

RACHAKONDA HRITHIK SAGAR

☎ +91 7893496639 • 📧 hrithiksagar • 🌐 hrithiksagar • 🎓 Google Scholar • ✉ hrithiksagar.rachakonda@gmail.com

Education

International Institute of Information Technology, Hyderabad

July 2025 – Present

MS by Research in Computer Science

– Focus: Vision-Language Models, Multimodal Document Understanding, Indic AI.

Sharda University, UP

Aug 2018 – May 2022

B.Tech in CSE (AI & ML)

GPA: 8.40/10

Experience

BharatGen

Jan 2025 – Present

AI Research Engineer

IIIT Hyderabad

- **Training & Inference at Scale:** Configured MS Swift from **ZeRO-1** → **ZeRO-3**, achieving **>90% GPU memory utilization** across **64 H100 GPUs**; optimized vLLM offline serving reducing QA generation latency **6×** (2 mins → 20 secs) and DeepSeek-R1 641B (**multi-node**) inference **5.6×** (2.5 mins → 27 secs) by tuning sampling parameters.
- **Data Pipeline at Scale:** Designed a three-stage synthetic data engine generating **10M QA pairs** across **400K documents** (layout-aware HTML, curved-text segmentation, chart rendering interception); multi-stage quality filtering via heuristic, **BERT-similarity**, and **LLM-as-Judge** at **200K-document scale**.
- **Patram-7B** (India's first foundational document VLM, publicly released on HuggingFace & IndiaAI AIKosh): Core contributor to full stack - data, training, and evaluation. Achieves **0.855 G-Eval on DocVQA**, outperforming **Gemma-3-12B-it**, **InternVL3-9B** and **DeepSeek-VL2** at the same parameter scale.
- **Patram-Grounder:** Led data engine and all benchmark evaluations for **mask-based pixel-level segmentation grounding**, surpassing DOGR by **6.9 F1 points** and setting new **state-of-the-art on 4 grounded DocVQA benchmarks**; extended to **grounded Chain-of-Thought** architecture with curated training data (**325K docs**, **1.5M grounded reasoning pairs**).

Centre for Visual Information Technology (CVIT), IIIT Hyderabad

May 2024 – Dec 2024

Research Assistant

- Designed and trained custom **multimodal architectures** integrating VLMs with **promptable segmentation modules** (Hi-SAM, Surya-OCR, DocTR) for fine-grained **Indic document understanding**.
- Built rigorous benchmarking pipelines comparing **10+** open and closed-source models (Claude 3.7 Sonnet, DeepSeek-VL2, Qwen2.5-VL, LLaMA) using heuristic metrics and GEval; established evaluation standards adopted across the project.
- Developed internal tooling: grounding dataset viewer, PDF raw dataset parser, and **data annotation tools** enabling in-depth **EDA** and error analysis across **complex document layouts**.

Stealth Startup

Mar 2022 – Aug 2024

Machine Learning Engineer

Hyderabad (part-time)

- Designed and deployed **production ML pipelines** for **text classification** and **information extraction**; co-invented a **patent-pending IoT device** (granted 2022).

Publications & Patents

M3Grounder: Mask-Based Multi-Span and Multi-Granular Grounding for Document QA

CVPR 2026 Main Conference | Paper

DoCoG: Mask-based Multi-Type Grounded Chain-of-Thought for Document QA

ECCV 2026 Main Conference

AI-Enabled Healthcare Chatbot for Patient Assistance

Applications of AI & ML, Springer, 2020

Women Assault Detection Using Pitch | Patent 202111040354: Oral Healthcare Management Device

Springer, AIR 2022; Patent 2021–22

Technical Skills

Languages & Frameworks: Python, SQL, PyTorch, TensorFlow, OpenCV, PEFT, vLLM, HuggingFace, DeepSpeed, MS Swift
AI / ML: VLMs, Multimodal Learning, Document AI, Synthetic Data Generation, Instruction Tuning, RLHF, LLM Evaluation, AI Agents & Tool Use, RAG
Infrastructure & Tools: SLURM, Docker, WandB, Git, Claude, GPT-4o, Gemini, Qwen, DeepSeek

Achievements

Patram-7B | Core contributor to India's first foundational document VLM

Publicly released on HuggingFace & IndiaAI AIKosh)

Selected merged PRs in official OpenAI (gpt-oss) and Alibaba (Qwen3-VL) repositories

vLLM offline serving fixes

~31 citations (h-index: 3) | Google Scholar