Sushvin Marimuthu *AI/NLP & Full-Stack Engineer*

sushvinmarimuthu@gmail.com +91-6381432504 sushvinmarimuthu.com LinkedIn GitHub

Summary

AI/NLP & Full-Stack Engineer pursuing an MS by Research at IIIT Hyderabad, with hands-on experience in building NLP/LLM models and integrating them into production web applications. Author of three NLP publications with shared-task results. Experience spans linguistic NLP, Healthcare NLP, and EdTech, with full-stack engineering as a consistent thread across projects - from research prototypes to production web applications.

Professional Experience

Rubrix.ai (Full-Stack Developer Intern) **May 2026 – Present**
React.js, FastAPI, MongoDB, Celery, FAISS, LLMs, RAG, Cross-Encoder reranking

- **Developed an AI-powered semantic recommendation system** that matches online resources to course content and created a database of resources with precomputed similarity scores, surfacing matches at $\geqslant 0.85$ similarity. It replaced a paid Perplexity API workflow, saving ~₹10K/month.
- **Implemented code and text plagiarism detection** for student submissions, triggered automatically by a Celery worker after every submission, replacing manual integrity checks.
- **Built an automated question paper generation feature** that lets instructors define templates such as sections, marks per question, units covered, number of sets, and AI-generated vs. previous-year question ratios, eliminating manual assembly and reducing API costs by ~50%.

LTRC, IIIT - Hyderabad (Junior Software Engineer) **Sep 2023 – Jul 2024**
Next.js (React), FastAPI, Trankit

- **Fine-tuned the Trankit framework** on an in-house annotated Tamil data to build an AI-powered Tamil syntactic parser that decomposes complex sentences and maps word relationships for Tamil language research.
- **Designed and built websites** for LTRC events (IASNLP, PLO) and course pages.

Mallow Technologies Pvt. Ltd. (Project Intern) **Feb 2023 – May 2023**
Ruby on Rails

- **Traced and fixed** a backend logic bug in the GIST CRM chat module that caused incorrect customer and message counts in production, restoring data accuracy across chat workflows.

Projects

Structured Radiology Intelligence (SRI) | MS by Research Project **May 2025 – Sep 2026**
Next.js (React), FastAPI, MongoDB, LLMs

- **Improved clinical data standardization for rectal cancer MRI reports** by fine-tuning LLaMA 3.1 8B Instruct for structured information extraction and building a curated domain-specific dataset using multiple LLMs.
- **Developed a web application** for dataset annotation, validation, quality control, and demo, and deployed the fine-tuned LLM as a scalable inference service with guardrails for reliable structured MRI report processing.

MediHelp AI (for CMC Vellore Doctors) | MS by Research Project **Aug 2024 – Dec 2025**
Next.js (React), FastAPI, MongoDB, Tiptap, WebSockets, BullMQ, Sarvam AI, LLaMA, Docker

- **Built a real-time collaborative, multi-format document editor** (Tiptap, WebSockets) and moved heavy operations to asynchronous queues (BullMQ) for better performance and reliability under load, reducing manual document work by ~60%.
- **Integrated LLM-powered** translation, Q&A, summarization, and text-to-table features to support doctors' report writing.
- **Deployed the application using Docker** and set up a CI/CD pipeline with GitHub Actions that deploys to AWS (EC2) for staging and CMC's internal infrastructure for production.

Karpagam e-Governance Portal - Centralized Academic Management System | B.E. Project **Oct 2021 – Feb 2023**
React.js, Django, MySQL, Redis Queue, WebSockets, RBAC

- **Developed an academic management system** for student records, attendance, schedules, and grading, with role-based access control and background jobs (Redis Queue) for asynchronous tasks.
- **Implemented leave modules** for students and faculty with online applications and approvals, synced with attendance such that approved leaves are reflected automatically.

Publications

Structured Radiology Intelligence: Extracting structured data from MRI Reports using LLMs **May 2026**
Third Workshop on Patient-Oriented Language Processing (CL4Health), LREC

- Achieved an 87.36% composite score and 18.42% performance improvement using LLaMA 3.1 8B Instruct for automated MRI report structuring.
- Developed and benchmarked a 55-variable clinical dataset, evaluating LLaMA, Qwen, and DeepSeek through prompt-tuned and fine-tuned LLM pipelines.

SpanSense - Perspective-specific span identification and Summarization **May 2025**
Second Workshop on Patient-Oriented Language Processing (CL4Health), NAACL

- Secured 6th place on the official PerAnsSumm 2025 leaderboard with an overall score of 49% for perspective-aware medical CQA summarization.
- Designed a RoBERTa-large + BART-large pipeline for perspective span classification and summarization, achieving 90% Macro-F1, 92% Weighted-F1, and 84% BERTScore.

Medical Visual Question Answering with Vision-Language Models **Jun 2024**
6th Clinical Natural Language Processing Workshop (Clinical NLP), NAACL

- Secured 6th place in the MEDIQA-M3G 2024 shared task for multilingual, multimodal medical answer generation.
- Fine-tuned a lightweight ViLT Vision-Language Transformer for clinical dermatology VQA and evaluated extended training with Gemini and LLaVA VLMs.

Education

International Institute of Information Technology - Hyderabad **Jul 2024 – Present**
MS by Research in Computer Science & Engineering *Hyderabad, India*

- **Research focus:** Medical NLP
- **Expected graduation:** Dec 2026
- **CGPA:** 8.33

Karpagam Institute of Technology (affiliated with Anna University) **Aug 2019 – Jun 2023**
B.E. in Computer Science & Engineering *Coimbatore, India*

- **CGPA:** 7.95

Skills

- **Programming Languages:** Python, JavaScript, TypeScript, Ruby
- **Frameworks:** React.js, Next.js, Django, FastAPI, Tailwind CSS, Ruby on Rails
- **Databases:** MongoDB, MySQL
- **DevOps & Tools:** Git, GitHub Actions, Docker
- **Cloud Platforms:** AWS (EC2, S3, IAM)
- **AI/LLM:** Transformers, LLMs, RAG, Prompt Engineering, AI CLI Tools (Claude Code, Codex)