GNANESWAR VILLURI

Lake Grove, USA ◇ +1 (934)642-0363 ◇ villurignanesh@gmail.com ◇ Google Scholar ◇ LinkedIn ◇ Portfolio

SUMMARY

AI researcher with experience at **Amazon AWS** and **Nokia Bell Labs**, focused on building and deploying agentic, tool-augmented LLM systems that deliver measurable impact across production and research settings.

EDUCATION

PhD in Computer Engineering, Stony Brook University, **GPA**: 4.0 — 08/2024 - 05/2027 (Expected)

- **Research Focus:** *Neuro-symbolic and tool-augmented large language models for reliable reasoning, with an emphasis on verification-aware prompting, reinforcement learning for tool-use policies, and agentic collaboration in open-ended coding and problem-solving tasks.*

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

SKILLS

Programming	Python, C++, JavaScript, SQL
LLMs & Reasoning	Tool-Augmented & Agentic Systems, Prompt Optimization (APO), Function Calling Reinforcement Learning, Chain-of-Thought, Verification-Aware Reasoning
Model Training & Adaptation	Fine-tuning (LoRA, QLoRA), Instruction Tuning, Evaluation & Benchmarking
Frameworks & Libraries	PyTorch, TensorFlow, Hugging Face, LangChain, NNsight, SpaCy, NLTK
Multimodal & Speech	Automatic Speech Recognition (Whisper), Multimodal Pipelines (Text–Audio–Code)
MLOps & Deployment	Git, Docker, Linux, MLflow, DVC, CML, SageMaker, Bedrock
Data & Knowledge Systems	Vector Databases (Qdrant, Pinecone), Knowledge Graphs, Neo4j, GraphRAG

PUBLICATIONS

- Shaik,H., Villuri,G., & Doboli,A.(2025). An Overview of LLMs and a Novel, LLM-Based Cognitive Architecture for Solving Open-Ended Problems. In *Machine Learning and Knowledge Extraction*, 7(4), 134. **(MAKE 2025)**
- Shaik,H., Villuri,G., & Doboli,A.(2025). Concept Combinations with Generator and Validator Agents Prompted Using Insights from Concept Networks. In *International Conference on Complex Networks*. **(Complex Networks 2025)**
- Villuri,G., & Doboli,A.(2025). An Experimental Study on the Interpretability of Transformer Models for Dialog Understanding. In 2025 IEEE Conference on Artificial Intelligence, Santa Clara, CA, May 05-07, 2025. **(CAI 2025)**
- 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, USA. **(CCN 2024)**.

RESEARCH EXPERIENCE

Graduate Research Assistant – Stony Brook University — 01/2023 – 05/2027

- Built LLM-based validation agents to detect errors in open-ended collaborative coding tasks.
- Implemented verification-aware generator–validator pipelines for static and real-time team evaluation.
- Designed neuro-symbolic dialogue and code representations using operational and axiomatic semantics.
- Developed tool-augmented LLM pipelines with knowledge graphs and multimodal NLP.
- Created graph-augmented agent frameworks for relational reasoning and agentic collaboration.
- Applied reinforcement learning and multi-agent methods for tool-use and coordination policies.
- Led research on reliable LLM reasoning by integrating validation, symbolic structure, and agentic workflows, resulting in peer-reviewed publications across multiple venues.

PROFESSIONAL EXPERIENCE

Applied Scientist Intern – Amazon AWS (Mountain View, CA) 09/2025 – 12/2025

- Conducted extensive empirical evaluation of the "GEPA" Automatic Prompt Optimization (APO) framework, assessing its efficacy in replacing hand-crafted prompts with automated optimization strategies.
- Validated improvements in structured reasoning accuracy of up to 12%, demonstrating that algorithmic search over instruction space significantly outperforms naive prompting on complex tasks.
- Benchmarked optimization performance across 10 production LLMs (1B–405B), identifying a critical "sweet spot" where mid-sized models yielded the highest relative gains.
- Investigated cross-model transferability, discovering an asymmetric pattern where prompts optimized for mid-sized models generalize effectively to larger architectures, enabling efficient prompt reuse.

ML Research Intern – Nokia Bell Labs (Murray Hill, NJ) 06/2025 - 08/2025

- Developed a multi-agent LLM pipeline using OPC UA to extract insights from industrial telemetry data.
- Designed hierarchical agents for high-level question generation and low-level insight extraction.
- Integrated knowledge graphs and function calling to correlate multi-layered industrial events.
- Investigated prompting strategies to enhance LLM accuracy in complex industrial NLP tasks.

Software Developer Intern - Zippi Delivery (Stony Brook, NY) 05/2024 - 08/2024

- Built AI-driven restaurant management platform using GoHighLevel MCP and LLMs for unified business operations.
- Developed multi-agent "Zippi" system with CrewAI achieving 95% task routing accuracy.
- Designed a hybrid LLM setup (Gemini + DeepSeek) cutting costs by 40% and manual oversight by 65%.
- Created webhook automation handling 10K+ social interactions daily, improving response times by 50%.

PROJECTS

SmartTutor: GraphRAG-based Learning Assistant

- Developed an AI-powered tutor leveraging GraphRAG, replacing traditional RAG with a knowledge graph for improved accuracy of 34% and enhanced interpretability.
- Processed 300 pages on Stereo Vision, generating a structured knowledge graph instead of word embeddings.
- Achieved a 2-second response time and cut user study time by 25%, simplifying complex concepts.
- Enabled real-time cache visualization, offering a transparent and explainable retrieval mechanism.
- Overcame hardware constraints by strategic model selection, deploying a functional prototype using distilGPT2 with an optimized graph-based retriever.

Agent Performance Platform

- Led platform development using text classification to analyze agent conversations, processing 1,000+ 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%.

OPEN-SOURCE CONTRIBUTIONS

GEPA – Automatic Prompt Optimization Framework

- Contributed upstream improvements to GEPA Automatic Prompt Optimization (APO) framework, enhancing prompt search robustness, evaluation consistency, and reproducibility across multi-model benchmarks (merged PR #147).
- Ongoing contributor focused on advancing automated prompt optimization, cross-model generalization, and agentic reasoning workflows.