

Sreevaatsav Bavana

DATA SCIENTIST

Hyderabad, Telangana, India

☎ +91 9177554960 | ✉ bavanasreevaatsav1@gmail.com | 🏠 sreevaatsavb.github.io/portfolio/ | 🌐 SreevaatsavB | in
sreevaatsav-bavana

Summary

Gold medalist in B.Tech Artificial Intelligence at Mahindra University and currently a Data Scientist at Penguin AI. Experience in developing reliable, production-grade ML systems at the intersection of applied machine learning, LLM engineering, and multimodal AI, with a focus on explainability, workflow automation, and measurable business impact.

Education

Mahindra University

Hyderabad, India

BACHELOR OF TECHNOLOGY IN ARTIFICIAL INTELLIGENCE

Aug 2020 - Aug 2024

- Gold Medalist in B.Tech AI, cumulative GPA: 8.76/10.
- Merit scholarship for 4 consecutive years (2020-2024); consistently in top 5% of the class.

Skills

Languages & Libraries

Python, C++, SQL, Pandas, NumPy, Scikit-learn, HuggingFace, PyTorch, TensorFlow, Algorithms, Data Structures

ML & GenAI

Supervised learning, Unsupervised learning, Reinforcement learning, LLMs, VLMs, LLM powered agentic systems

AI Engineering & Infra

LangChain, vLLM, MCP, MLflow, Ray, Quantization, Inference optimization

Cloud & Deployment

AWS, Docker, Kubernetes, FastAPI, Async programming, Linux, GitHub Actions, MongoDB

Experience

Penguin AI

Hyderabad, India

DATA SCIENTIST

Jun 2025 - Present

- Built production-grade LLM workflow systems for prior authorization, reducing medical nurse review time by approximately 30-40% with around 80-85% accuracy; now improving with a human-in-the-loop workflow.
- Driving applied research methodologies to enhance automated medical coding quality.
- Developed an ML + ETL system to predict likely appeal denials and prioritize case review workflows, cutting average appeals processing time by 56%.

Awone AI

Hyderabad, India

ASSOCIATE DATA SCIENTIST

Jun 2024 - Jun 2025

- Improved production LLM inference speed by 15% through quantization and speculative decoding with an n-gram head.
- Built and deployed a BERT-based NER model for healthcare PII redaction using a custom loss function.

Awone AI

Hyderabad, India

DATA SCIENCE INTERN

Jan 2024 - Jun 2024

- Implemented transformer models inspired by Llama-2 and Mistral from scratch to deepen architecture and optimization understanding.
- Experimented with multiple LLM optimization and quantization approaches to reduce latency on development platforms.
- Finetuned task-specific small language models for medical insurance workflows.

Mahindra University

Hyderabad, India

RESEARCH INTERN

Jul 2023 - Sep 2023

- Collaborated in a four-member team to conduct a comprehensive survey on video summarization techniques.
- Co-authored a survey paper published at the ACM Multimedia Workshop (2023).

Projects

Generative AI Fundamentals from Scratch

Self-directed
Apr 2024 - May 2024

- Implemented core transformer and language model modules from scratch in PyTorch.
- Validated intermediate outputs against HuggingFace implementations.
- Code: github.com/SreevaatsavB/Generative_AI_fundamentals.

AI-Assisted Learning for NVIDIA SDKs and Toolkits

ICETCI Hackathon, Mahindra University
Aug 2023 - Sep 2023

- Developed a fine-tuned LLM to help users understand and use NVIDIA SDKs and toolkits.
- Secured 3rd place in ICETCI Hackathon.
- Code: github.com/SreevaatsavB/ICETCI-Hackathon.

Enhancing Video Summarization with a Text-to-Image Module

Mahindra University
Aug 2023 - Dec 2023

- Improved text-to-image projection by adding a learnable projection layer.
- Finetuned baseline projection modules using the QVHighlights dataset.

Publications & Research

GIV-CXR: Densely Grounded, Visually Interpretable, Chest X-Ray Question Answering Dataset

KDD 2026
SUBMITTED

- Authors: Sreevaatsav Bavana, Adit Rushil Potta, Sai Amrit Patnaik, Nidhi Goyal.
- Built a densely grounded chest X-ray QA dataset to support explainable visual reasoning in medical imaging.
- Contributed to baseline multimodal modeling and evaluation design for clinically relevant interpretability.

From Layoff to Reemployment: Analyzing Social Media Narratives

CSoNet 2025
ACCEPTED

- Authors: Yayati Gupta, Sreevaatsav Bavana, Sudhira Chegu, Sujith Sai Kalakonda, Yash Mathur, Manmitha Polsani, Sanatan Sukhija.
- Studied large-scale social media discourse to trace how layoff experiences evolve toward reemployment outcomes.
- Applied NLP-based trend and narrative analysis to identify recurrent themes and temporal shifts.

A Systematic Study on Video Summarization: Approaches, Challenges, and Future Directions

ACM Workshop on User-centric Narrative Summarization of Long Videos
PUBLISHED

- Authors: Kajal Kansal, Nikita Kansal, Sreevaatsav Bavana, Bodla Krishna Vamshi, Nidhi Goyal.
- Co-authored a structured survey comparing classical and deep-learning approaches for long-video summarization.
- Synthesized open challenges in evaluation, narrative coherence, and multimodal alignment to guide future work.

Achievements

2025	Performance Excellence Award , Awarded Performance Appreciation at Penguin AI (2025) as one of the selected contributors for high-impact execution	Hyderabad, India
2020-24	Merit Scholarship (Top 5% of branch) , Mahindra University	Hyderabad, India
2023	3rd Place, ICETCI Hackathon , Mahindra University	Hyderabad, India
2023	Rank 111 (Top 150 of 4000+ teams), Amazon ML Challenge , Amazon	Global