Pranav Saran

US Citizen | 510-709-7661 | pxs878@case.edu | linkedin.com/in/pranavsaranaway | github.com/pranavsaranaway

Education

Case Western Reserve University (GPA: 4.0/4.0) **Cleveland, OH**

Computer Science B.S, Computer Engineering B.S.E *Aug. 2024 – May 2028*

- **Coursework**: Data Structures, Algorithms, Computer Security, Discrete Math, Logic Design and Computer Organization, Linear Algebra, Database Systems
- **Involvement**: Deans High Honor List, University Program Board, Club Badminton Events and Outreach Manager, Club Badminton Tournament Team, Global Health Design Collaborative - Pulse Ox Team

Experience

Machine Learning Researcher **Jan. 2025 - Jun. 2025**

Algoverse *Palo Alto, CA*

- Achieved 83% idiom and 78% metaphor accuracy with a hybrid mBERT+BiLSTM for low-resource Konkani.
- Reduced model size via pruning while retaining 100% idiom and 88% metaphor accuracy.
- Benchmarked robustness with ablation studies on mBERT, IndicBERT, and XLM-R.

Software and Electronics Engineer **Sep. 2024 – Present**

Case Global Health Design Collaborative *Cleveland, OH*

- Built a custom ESP32-C3 Soft AP to stream MAX30102 data, boosting real-time monitoring efficiency by 40%.
- Developed a Python/Kivy app for heart rate and SpO_2 visualization from 500+ sessions, raising engagement 60%.
- Improved data reliability 40% with a detrending algorithm; MVP award, and 2nd Place at CWRU Intersections.

Projects

GenMe - A ReadME Generator Agent | *Python, SmolAgents, OpenAI GPT-4, Gradio*

- Built an autonomous LLM agent to analyze large codebases and auto-generate technical documentation.
- Developed a toolchain with Gradio for code parsing, dependency inference, and human–agent collaboration.

Med Lens - Multi Task Clinical NLP Pipeline | *BERT, HuggingFace, MIMIC-III*

- Fine-tuned BioBERT on 40K+ MIMIC-III notes, reaching 91% ICD F1 and 89% NER accuracy.
- Boosted summarization (+18%) and cut model size 30% with ROUGE benchmarking and pruning.

Clarence: LLM Benchmark Agent | *Python, SmolAgents, OpenAI API, LLM Tooling*

- Built a multimodal reasoning agent with smolagents and GPT-4 for automated search, code, and file tasks.
- Implemented tool orchestration and REST API pipelines for benchmark evaluation via HF spaces.

Publications

- **Pruning For Performance: Efficient Idiom and Metaphor Classification in Low-Resource Konkani Using mBERT**. *The 2025 Conference on Language Modeling, Workshop in Multilingual Data Quality Signals(COLM 2025)*, 2nd Author

Technical Skills

Languages: Python, Java, C++, C#, JavaScript, SQL(Postgres), Bash
Machine Learning: PyTorch, TensorFlow, Scikit-learn, Transformers, HuggingFace, Jupyter
Frameworks: LangChain, Next.js, FastAPI, React, REST APIs, WebSockets, Node.js, Express
Data Systems: Hive, PySpark, Apache Airflow, Milvus, BigQuery, Azure SQL, GCS, PostgreSQL
Infrastructure & Dev Tools: Docker, CUDA, GPU Programming, Git
Concepts: Data Pipelines, LLM Inference, RAG Architectures, Agentic Systems, Attention Mechanisms, Production Deployment, Backend/Full-Stack Development, CI/CD, Software Engineering, Microservices, Distributed Systems