Anurag Mishra

Medium: anuragmishra_27746
Github: github.com/anurag-mishra899

Email: anurag.mishra899@gmail.com
Mobile: +91-9456243057
LinkedIn: in/anurag-mishra-660961b7

PROFILE SUMMARY

- **Experience**: **6.5+ YoE** in designing and developing **Machine Learning, Deep Learning, Generative AI, and NLP** solutions for various industries, including retail, banking, telecommunications, and chemicals
- **Frameworks**: PyTorch, Langchain, LlamaIndex, Fastapi, Scikit, NLTK, SpaCy, TensorFlow, Keras
- **Tools**: Kubernetes, Docker, GIT, PostgreSQL, MySQL, SQLite
- **Platforms**: Azure Cloud, AWS, GCP, Linux
- **Certified in**: **Azure Fundamental, Azure Data Scientist, Databricks ML Professional**
- **Publications**: Article 1, Article 2, Article 3, GenAI Conference

EXPERIENCE

- **EY GDS** — Hybrid
  Associate Manager (Full-time) — *Feb 2022 - Present*

  - **Deployed GenAI Chat-bot Solution for 40 Branches, Serving 2,000-3,000 Users**:
    * **Led a team of six engineers** to **design** and **deploy** a Chat-bot solution for **2,000-3,000 users** across **40 branches** of the largest bank in the Philippines.
    * **Reduced query resolution time** from 5-10 minutes to under 30 seconds, **boosting customer satisfaction by 45%**.

  - **Developed LLM Production Pipeline, Reducing Downtime by 50%**:
    * Engineered a custom preprocessing and chunking pipeline to **manage complex tables, PDF formatting, and flowcharts, increasing retrieval performance by 55%**.
    * Built a scalable deployment pipeline using FastAPI, Langchain, Azure API Gateway, and Kubernetes, **supporting throttling, fallbacks, and region-wise load balancing OpenAI models, reducing downtime by 50%**.

  - **Implemented LLMOps and Evaluation Pipeline, Boosting Accuracy by 46%**:
    * Introduced **Custom RAG evaluation framework for automated end-to-end assessment**, enhancing performance and optimizing solution components such as routing, retrieval, and generation, resulting in a 46% increase in accuracy.
    * Developed a monitoring pipeline using **Langfuse for observability and logging**, reducing issue resolution time by 30%.
    * Established **feedback mechanism, achieving a 25% improvement** in solution accuracy over six months.

  - **Smart Call-Centre Solution for US-Based Client, Increasing Agent Productivity by 60%**:
    * Built LLM and data ingestion pipeline integrating Azure DevOps, Langchain, and ServiceNow, managing post-call actions including summarization, ticket creation, updates, and quality assurance.
    * Designed **pipeline to handle 300-400 calls per center**, increasing agent productivity by 60% and reducing average call handling time by 40%.

  - **Fine-tuning Transformer Models Reducing Operational Costs by 70%**:
    * Delivered advanced techniques in deep learning model tuning, leading to significant improvements in **automation extraction workflows** across the organization; measured performance gains resulted in reduced processing times of critical data inputs by over 80% hours monthly.
    * Deployed deep learning models as **batch endpoints utilizing Azure Web Apps, Databricks Jobs, and MLflow**, ensuring streamlined integration and efficient model management, reducing operational costs by 70%.

- ○ **Executed Go-to-Market Strategy, Increasing Client Engagement by 35%**:
  - ∗ Contributed to the "Go to Market" strategy for Generative AI across various clients, developing prototypes for industry use-cases using models like OpenAI, LLaMa, MPT, and Dolly, resulting in a 35% increase in client engagement.
  - ∗ Conducted client awareness sessions on **"Industry Implementation and Risks of GenAI," educating over 200 clients** on the practical applications and potential risks of generative AI technologies.
  - ∗ Successfully drove the adoption of Generative AI solutions in **three major client projects, increasing revenue by \$500K**.

- **Deloitte (Office of US)** <div align="right">Remote</div>
  Consultant - AI (Full-time) <div align="right">*Feb 2021 - Feb 2022*</div>
  - ○ **Fixture Optimization for Inventory Management, Achieved \$5 Million Cost Reduction**: Devised optimization solution to allocate the optimal number of fixtures for each product and brand in retail chain stores, resulting in a \$5 million reduction in costs and expenses.
  - ○ **Sales Forecasting for Retail Client, Enhanced Demand Forecasting by 45%**: Developed sales forecasting model for each department per store to assess demand. Utilized tree-based algorithms such as Gradient Boosting, XGBoost, and LightGBM, as well as regression-based models like ARIMA & SARIMA.

- **Virtusa Polaris** <div align="right">Hybrid</div>
  Machine Learning Engineer (Full-time) <div align="right">*June 2018 - Feb 2021*</div>
  - ○ **Multi-Class classification of Trading Products, Reducing Manual Tagging from 15k/hour to 1 Million/hour**: Devised solution for Product Controllers to reduce person-hours spent on tasks of identifying breach in trading activity.
  - ○ **NLP Models to Assess Commentary using SpaCy, Improving Comments Quality by 80%**: Developed NLP solution to identify the inconsistency in PnL commentary generated by Product Controllers. Milestones included building Rest API to render user request and ETL operations to handle user requests.
  - ○ **ML Models & Techniques**: Utilized various ML algorithms like Random Forest, Logistic regression Gradient Boosting and Hyper-parameter tuning to get 95% accuracy.
  - ○ **Tools & Technologies**: Used Python programming and worked on libraries like Pandas, Scikit-learn, NumPy, Matplotlib, Seaborn, Spacy

PROJECTS

- **Fine-Tuned LLM for Domain Specific tasks using LoRA, Increasing Domain-based NER By 55%**: Performed Fine-tuning of LLM models(LLaMA 3.1 8B, Mistral 7B) on 80 GB A-100 GPU on custom build dataset for Manufacturing domain.

- **LLM Based Agentic flow for Text to Code Solution**: Agent based solution that convert requirements to code. The solution uses multi-agents (debugger, coder, tester etc) to write the bug-free and optimized code.

EDUCATION

- **National Institute of Technology, Manipur** <div align="right">India</div>
  Bachelor of Technology - Information Technology; GPA: 8.5 <div align="right">*July 2014 - June 2018*</div>
  Courses: *Operating Systems, Data Structures, Analysis Of Algorithms, Artificial Intelligence, Machine Learning, Networking, Databases*

HONORS AND AWARDS

- Awarded Excellence Client Performance Award in EY

- Honored to be among top 10 students in CSE Department

- Achieved (Among Top 1 % in board) in the entire district in HSC exam.