Ritesh Thawkar

+91 9552361592 | riteshthawkar2003@gmail.com

Ritesh Thawkar | Ritesh Thawkar | Google Scholar | Hugging Face

Professional Summary

Multimodal AI researcher developing self-evolving vision-language models and unified systems for understanding and generation. Research spans multimodal reasoning and analysis of model capabilities across perception, cognition, grounding, and long-horizon reasoning. Work also includes agentic RAG, retrieval pipelines, and applied LLM systems that connect research advances with real-world applications.

Education

- **Mohamed Bin Zayed University of Artificial Intelligence** *Aug 2024 - Present*
  Master of Science in Computer Vision Abu Dhabi, UAE
  - **GPA:** 3.60/4.00
  - **Relevant Coursework:** Machine Learning with Python, Introduction to Deep Learning, Human and Computer Vision, Probabilistic and Statistical Inference, Mathematical Foundations of Artificial Intelligence
- **Dr. Babasaheb Ambedkar Technological University** *Feb 2021 - May 2024*
  Bachelor of Technology in Computer Science and Engineering Lonere, India
  - **GPA:** 8.51/10.00
  - **Relevant Coursework:** Data Structures, Design & Analysis of Algorithms, Database Systems, Machine Learning, Artificial Intelligence, Cloud Computing

Publications

C = Conference, S = Preprint

[C1] Shravan Venkatraman, **Ritesh Thawkar**, Omkar Thawakar, Rao Muhammad Anwer, Hisham Cholakkal, Salman Khan, Fahad Shahbaz Khan. (2026). **Paying More Attention to Visual Tokens in Self-Evolving Large Multimodal Models.** In *ECCV 2026*.

[C2] Wafa Al Ghallabi, **Ritesh Thawkar**, Sara Ghaboura, Omkar Thawakar, Numan Saeed, Dana Al Nuaimi, Ajnas Alkatheeri, Salman Khan, Fahad Shahbaz Khan. (2026). **How Good are Foundation Models in Longitudinal MRI Disease Progression Reasoning?** In *MICCAI 2026*. **Spotlight presentation**.

[C3] Shravan Venkatraman, Omkar Thawakar, **Ritesh Thawkar**, Abdelrahman Shaker, Rao Muhammad Anwer. (2026). **Perception Before Supervision: Self-Contained Visual Distillation from Counterfactual Blind Spots.** In *BMVC 2026*.

[S1] **Ritesh Thawkar**, Shravan Venkatraman, Omkar Thawakar, Abdelrahman Shaker, Fahad Khan, Hisham Cholakkal, Salman Khan, Rao Muhammad Anwer. (2026). **Ask, Solve, Generate: Self-Evolving Unified Multimodal Understanding and Generation via Self-Consistency Rewards.** *arXiv preprint*.

[C4] Wafa Al Ghallabi, Muhammad Zaigham Zaheer, **Ritesh Thawkar**, Omkar Thawakar, Salman Khan, Fahad Shahbaz Khan. (2026). **A Multi-Agent Diffusion Approach for MRI Anomaly Segmentation via Modality-Specific LoRA Specialization.** In *WACV 2026*.

[S2] Abdelrahman Shaker, Ahmed Heakl, Jaseel Muhammad, **Ritesh Thawkar**, Omkar Thawakar, Senmao Li, Hisham Cholakkal, Ian Reid, Eric P. Xing, Salman Khan, Fahad Shahbaz Khan. (2026). **Mobile-O: Unified Multimodal Understanding and Generation on Mobile Device.** *arXiv preprint*.

[S3] Omkar Thawakar, Shravan Venkatraman, **Ritesh Thawkar**, Abdelrahman Shaker, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, Fahad Khan. (2025). **EvoLMM: Self-Evolving Large Multimodal Models with Continuous Rewards.** *arXiv preprint*.

[S4] Dinura Dissanayake, Ahmed Heakl, Omkar Thawakar, Noor Ahsan, **Ritesh Thawkar**, Ketan More, Jean Lahoud, Rao Anwer, Hisham Cholakkal, Ivan Laptev, Fahad Shahbaz Khan, Salman Khan. (2025). **How Good are Foundation Models in Step-by-Step Embodied Reasoning?** *arXiv preprint*.

[C5] Omkar Thawakar, Dmitry Demidov, **Ritesh Thawkar**, Rao Muhammad Anwer, Mubarak Shah, Fahad Shahbaz Khan, Salman Khan. (2025). **Beyond Simple Edits: Composed Video Retrieval with Dense Modifications.** In *ICCV 2025*.

[C6] Wafa Al Ghallabi, **Ritesh Thawkar**, Sara Ghaboura, Ketan Pravin More, Omkar Thawakar, Hisham Cholakkal, Salman Khan, Rao Muhammad Anwer. (2025). **Fann or Flop: A Multigenre, Multiera Benchmark for Arabic Poetry Understanding in LLMs.** In *EMNLP 2025*.

[S5] Ayesha Ishaq, Jean Lahoud, Ketan More, Omkar Thawakar, **Ritesh Thawkar**, Dinura Dissanayake, Noor Ahsan, Yuhao Li, Fahad Shahbaz Khan, Hisham Cholakkal, Ivan Laptev, Rao Muhammad Anwer, Salman Khan. (2025). **DriveLMM-o1: A Step-by-Step Reasoning Dataset and Large Multimodal Model for Driving Scenario Understanding.** *arXiv preprint*.

[C7] Sara Ghaboura, Ketan Pravin More, **Ritesh Thawkar**, Wafa Al Ghallabi, Omkar Thawakar, Fahad Shahbaz Khan, Hisham Cholakkal, Salman Khan, Rao Muhammad Anwer. (2025). **Time Travel: A Comprehensive Benchmark to Evaluate LMMs on Historical and Cultural Artifacts.** In *Findings of ACL 2025*.

[C8] Omkar Thawakar, Dinura Dissanayake, Ketan Pravin More, **Ritesh Thawkar**, et al. (2025). **LlamaV-o1: Rethinking Step-by-step Visual Reasoning in LLMs.** In *Findings of ACL 2025*.

Experience

- **Microsoft Research India** *Jun 2026 - Jul 2026*
  Research Intern Bengaluru, India
  - Conducted a systematic evaluation of multimodal models through MS-VISTA, integrating two visual benchmarks and one reasoning-analysis study across perception, cognition, and spatial intelligence.
  - Developed reproducible VLM/MLLM evaluation pipelines covering model inference, answer extraction, experiment tracking, and comparative performance analysis.
  - Built a Flask and React research leaderboard with authenticated submissions, benchmark scoring, capability-level analysis, and auditable evaluation reports.
- **Mohamed Bin Zayed University of Artificial Intelligence** *Sep 2024 - Present*
  Research Assistant Abu Dhabi, UAE
  - Engineered visual-reasoning benchmarks for VLMs/MLLMs, evaluating sequential logic, grounding, and multimodal reasoning across research tasks.
  - Developed composed-video-retrieval methods using dense video descriptions and multimodal embeddings for fine-grained retrieval and temporal understanding.
  - Deployed document-retrieval pipelines and scalable LLM web agents for Lawa.AI using FastAPI, vector databases, and retrieval-augmented generation.
  - Optimized Django and PostgreSQL AI infrastructure for Nutrigenics.care, integrating image-aware LLM recommendations into personalized nutrition workflows.
- **Quantiphi Analytics** *Feb 2024 - May 2024*
  Framework Engineer Intern Remote
  - Engineered data-ingestion pipelines for small and medium business operations, supporting reliable multi-source operational tracking.
  - Automated business-insight extraction with ML-driven analytics workflows, converting operational data into decision-ready outputs.
  - Optimized enterprise data-processing workflows, reducing end-to-end processing time by 20%.

Selected Projects

- **MBZUAI Web Agent Suite:** *Tools: [Agentic RAG, Web Agents, LLM Orchestration, FastAPI]*
  - Built an LLM-driven agentic pipeline that resolves MBZUAI website queries using retrieval, tool-use, and response verification.
  - Deployed the system in production, serving thousands of users and iterating prompts, tools, and routing based on real traffic.
  - Developed an interaction analytics dashboard to track user intents, content gaps, and unresolved queries for continuous Agentic RAG tuning.

- **Lawa Agent Studio:** *Tools: [LLM, RAG, Workflow Orchestration, React, FastAPI]*
  - Built a platform for creating customizable LLM-powered agentic chatbots with configurable tools, tone, and knowledge sources.
  - Implemented ingestion and indexing pipelines for website and document sources to enable fast, accurate LLM responses.
  - Added analytics and evaluation hooks to measure resolution rates, user satisfaction, and retrieval quality for LLM improvements.
- **Lawa.ai:** [🌐] *Tools: [LLM, RAG, Workflow Orchestration, FastAPI, React]*
  - Designed and developed LLM-powered web assistants to improve user interaction and engagement for websites.
  - Implemented a document retrieval system using vector databases to ground LLM responses and improve query resolution accuracy.
  - Developed backend workflows and API integration using FastAPI and websockets for scalable deployment.
- **Nutrigenics.care:** [🌐] *Tools: [LLM, Django, PostgreSQL, Recommendation System]*
  - Designed and developed the platform's web infrastructure to manage user and recipe data efficiently, enhancing scalability and usability.
  - Integrated NutritionGPT to power LLM-based recipe recommendations from image inputs and user preferences.
  - Implemented a personalized recommendation system leveraging LLMs to tailor nutrition and recipe suggestions to individual needs.
  - Optimized backend workflows using Django and PostgreSQL to ensure efficient data handling and faster response times.

Skills

- **AI Research & Evaluation:** VLMs / MLLMs, Multimodal Benchmarking, Benchmark Design, Capability Analysis, Error Analysis
- **LLM Systems:** Agentic RAG, Tool-Using Agents, Retrieval, Vector Databases
- **Languages:** Python, TypeScript, JavaScript, C
- **ML / Deep Learning:** PyTorch, Transformers, TensorFlow, Scikit-learn
- **Research Engineering:** Evaluation Pipelines, Experiment Tracking, LaTeX, Git, Docker
- **Web & Data:** FastAPI, Flask, Django, React, PostgreSQL, SQLite
- **Cloud:** AWS, Azure, GitHub Actions

Research Service & Languages

Research Service: Reviewer, ACCV 2026
Languages: English (Intermediate), Hindi (Advanced), Marathi (Advanced)