Vivek Maurya

linkedin/vivekcmaurya | github/bibekku

Email : vivekcmaurya30@gmail.com

Mobile : +1(516)864-9896

Programming Skills

- **Languages**: Python, SQL, Java, Javascript, TypeScript, HTML, CSS, Bash, Scala, LaTeX
- **Libraries/Frameworks**: pandas, NumPy, scikit-learn, TensorFlow, PyTorch, matplotlib, Spring, node.js, react.js, d3.js
- **MLOps/DevOps**: Docker, Kubernetes, Linux, Git, AWS, CI/CD, Kafka, Integration Testing, Automated Testing
- **Databases**: PostgreSQL, MongoDB, AWS DynamoDB, AWS OpenSearch Serverless, Hadoop, Pinecone, ElasticSearch
- **Paradigms**: Data Science, Machine Learning, GenAI, Object Oriented Programming, Microservices, REST, Scrum, Agile
- **Tools**: PyCharm, Visual Studio Code, IntelliJ, Kibana, Android Studio, Xcode, DBeaver, regex, Postman, VirtualBox

Work Experience

- **Stony Brook University** – Visiting Scholar | *Sep 2024 - Present*
  - Enhanced Retrieval-Augmented Generation (RAG) systems, ensuring seamless integration with Generative Engine results.
  - Conducted research on foundation models and fine-tuned LLAMA3, Mistral 7B, and Claude OPUS models for optimized performance using Parameter-Efficient Fine-Tuning (PEFT) including p-tuning, LoRA, and IA3.
  - Improved user experience by optimizing prompt engineering and reducing system latency.
  - Developed a Retrieval-Augmented Generation (RAG) system using FastAPI, integrating Hugging Face LLM, LangChain, and LlamaIndex for efficient querying and document clustering and information retrieval across multiple PDFs.
  - **Technologies**: GenAI, RAG, LangChain, FastAPI, HuggingFace
- **Amazon** – Software Development Engineer (Alexa NLU) | *Jun 2018 - Jan 2021*
  - Integrated Alexa compatibility into Amazon Pay backend (Java) microservices, A/B tested and launched in NA, EU, and Japan.
  - Built reusable React-Native components for Amazon Pay in the Alexa app, tested on Android/iOS, became SME for future mobile app changes.
  - Developed NLU and backend (Java) components for US 2020 election donations via Alexa, ensuring legal compliance. Reduced dev effort by 2 SDE-months by identifying requirement gaps early.
  - Wrote NLU grammars for fuel payments via Alexa Auto, resolved conflicts with existing utterances, collaborated with Fiserv, and integrated JWT-based API authentication.
  - Maintained CI/CD, integration tests, and handled on-call rotations. Automated merchant onboarding with bash scripts, eliminating human error and cutting onboarding time from 2 hours to 10 minutes.
  - **Technologies**: Java, Spring, JUnit, CI/CD, AWS, Linux, React-Native, NLU, Test-Driven Development, bash, Android, iOS
- **Amazon** – Software Development Engineer - Intern | *May 2017 - Jul 2017*
  - Built PoC for a data migration pipeline backfilling historical data from legacy RDBMS to AWS DynamoDB, earning commendations as an intern.
  - Created a Java library to generate complex JSON for testing data migration.
  - **Technologies**: Java, Spring, CI/CD, AWS DynamoDB, Linux, Test-Driven Development, JSON

Projects

- **Improving AI Code Generation**: Worked on an AI training project via Outlier to identify and improve model weaknesses in code generation. Evaluated responses, provided targeted feedback, and refined outputs to enhance instruction following, executability, and correctness.
- **Optiver Trading at the Close**: Performed time-series analysis and feature engineering to develop an LSTM model for financial market prediction.
- **Django Portfolio Website**: Created a portfolio website using Django and PostgreSQL for an economics professional, hosted on a Hostinger Ubuntu VPS.
- **Music Taste Visualizer**: Used node.js and d3.js to visualize the user's Spotify likes and provide an interactive taste profile which can be used to request more specific recommendations from Spotify.

Education

- **Stony Brook University** – Master of Science in Computer Science; GPA: 3.81/4.00 | *Aug. 2022 – May 2024*
- **Veermata Jijabai Technological Institute** – B.Tech. in Information Technology; CGPA: 8.63/10.00 | *July 2014 – May 2018*