Saaketh Chaganty

Multi Agent LLM Inference · Scalable NoSQL · Full Stack AI · Backend Expert

Boston, MA

(551) 280 6937 | chaganty.s@northeastern.edu | saakethtypes.github.io | github.com/saakethtypes | linkedin.com/in/saakethchaganty

Education

3.5/4.0 **MS in Information Systems**, *Northeastern University* | Boston, MA *2022-24*

7.5/10 **Bachelor's in Mechanical Engineering**, *Mahindra University* | Hyderabad, India *2018-22*

Courses: DL | ML | Algorithmic Digital Marketing | Neural Modeling | Data Science Eng. Methods | Program Structures | Advanced Data Analytics | Research Methods in AI

Skills

Programming Python, JavaScript, C/C++, R, Java, Git, HTML

Software Linux, PyTorch, TensorFlow, OpenAI, SciPy, Matplotlib, Docker, Alexa Skills Kit, Keras, GCP, Azure, MongoDB, OpenCV, AWS Lambda, DynamoDB, NodeJs, ReactJS, Pandas, Snowflake, Supabase, LangChain, Weaviate, CrewAI, LangGraph

Soft Skills Time Management, Problem-solving, Leadership, Documentation, On-site coordination, Volunteer Developer.

Experience

Devocratic Technologies, *Lead Backend Engineer* | Delhi, India *April 2021 - Aug 2022*

- Created a highly scalable NoSQL single-table database and developed rate limited REST APIs for a prediction market trading platform (similar to nadex) in DynamoDB achieving secure concurrent 10K+ transactions for over 40,000+ users on 20k+ USD.
- Led cross functional teams constructing CI/CD pipelines automating content creation and customer support resulting in a cost savings of $850 per month

OTSI *Data Science Intern* | Hyderabad, India *March 2021 - May 2021*

- Deployed a robust face recognition OpenCV model for a facial attendance system at scale with successful prototyping within the company with 200+ employees.

UrbanKisaan, *AI Solutions Architect Intern* | Hyderabad, India *April 2020 - Sep 2020*

- Developed a faster R-CNN model for a smart hydroponics farm (92% accuracy) for precise identification of ripened tomatoes from plant images captured by a rover deployed in the farm achieving accurate results for 4 tonnes of harvest.
- Led a team of 4 people for streamlining end-to-end integration deployment for offering dynamic discounts on yield, acquiring a feature in TechCruch 2021 event presentations as an autonomous large scale farm and resulting in 23% increase in sales orders.

Projects

Multi-Agent-LLM System *Nov 2023 - Jan 2024*

- Utilized multi-agent LLM systems with self editing LTM, STM memory stores & dynamic role specifications for task orchestration to perform data analysis.
- Programmed RAG & prompt chaining techniques over the graph of thoughts framework for reiteration and efficient goal completion.
- Installed MAS on a smart hydroponics farm to enable natural language query pipeline for plant health check, anomaly & germination detection, control IoT devices and querying PostgresDB data.

Multi-Model Transformer for Weather Prediction *Sep 2023 - Nov 2023*

- Modeled a CNN Transformer architecture for satellite images & weather station data with Conv2D LSTM & GRU attention layers performing predictive forecasting of lake-effect precipitation at Lake Michigan.
- Designed 12 visual and statistical features and trained the model within 100 epochs to generate 3 days forecast for light rainfall with a 76% accuracy.

AI Powered Social Media *Mar 2023 - May 2023*

- Built a social media platform with personalized ad delivery by fine tuned a BLIP-2 ViT-G image captioning model to extract nouns from user posted images, enabling targeted ad placement based on image liking behavior.
- Developed a high-performance vector database using ImageNet pretrained weights for efficient image search over 10K+ social media images within 21ms in the DynamoDB & Streamlit stack to provide seamless state management for ad placements.

E-commerce Customer Churn Prediction *Mar 2023 - May 2023*

- Applied Logistic Regression using Snowflake & Scikit-Learn to predict churn customers expenditure.
- Performed GridSearchCV, XGBoost. oversampling, and SMOTE to enhance model performance and achieved a 90% accuracy on the sales data.

Plagiarism Paraphrase Detection *Mar 2023 - May 2023*

- Developed a multilingual plagiarism detection tool using bert-base-uncased tokenizer, cohere reranking, and HuggingFace models, applied to a diverse dataset of Arxiv papers.
- Enhanced paraphrase analysis using a PEFT adapter on LLama, enabling detailed detection of paraphrasing techniques like coordination changes and negation switching among authors in the pool.