GNANESWAR VILLURI

Stony Brook, USA ◇ +1 (934)642-0363 ◇ villurignanesh@gmail.com ◇ linkedin.com/in/villurignanesh

SUMMARY

Innovative AI researcher and engineer with a proven track record in developing cutting-edge ML solutions, optimizing NLP pipelines, and driving technological advancements across academia and industry.

EDUCATION

PhD in Computer Engineering, Stony Brook University, **GPA**: 4.0 — 08/2024 - 12/2026 (Expected)

- **Research Focus:** *Scalable ML algorithms, deep learning, and reinforcement learning for multi-modal data analysis.*

MS in Computer Engineering, Stony Brook University, **GPA**: 3.8 — 08/2022 - 05/2024

BS in Computer Science and Information Technology, JNTU Kakinada, **GPA**: 8/10 — 08/2017 - 05/2021

SKILLS

Programming Languages	Python, C++, JavaScript, SQL
AI & Machine Learning	Machine Learning, Deep Learning, Natural Language Processing (NLP), Large Language Models (LLM), Text Classification, Text Generation, Translation, Automatic Speech Recognition (ASR)
LLM Techniques	Finetuning (LoRA, QLoRA), Reinforcement Learning (RLHF, PPO, DPO)
Libraries & Frameworks	PyTorch, TensorFlow, Scikit-learn, OpenCV, Hugging Face, SpaCy, NLTK, Langchain
Data Science Tools	Pandas, NumPy, Matplotlib
Mobile Development	React Native, Flutter
MLOps & DevOps	Git, Docker, Linux, MLflow, DVC, CML
Data Management & Storage	Firebase, Redis, Vector Databases (Qdrant, Pinecone), Hadoop, MapReduce

PUBLICATIONS

- Villuri,G., Shaik,H., & Doboli,A.(2025). A Stacked Multi-Layered Perceptron - LLM Model for Extracting the Relations in Textual Descriptions. In 2025 IEEE Symposium Series on Computational Intelligence , Trondheim, Norway, March 17-20, 2025. **(SSCI 2025)**
- Villuri,G., & Doboli,A.(2024). Towards Semantic Classification of Dialog using Contextual Prediction Networks. In Proceedings of the 7th Annual Conference on Cognitive Computational Neuroscience. Cambridge, MA, USA. **(CCN 2024)**.
- Villuri,G., & Doboli,A.(2024). Using Speech Data to Automatically Characterize Team Effectiveness to Optimize Power Distribution in Internet-of-Things Applications. 2024 Conference on Information Technology and Data Science. **(CITDS 2024)**.
- Villuri,G., Pallapu,H., Simona,D., & Doboli,A.(2024). Automatically Understanding Human Behavior for IoT Applications with Optimized HITL Control. 2024 Int. Conf. Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design , Volos, Greece, Jul. 2-5, 2024. **(SMACD 2024)**.

RESEARCH EXPERIENCE

Graduate Research Assistant - Stony Brook University — 01/2023 - Present

- Designed and implemented a hybrid ML pipeline integrating DistilBERT embeddings with gradient-boosted decision trees, improving accuracy by 24% and reducing computational overhead by 30%.
- Implemented NLP pipelines using CoreNLP, SpaCy, and Whisper for multi-modal conversational analytics.
- Researched SLU methodologies, focusing on transformer architectures for intent recognition and slot filling.
- Applied causal tracing and ROME for model interpretability in large language models.
- Contributed to ONR project on team optimization using RL and multi-agent systems, improving performance by 22%.

PROFESSIONAL EXPERIENCE

Software Developer Intern - Zippi Delivery (Stony Brook, NY) 05/2023 - 08/2023

- Developed cross-platform applications using Flutter and Firebase, increasing order volume by 60%.
- Designed branded restaurant landing pages, attracting 1,000+ new customers and boosting daily orders by 40%.
- Integrated DoorDash Drive API, reducing delivery delays by 25% during peak hours.
- Created and launched a loyalty program, increasing repeat orders by 30%.
- Integrated ACH payment option, reducing transaction fees by 15%, saving 10% on processing costs.

Machine Learning Engineer - Accenture PLC (Bengaluru, India) 01/2022 - 07/2022

- Leveraged frequent pattern mining to segment e-commerce customers based on browsing and purchase behavior.
- Improved client's customer retention by 18% through personalized product recommendations.
- Increased average order value by 12% among segmented customer groups using advanced ML techniques.
- Collaborated with cross-functional teams to integrate ML models into production systems, ensuring scalability and performance.

Software Engineer - ABDA Digital (Hyderabad, India) 11/2020 - 01/2022

- Developed Hola Enterprise, a SaaS platform for customizable video generation, increasing user engagement by 30%.
- Engineered modular web components, reducing video rendering time by 25% and boosting adoption rates.
- Implemented advanced designer features, improving UX scores by 45% and decreasing design time by 35%.

PROJECTS

SmartTutor: RAG-based Learning Assistant

- Engineered AI-powered tutor utilizing RAG for Stereo Vision, processing 300 pages and generating 10,000 vector embeddings.
- Integrated Mistral AI (92% accuracy) and custom RAG model (94% accuracy), enhancing overall performance.
- Boosted retrieval accuracy to 95%, achieved 2-second average response time, and slashed user study time by 25%.
- Circumvented hardware constraints through strategic model selection, implementing a functional system with distil-GPT2.

Language Neutralization System

- Architected real-time translation system for call centers, supporting 10+ languages with 95% accuracy.
- Constructed streaming pipeline incorporating Speech Recognition, Text Translation, and Text-to-Speech models, reduced processing delay by 40%.
- Streamlined system for accelerated machines, yielding 30% performance boost.
- Orchestrated system deployment via Docker, minimizing setup time by 60%.
- Established API web socket, amplifying system accessibility by 80% for call center applications.

Agent Performance Platform

- Led the development of the platform creation using text classification to analyze agent conversations, handling 1000+ daily interactions.
- Fine-tuned model to attain 85% accuracy in identifying and scoring 15 distinct soft skill traits.
- Executed MLOps practices, accelerating model update time by 70% and enhancing overall accuracy by 10%.
- Revolutionized agent performance evaluation, boosting efficiency by 50% and trimming manual review time by 75%.