Avinash Anand, PhD

avinasha@iiitd.ac.in avinanand
Google Scholar

Summary

Research Interests: Natural language processing, large language models (pre-training, alignment, and multilingual systems), multimodal LLMs for accident detection and medical imaging.
As my current focus is on building LLM pre-training pipelines—from kernel-level optimization and ultrascale GPU parallelism to LLM algorithmic improvements like FlashAttention-4 and post-training methods such as GRPO—my research centers on advancing Large Language Models and multimodal generative AI. The aim is to design robust and efficient systems that integrate text and vision for generalized reasoning and real-world applicability. My work spans LLM architecture, optimization, and task-specific model development, emphasizing multilingual, cross-lingual, and mathematical reasoning capabilities. I also explore multimodal mental health detection and vision-based applications, such as road accident detection, traffic analysis, and maritime tracking. In general, my goal is to build scalable and impactful AI systems that advance healthcare, transportation, and software engineering while meaningfully improving quality of life.

Education

2021 - 2024 **PhD Research Scholar, IIIT Delhi, India.**
Thesis title: *Multimodal Systems For Scientific and Educational Applications*,
Courses: Large Language Models(LLMs), Natural Language Processing, Knowledge Graphs, Collaborative Filtering, Data structures and algorithms, Information Retrieval

2010 – 2014 **Bachelor of Technology, IIIT Jabalpur** in Computer Science and Engineering

Employment History

Dec-2024 - present **Postdoctoral Researcher, Singapore Institute of Technology (SIT)** My research focuses on advancing Large Language Models (LLMs) by developing generative models that integrate text, audio, and visual data to create optimized, task-specific AI systems. I work on improving LLM architecture, training, and reasoning capabilities, with applications in healthcare, software engineering, and AI problem solving. My goal is to push the boundaries of LLMs, enhancing their scalability and real-world impact.

Jan-2024 - June-2024 **National University of Singapore, Research Intern**,
Worked in Prof. Roger ZIMMERMANN's Lab.
Worked and Publsihed 4 Papers on LLMs and Multimodal LLMs.

2021 - 2024 **Global Standards India, PhD Fellow/Lead AI Scientist**,
Worked on OCR pipeline for 1.2 million Indian retail products. Responsible for Leading the entire AI team and building Intelligent systems for the company.

Employment History (continued)

2022 - 2023 **National Institute of Informatics, Tokyo, Japan**,
Worked in Prof. Shin'ichi Satoh's Lab.
Worked on OCR pipeline for 1.2 million Indian retail products and built a product for scientific and financial document analysis and summarization, also worked on improving scientific text generation using deep learning techniques.

2018 - 2021 **Lead Research Engineer, MIDAS Labs, IIIT-Delhi,**
For a Japanese MNC, I built and deployed grammar error correction models using Transformers. These models improved correction speed by 10 times compared to conventional Transformers, enabling real-time corrections.

2016 - 2017 **Quittung Labs Pvt. Ltd.(BillFree), Noida** Co-Founder & Data Scientist
I created a system for personalized product recommendations, inventory forecasting, and customer satisfaction analysis. I also developed an end-to-end bill management system, including backend development for the "BillFree" app on Android and iOS.

2014 - 2016 **Appcon Solutions, Jabalpur** Co-founder & Developer
I automated "e-challan" generation for the Traffic Police Department, developed pricing and categorization models for online rentals, created prediction models for online publishers, and built embedded RFID products for the State Government of Madhya Pradesh.

Current Projects, 2025

NutriAI: Multimodal Transformer System for Personalized Nutrition and Food Intake Estimation
Designed *NutriAI*, a multimodal AI system combining LLMs and food image analysis to deliver personalized nutrition recommendations based on user health data and meal intake. The system leverages transformer-based models to assess individual health conditions and generate tailored dietary plans, estimates nutritional intake from RGB-D meal images using deep learning–based mass estimation, and offers real-time coaching with localized food suggestions in Singapore. Contributed to research on personalized health modeling and AI-driven nutrition tracking.
Industrial Application: NutriAI holds strong potential for deployment in smart hospitals and health insurance platforms enabling scalable, personalized nutrition interventions to reduce chronic disease risks.

Spatio-Temporal Vision-Language Model for Causal Safety Reasoning in Industrial Environments
Developed a real-time Vision-Language Model (VLM) leveraging temporal video transformers for causal action-level reasoning on warehouse safety compliance. Integrated multi-modal alignment of video and language data to detect violations, reconstruct incidents, and trigger explainable alerts using existing CCTV infrastructure. Advanced research in spatio-temporal grounding and agentic capabilities for industrial safety automation.
Industrial Application: This system can be deployed in manufacturing plants, and smart warehouses to automate safety audits, and prevent workplace incidents through AI-driven monitoring.

Current Projects, 2025 (continued)

- **Advancing Multilingual Mathematical Reasoning for Open-Source LLMs via Hierarchical Curriculum Learning, EACL'26(submitted)**
  Proposed a two-axis curriculum learning framework combining difficulty-stratified training and reverse reinforcement learning–based step-wise reasoning to enhance multilingual mathematical problem-solving in LLMs. Fine-tuned open-source models like Qwen2.5-Math using bilingual instruction tuning across English, Hindi, and Marathi, enabling robust cross-lingual transfer of mathematical reasoning capabilities. Designed evaluation protocols to assess symbolic correctness, step alignment, and reasoning faithfulness in low-resource language contexts.
  Industrial application: Enables deployment of multilingual math tutoring agents and assessment tools for regional education systems and edtech platforms.

- **Cross-Lingual Code Repair & Generation via Reasoning-Driven Curriculum Learning, ICML-26(to be submitted)**
  Developed a multilingual pipeline for bug localization and code repair in Python, Java, and C++ using a reasoning-aware curriculum learning framework. Incorporated natural language explanations for each code transformation step to enhance model transparency and user trust. Introduced bilingual combined training to improve cross-lingual generalization in code generation and created a task-specific evaluation framework measuring functional correctness, reasoning quality, and multilingual robustness.
  Industrial application: Lays the foundation for intelligent coding assistants and automated test-case generation tools, enhancing code debugging and validation in enterprise software engineering.

- **Optimized Stereo Depth Estimation for Real-Time, Long-Range Accuracy in Autonomous Systems**
  This paper introduces a novel stereo depth estimation pipeline designed for real-time, long-range accuracy in autonomous systems. By combining optimized feature matching, robust synchronization, and temporal smoothing, the system achieves <3% error at 50m and <10% error at 100m, outperforming conventional monocular methods. Operating at 10 FPS on CPU, it provides significant improvements in processing speed and accuracy, making it suitable for applications in autonomous systems, robotics, and 3D reconstruction.
  Industrial application: Enables low-cost depth perception for autonomous vehicles, drones, and industrial robots operating in real-world environments.

- **Multimodal LLMs for Scalable and Explainable Mental Health Support**
  Developed a multimodal architecture leveraging large language models (LLMs) to enable accurate detection, monitoring, and diagnosis of mental health conditions like depression. The system integrates speech, vision, text, and behavioural signals to detect fine-grained depression levels and delivers interpretable outputs for clinical decision support. Emphasis was placed on real-time, scalable mental health monitoring and personalised feedback. Ongoing work focuses on explainability and CBT-driven conversational support and clinical usability in mind, aiding both users and therapists.
  Industrial application: Enables AI-driven mental health assistants that offer continuous, multi-modal monitoring, early risk detection, and explainable clinical insights for integration into clinical screening in Singapore.

Current Projects, 2025 (continued)

- **FinAgent-plan is to design 6 Agents, which are individually performing specific tasks mentioned below-Top-5 IN SEBI Hackathon-25**
  Agent 1 - Financial Data Scrapper this would be scrapping Financial Data such as Balance Sheets, Cash Flow, Peer Comparison Table, Profit Loss Statement which are numbers, which would be further store in context file.
  Agent 2 - Financial Document Analysis Agent its purpose is to extract, interpret, and summarize critical information from unstructured financial documents, which are conceal reports, annual reports, investor presentations, stock exchange fillings, etc.
  Agent 3 - Sector Specific Digital Twin, it's purpose is to identifying real-time market opportunities and providing sector-specific insights, as it will be trained on Sector Specific Expert like Banking and Finance for Warren Buffet, It will try to replicate his strategy as per real time market scenario.
  Agent 4 - Broker Integration and Strategy Agent, it will be integrated with user's trading account and will study the patters of his/her style of investing and portfolio allocation, and then will suggest portfolio restructuring strategies.
  Agent 5 - Deep Web Research and Sentiment Analysis Agent, experts believe market works on 2 things. One is numbers which are Financial data and other is sentiment. Deep Web Research agent will use it's Deep Research Models to Fetch Real time News and will feed to the LLM's context window to Judge the impact of such events based on Past Scenarios.
  Agent 6 - Fiduciary Agent and Trust Layer it will act as an ethical compass for the entire Financial LLM system, ensuring its operations and recommendations prioritize the client's best interests and promote trustworthiness.
  Industrial application: This system of agents enables automated financial data collection, analysis, and real-time strategy recommendations, ensuring ethical oversight and personalized portfolio management. It offers wealth management firms and financial institutions advanced tools for market insights, sentiment analysis, and regulatory compliance.

Recently Accepted Papers, 2024-25

Avinash Anand, ..,Rajiv Ratn Shah, *"BRI-MH: Behavioral Risk Index for Mental Health — An interpretable multimodal LLM-augmented framework", Accepted at AAAI-26(SA).*

Avinash Anand, ..,Rajiv Ratn Shah, *"When Equal Isn't Fair: Mitigating Over-Normalization in Large Language Models", Accepted at AAAI-26(SA).*

Avinash Anand, ..,Rajiv Ratn Shah, *"IMPACT: Integrated Multimodal Pipeline for Rapid Accident Causality Tracking", Accepted at AAAI-26(SA).*

Avinash Anand, ..,Rajiv Ratn Shah, *"Enhancing Scientific Visual Question Answering via Vision-Caption aware Supervised Fine-Tuning", Accepted at ACMMM-LAVA-25.*

Avinash Anand, Avni Mittal,.., Rajiv Ratn Shah, Roger Zimmermann, and Shin'ichi Satoh, *"Unveiling Learner Dynamics: The ECLIPSE Dataset and NeuralGaze Framework for Prolonged Engagement Assessment in Online Learning", Accepted at ECAI-24.*

Avinash Anand, Naman Lal,.., Rajiv Ratn Shah, *"Advances in Citation Text Generation: Leveraging Multi-Source Seq2Seq Models and Large Language Models", Accepted at CIKM-24.*

Avinash Anand, Raj Jaisawal, Rajiv Ratn Shah, and Roger Zimmermann, *"Avinash Anand, Raj Jaisawal, Rajiv Ratn Shah, "Advancing Multimodal LLMs: A Focus on Geometry Problem Solving, Reasoning, and Sequential Scoring". ACM Multimedia Asia 2024 .*

Avinash Anand, Raj Jaisawal,.., Rajiv Ratn Shah, and Roger Zimmermann, *"Enhancing Geometric Problem-Solving using GeoVQA: A Multimodal Geometry Dataset for Advanced Vision Language Model", Accepted at MIPR-24.*

Avinash Anand, Avni Mittal,.. , Rajiv Ratn Shah, Roger Zimmermann, and Shin'ichi Satoh, *"ExCEDA: Unlocking Attention Paradigms in Extended Duration E-classrooms by leveraging Attention-mechanism models", Accepted at MIPR-24.*

Research Publications

Conference Proceedings

1. A. Anand, K. Prasad, C. Kirtani, *et al.*, "Multilingual mathematical reasoning: Advancing open-source llms in hindi and english," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 23 415–23 423. DOI: 10.1609/aaai.v39i22.34509.
2. A. Anand, Janak, Apoorv, A. Verma, and R. R. Shah, "Mm-phyqa: Multimodal physics question-answering with multi-image cot prompting," in *PAKDD-24*, Taiwan: PAKDD, 2024. DOI: https://doi.org/10.1007/978-981-97-2262-4_5.
3. A. Anand, K. Addala, K. Baghel, *et al.*, "Revolutionizing high school physics education: A novel dataset," in *2023 Conference on Big Data and Artificial Intelligence*, New Delhi, India: IEEE Computer Society, 2023. DOI: 10.1007/978-3-031-49601-1_5.
4. A. Anand, A. Goel, M. Hira, *et al.*, "Sciphyrag - retrieval augmentation to improve llms on physics q&a," in *2023 Conference on Big Data and Artificial Intelligence*, New Delhi, India: IEEE Computer Society, 2023. DOI: 10.1007/978-3-031-49601-1_4.
5. A. Anand, M. Gupta, K. Prasad, *et al.*, "Kg-ctg: Citation generation through knowledge graph-guided large language models," in *2023 Conference on Big Data and Artificial Intelligence*, New Delhi, India: IEEE Computer Society, 2023. DOI: 10.1007/978-3-031-49601-1_3.
6. A. Anand, M. Gupta, K. Prasad, *et al.*, "Mathify: Evaluating large language models on mathematical problem-solvingtasks.," in *GAIED, 2023*, New Orleans: Neural Information Processing Systems, 2023.
7. A. Anand, A. Jairath, N. Lal, *et al.*, "Gec-dcl: Grammatical error correction model with dynamic context learning for paragraphs & scholarly papers," in *2023 Conference on Big Data and Artificial Intelligence*, New Delhi, India: IEEE Computer Society, 2023. DOI: 10.1007/978-3-031-49601-1_7.
8. A. Anand, R. Jaiswal, P. Bhuyan, *et al.*, "Tc-ocr: Tablecraft ocr for efficient detection & recognition of table structure & content," in *Proceedings of the 1st International Workshop on Deep Multimodal Learning for Information Retrieval*, ser. MMIR '23, Ottawa ON, Canada: Association for Computing Machinery, 2023, pp. 11–18, ISBN: 9798400702716. DOI: 10.1145/3606040.3617444.
9. A. Anand, R. Jaiswal, M. Gupta, *et al.*, "Ranlaynet: A dataset for document layout detection used for domain adaptation and generalization," in *ACM Multimedia Asia 2023 (MMAsia '23)*, Tainan, Taiwan: Association for Computing Machinery, 2023.
10. A. Anand, K. Prasad, U. Goel, *et al.*, "Context-enhanced language models for generating multi-paper citations," in *2023 Conference on Big Data and Artificial Intelligence*, New Delhi, India: IEEE Computer Society, 2023. DOI: 10.1007/978-3-031-49601-1_6.
11. A. Goel, M. Hira, A. Anand, S. Bangar, and D. R. R. Shah, "Advancements in scientific controllable text generation methods," 2023. arXiv: 2307.05538 [cs.CL].

Skills

Frameworks — Artificial Intelligence, Advanced Natural Language Processing, Advanced Computer vision, Data Preprocessing and Feature Engineering, Model Development, Training and Evaluation, OCR, AI Frameworks and libraries

Languages — English, Hindi.

Awards and Achievements

- **Best paper Award-2024 IEEE International Joint Conference on Biometrics (IJCB)**
- **Kaggle Competitions Expert** Rank in the top 2% Worldwide among 175K Active Data Scientists

Awards and Achievements (continued)

- **IIIT-Delhi Dean IRD** Research Excellence Award for my PhD research- 2023

References

Dr. Rajiv Ratn Shah, Associate/Chair Professor, IIIT Delhi.

Dr. Yaman Kumar Singla, Adobe, IIIT Delhi

Dr. Shin'ichi Satoh, Professor, NII, Tokyo, Japan.