

Shivam Sharma

Email: shivam.klt77@gmail.com | GitHub: [theshivam7](https://github.com/theshivam7) | LinkedIn: [theshivam7](https://www.linkedin.com/in/theshivam7) | Scholar | [HuggingFace](https://huggingface.co/theshivam7) | Portfolio: theshivam.online

Education

Indian Institute of Technology Madras
Bachelor of Science, Data Science and Applications

Chennai, India
Sep 2022 – Dec 2026

Research Experience

VideoDB Jan 2026 – Apr 2026
Software Engineer Intern Remote

- Built and open-sourced **VideoDB Claude Skills** [\[link\]](#), enabling natural-language video search, transcription, and processing via MCP integration; distributed to developer marketplace.
- Engineered a real-time **LLM-powered fact-checking pipeline** for live streams (YouTube Live, Google Meet); co-authored a **Gemini VLM benchmarking paper** [\[link\]](#) evaluating multimodal reasoning on video scene understanding.

Nanyang Technological University Singapore Jan 2025 – Dec 2025
Research Intern, Speech and Multilingual NLP Singapore (Remote)

- Designed multilingual **ASR** and speaker diarization pipelines over GigaSpeech 2, Emilia, and NVIDIA Granary; analyzed normalization strategies and benchmark trade-offs across regional accents and speaker demographics.
- Built **Indian-ASR-Bench** [\[link\]](#), a 5-model WER benchmark on Indian English academic speech (**986 TIE clips**); evaluated Whisper Base/Medium/Large, Parakeet-TDT-0.6B, and Qwen3-ASR-1.7B; paper in progress. **Whisper Medium achieved 14.72% corpus WER**, while normalization choices shifted WER by 10–17 pp.

BITS Pilani, Goa Campus Sep 2024 – Dec 2025
Research Intern, NLP and LLM Evaluation Goa, India

- Built **PustakAI** [\[link\]](#), a curriculum-aligned QA pipeline with RAG-based prompting and multi-metric LLM evaluation; fine-tuned **Gemma 3 1B** and **LLaMA 3.2 3B** for educational QA; accepted at **ACM COMPUTE 2025**; released NCERT dataset on HuggingFace.
- Curated **ICH-QA**, a **131K-pair QA dataset** from 7,506 Wikipedia articles on Indian Cultural Heritage; **RAG achieved 48% Exact Match**, validating the dataset requires genuine domain-specific knowledge.

Publications

- PustakAI: Curriculum-Aligned and Interactive Textbooks Using Large Language Models** [\[link\]](#) *ACM COMPUTE 2025*
- Do Thought Streams Matter? Evaluating Reasoning in Gemini Vision-Language Models for Video Scene Understanding** [\[link\]](#) *arXiv 2026*

Projects and Open Source

Video Fact Checker [\[link\]](#) *Electron, React, Hono, Gemini, VideoDB*

- Desktop app that monitors YouTube, Google Meet, and podcasts in real time; captures system audio via VideoDB, extracts claims every 20 seconds, and classifies each with **Gemini** as verified, misleading, or missing context.

BhashaSarthi, Multilingual Translator [\[link\]](#) *React Native, Expo, Gemini 2.5 Flash*

- Open-source mobile app covering all **22 official Indian languages**; voice input with auto-transcription, text-to-speech output, and on-device translation history via AsyncStorage.

Open Source: Open WebUI [\[link\]](#) (90k+ stars), Windows path compatibility fix in audio module, **PR merged**; **OpenAI Codex** [\[link\]](#), CLI UX contribution for skill display in terminal interface, accepted by maintainers.

Technical Skills

- ML / NLP:** PyTorch, Hugging Face Transformers, LLMs (Gemma, LLaMA, Gemini), fine-tuning (LoRA/PEFT), RAG, Whisper ASR, speaker diarization, prompt engineering, LangChain, evaluation (BLEU, ROUGE, BERTScore, RAGAS, LLM-as-judge)
- Languages:** Python, TypeScript, JavaScript, SQL, Java
- Frameworks and Tools:** React Native, Flask, REST APIs, MongoDB, Git, GitHub, Jupyter Notebook

Honors and Awards

Research Presenter, **GLEX 2025**, New Delhi | Finalist, Smart India Hackathon 2023 and Karnataka State Police Hackathon 2024 | Scholarships: IIT Madras B.S. Degree; Google PM Scholarship