BALA GUHANESH

Chennai, India | (+91) 95006 64509 | gsbalaguhanesh@gmail.com | LinkedIn | Github | Website

EDUCATION

Vellore Institute of Technology **Chennai, India**
B.Tech in Computer Science in specialization with AI and Robotics *Sep 2021 - May 2025*
Cumulative GPA: **8.52/10**

SUMMARY

Highly motivated and skilled B.Tech Computer Science student specializing in AI and Robotics with a strong foundation in full-stack web development and machine learning. Proven ability to develop and deploy web applications, Agentic workflows, manage LLMs, and contribute to innovative research projects. Seeking opportunities to leverage expertise in AI, machine learning, and web development to create impactful solutions. Possessing 2 years of experience in managing LLMs using creating web development interfaces. My greatest strength is my ability to learn and adapt quickly. Let's connect and build something meaningful, even if the journey is challenging.

EXPERIENCE

AI/ML Engineer **Remote**
Freelance *Jan 2023 - Present*

- Implemented a microservices architecture using FastAPI for LLM deployment, optimizing API endpoints for speed. Developed a Django-based front-end with a focus on usability and data visualization. Employed techniques like caching and asynchronous task processing to handle high request volumes. Utilized the Model-View-Template (MVT) architecture in Django to maintain a clean and scalable codebase.
- Achieved an average accuracy improvement of 18% on downstream tasks after fine-tuning (evaluated using metrics like F1-score and BLEU score). Reduced model size by 60% through quantization and pruning, enabling deployment on resource-constrained environments (e.g., edge devices). Data augmentation increased dataset size by a factor of 2.5x, significantly improving model generalization. Django query optimization reduced database query time by 25%.

Research Intern **Alberta, Canada**
MITACS Globalink Research Internship *May 2024 - Aug 2024*

- Worked under the supervision of Dr. Maiga Chang at Athabasca University, Edmonton, on the research project titled "Key Phrase Extraction for Enhanced User Question Answering". Developed a Next Word Prediction Model utilizing N-grams and Achieved 94% accuracy with the Binary Neural Network. Created a custom Tokenizer on DB-Pedia ENG 3.8 dataset and enhanced the Tokenizer's subword handling efficiency by 25%, improving model throughput by 15%.
- Developed a Python library that enables novice users to interact with various large language models (LLMs) for integration into their own applications. My work also involved prompt engineering to enhance response accuracy and improve user interactions.

Research Lab Assistant **Chennai, India**
Data Analytics and Robotics Lab, VIT Chennai *Sept 2022 - May 2024*

- Led research in pattern recognition and Developed algorithms that achieved classification accuracy of 87% for high-resolution raster data. Enhanced operational efficiency by 30% through Sparse Attention and Transformer integration.
- Contributed to the development of web-based data visualization tools for research findings, improving accessibility and understanding of complex data.

Machine Learning Intern **Mumbai, India**
Digital EPCS Pvt Ltd *Aug 2022 - Sep 2022*

- Developed an E-learning chat bot utilizing 3D CNNs to capture both spatial and temporal features from video frames. This reduces learning time for learners by 40% through targeted video queries.
- Integrated Retrieval-Augmented Generation (RAG) with a web scraping mechanism to enhance the chatbot's ability to generate context-aware answers from online sources, enriching the overall user experience and improving the contextual accuracy by 20%.
- Designed and implemented a REST API endpoint to serve chatbot responses, ensuring scalability and reliability.

Research Intern **Bangalore, India**
Ratan NVP groups *May 2023 - July 2023*

- Developed a continual learning model to address frequent retraining challenges, reducing retraining frequency by 35%. Leveraged a Neuro-modulated Meta-Learning architecture with selective activation mechanisms, achieving an average task accuracy of 92.4% across 10 benchmarks. Optimized Diffusion model performance by refining Gaussian noise addition (variance tuned to 0.01) and retrieval processes, resulting faster convergence during training and a 15% reduction in sample generation time.

LEADERSHIP EXPERIENCE

Research Project **Chennai, India**
AI/AR Integrated Smart Glasses for Visually Impaired *Feb 2023*

- Developed AI/AR smart glasses with ESP32 cameras, multi-sensor arrays, and GPS, achieving a 65% improvement in usability. Implemented object detection with 95.3% accuracy and optimal path recognition with less than 150 ms latency using transformer-based models. Integrated web scraping models to provide real-time contextual insights, reducing response time by 20% and enhancing environmental understanding for visually impaired users.
- Created a web-based interface for configuring and monitoring the smart glasses, allowing for remote management and updates.

IoTHINC Club **Chennai, India**
Club Member/ Secretary, VIT Chennai *Jan 2023 - Jan 2024*

- Led a team of 5 members in developing a LiDAR drone system capable of processing terrain data and performing autonomous navigation in complex environments with 85% accuracy. Designed and implemented a real-time 3D mapping and terrain analysis pipeline, reducing terrain processing time by 32.3% using LiDAR data. Integrated QGIS for geospatial data processing, improving data handling efficiency by 25.1%, which significantly enhanced the drone's adaptability to diverse topographies.
- Developed a web application for visualizing drone-collected data, providing interactive maps and analysis tools.

QuantumML Hackathon **Chennai, India**
Runner up, VTT Chennai *Nov 2024*

- During a hackathon, I led my team in tackling the challenge of efficiently clustering high-dimensional data using quantum computing. We implemented a quantum version of the K-means clustering algorithm using Quantum Approximate Optimization Algorithm (QAOA) to encode data points and perform distance calculations, while leveraging a classical optimization step to update cluster centers.
- As a result, under my leadership, we significantly reduced the time complexity of the clustering process, improving performance by approximately 30% compared to the classical approach on datasets with over 1,000 features, showcasing the potential of QAOA-based quantum algorithms to enhance machine learning tasks.

PUBLICATIONS

G S Bala Guhanesh, "A Comprehensive overview of Code Language models" (Under Review) **May 2024**
- Authored and submitted a review paper on the progression of code-based language models, from the introduction of attention mechanisms to the latest innovations in code generation providing a comprehensive overview on Code Language models (CLMs).

G S Bala Guhanesh, M Vignaraja Sundar, I Sherwyn John "TerraFlyer: Aerial and Terrestrial Drone with Optimal path recognition", in IEEE Internet of Things journal (Under Review) **Nov 2024**
- Authored a research paper on developing an ML-integrated ground and aerial vehicle. Contributed to designing the optimal path recognition algorithm, achieving 87% accuracy by analyzing stability, thrust, and aerodynamic factors. Optimized hardware selection reduced resource usage by 15%, ensuring efficient navigation and performance.

SKILLS AND INTERESTS

Programming Languages: Python, C, C++, Java
Full stack Development: HTML, CSS, JavaScript, PHP, React.js, Three.js, Next.js, Flask, Django, REST API, FAST API.
ML Libraries and tools: TensorFlow, PyTorch, Keras, Scikit-learn, Matplotlib, NumPy, Pandas, OpenCV, NLTK, Transformers, Diffusers, LangChain, Dialogflow, MATLAB, XGBoost, LightGBM, spaCy, Perplexity
Other: Github, BitBucket, Apache, Putty, Linux, AWS, Docker
Communication: English(Proficient), Tamil(Native), German(A1 level)
Hardware: Arduino, Node Red, Raspberry pi, ESP32

CERTIFICATIONS

- Data Scientist in R language (Simple Learn) **Jan 2022**
- Transformers and Attention (Coursera) **May 2022**
- Introduction to Natural Language processing (Coursera) **Mar 2023**
- Geo information systems (Udemy) **May 2023**