E., Electrical and Electronics Engineering + M.Sc., Mathematics, Jul 2015

PROJECTS

Cross Task Collaboration in Large Language Models, Aug 2022 – Apr 2023
- Developed a novel approach for summary generation by using questions as cues.
- Combined T5 encoder, trained on question answering task, with T5, trained on passage summarization, using multi-headed self-attention.

Dialog Agent for Home Recommendation, Apr 2023 – Aug 2023
- Developed a chat-based requirement gathering tool for home recommendations.
- Created detailed dialog flow and NLU module to capture useful information from user responses.
- Used ElasticSearch to retrieve properties aligned with user preferences.