

PRATHAMESH PRAVIN MORE

prpmore@gmail.com (930)-333-4542 linkedin.com/in/more-prathamesh github.com/Spidey13 Publications

PROFESSIONAL SUMMARY

M.S. Data Science graduate specializing in productionizing end-to-end Generative AI and Agentic AI solutions. Proven experience architecting advanced Retrieval-Augmented Generation (RAG) pipelines, developing with Agentic AI frameworks (LangChain), and deploying full-stack ML applications using Python, FastAPI/Flask, and vector databases (Pinecone). Eager to define GenAI architecture and guide development teams at Infosys.

EDUCATION

Master of Science in Data Science Indiana University, Bloomington, IN 05/2025
Relevant Coursework: Computational Linguistics, LLMs, Big Data Economics, Statistics, Data Mining, Data Visualization
Bachelor of Technology in Computer Engineering University of Mumbai, Mumbai, India 07/2023
Relevant Coursework: Data Structures and Algorithms, Big Data Analytics, Artificial Intelligence and NLP

PROFESSIONAL EXPERIENCE

Data Science Intern *Indiana University — Bloomington, IN* 05/2024 to 08/2024

- Architected advanced topic modeling pipelines using **LDA, NMF, K-Means, and Spectral Clustering** to analyze **550+ architectural and design publications** for **DEI themes** achieving **93% Topic Coherence** and developing a modular text-mining framework whose findings were published in *The Design Journal (Taylor & Francis, Feb 2025)*.
- Engineered interactive **Plotly and Dash** dashboards, for temporal analysis of **150+ DEI-related terms**, with dimensionality reduction and **word co-occurrence analysis** via and visualize emerging semantic trends.
- Enhanced insight granularity and actionability by partnering with **domain experts** to iteratively refine feature selection and interpret topic modeling outputs.

Machine Learning Engineer Intern *Dimensionless Technologies — Mumbai, India* 05/2023 to 07/2023

- Devised an **NLP pipeline** using **BERT** with AWS, achieving precision of **0.84** to continuously analyze articles from CNN and update a **real-time dashboard** with categorized insights for **trend monitoring and market analysis**.
- Implemented an **OCR model** for industrial documents and deployed a **QA system** that improved **information retrieval speed by 70%**, automating data extraction and streamlining the Tender Document creation process.
- Partnered with clients and cross-functional teams to refine **NLP models** based on business needs and contributed to **data strategy, model interpretability, and deployment discussions**.

Full Stack Developer Intern *Benchmark Computer Solutions Pvt. Ltd. — Mumbai, India* 08/2022 to 10/2022

- Engineered advanced NLP modules using **SpaCy and Transformers** for automated resume parsing, improving entity extraction F1-score from 0.78 to **0.91** and cutting manual annotation cycles by **40%** through active learning.
- Developed and deployed a full-stack **Flask web application** and a **RESTful API** to process over 1,000 resumes, automating the data ingestion and analysis pipeline and reducing recruiter screening time by **30%**.

SKILLS

Programming Languages & Data Tools: Python, R, SQL, Java, C++, Scala, MySQL, PostgreSQL, MongoDB, Firebase, Pinecone, Spark, Hadoop, Airflow

Machine Learning & NLP: Transformers, LLMs, LSTMs, GRUs, CNNs, RNNs, YOLO, SMOTE, Feature Engineering, Hyperparameter Tuning, PCA, t-SNE, SHAP, Named Entity Recognition, Sentiment Analysis, Topic Modeling, Text Summarization, OCR, SpaCy, NLTK, LangChain

Cloud & MLOps: AWS (S3, SageMaker, ECS, Lambda), Azure (Blob Storage, Cognitive Services), MLflow, DVC, Kubeflow, Docker, Kubernetes, GitHub Actions, CI/CD, FastAPI, REST APIs

Frameworks & Visualization: TensorFlow, PyTorch, Scikit-Learn, NumPy, Pandas, SQLAlchemy, Flask, Django, Matplotlib, Seaborn, Plotly, Grafana, NetworkX, Power BI, Tableau

Projects

Audio-Lyric Emotional Alignment — OpenL3, SentenceTransformers, Librosa, PCA, K-Means, NLTK, Matplotlib

- Engineered an **agentic AI workflow** using **LangChain** to automate multimodal music analysis, orchestrating multiple LLM calls and data processing steps to quantify emotional congruence between audio and text.
- Implemented an advanced **Retrieval-Augmented Generation (RAG)** pipeline to provide rich, contextual data from a vector database to LLMs, significantly improving the quality and relevance of generated insights.
- Leveraged generative models for zero-shot contextual summarization of lyrics, enriching textual features and boosting downstream emotion classification accuracy to **92%** against human-annotated benchmarks.

F1 Race Strategy Simulator — CatBoost, XGBoost, MLflow, DVC, Docker, FastAPI, MongoDB, Streamlit, Plotly

- Engineered a **high-fidelity Digital Twin** of F1 race dynamics, leveraging **CatBoost and XGBoost** on high-frequency telemetry data to accurately model and predict critical variables like lap times and tire degradation curves.
- Developed a **multi-agent simulation environment** to test and optimize race strategies, exploring reinforcement learning principles to identify optimal pit stop timings and driver tactics under various competitive scenarios.
- Operationalized the full simulation and inference pipeline using **MLflow for versioning** and **Docker for deployment**, delivering actionable strategic insights to a **Streamlit-based Decision Intelligence dashboard**.