Avni Mittal

+91-9667730877 | avni.mittal2002@gmail.com | Google Scholar | avni-mittal-5a0b3b198 | AvniMittal13

Education

- **Indian Institute of Technology Mandi** *Nov. 2020 – May 2024*
  Bachelor of Technology in Computer Science and Engineering - with Honors in Management Himachal Pradesh, India
  - CGPA: 8.93/10

- **Technische Universität Darmstadt** *Oct. 2022 – Mar. 2023*
  Semester Exchange (5th Semester), B.Tech CSE Darmstadt, Germany
  - CGPA: 9.33/10

Research Interests

Multi-agent systems, AI alignment, interpretability, and reasoning in large language models.

Patents and Publications

C=Conference, P=Patent

[C.1] ***Avni Mittal***, Arka Mukhopadhyay, Arnav Kulkarni, Anhad Thukral, Ayuj Aryan, Arnav Bhavsar. "Grapher-NCA: Integrating Global Context via Graph-Based Updates in Neural Cellular Automata." [In Preparation]

[C.2] ***Avni Mittal***, Shanu Kumar, Sandipan Dandapat, Monojit Choudhary. "LITMUS++: An Agentic System for Predictive Analysis of Low-Resource Languages Across Tasks and Models." *IJCNLP–AACL 2025, System Demonstration Track*. [To Appear]

[C.3] ***Avni Mittal***, Sree Hari Nagaralu, Sandipan Dandapat. "PROTECT: Policy-Related Organizational Value Taxonomy for Ethical Compliance and Trust." *SICon 2025 (ACL Workshop)*. [Link]

[C.4] ***Avni Mittal***, John Kalkhof, Anirban Mukhopadhyay, Arnav Bhavsar. "MedSegDiffNCA: Diffusion Models with Neural Cellular Automata for Skin Lesion Segmentation." *IEEE CBMS 2025*. [Link]

[C.5] Avinash Anand, ***Avni Mittal***, Laavanaya Dhawan, Mahisha Ramesh, Juhi Krishnamurthy, Naman Lal, Raj Jaiswal, Pijush Bhuyan, Himani, Astha Verma, Rajiv Ratn Shah, Roger Zimmermann, Shin'ichi Satoh. "Unveiling Learner Dynamics: The ECLIPSE Dataset and NeuralGaze Framework for Prolonged Engagement Assessment in Online Learning." *ECAI 2024*. [Link]

[C.6] Avinash Anand, ***Avni Mittal***, Laavanaya Dhawan, Juhi Krishnamurthy, Mahisha Ramesh, Naman Lal, Astha Verma, Pijush Bhuyan, Himani, Rajiv Ratn Shah, Roger Zimmermann, Shin'ichi Satoh. "Exceda: Unlocking Attention Paradigms in Extended Duration E-Classrooms with Attention Mechanism." *IEEE MIPR 2024*. [Link]

[P.1] ***Avni Mittal***, Shanu Kumar, Sandipan Dandapat, Monojit Choudhary. "Graph-Based Multi-Agent Predictive Analysis System for Trustworthy Language Model Evaluation." *Accepted for filing (patent being filed).*

[P.2] Gautam Prasad, ***Avni Mittal***, Yugal Sachdev. "A System of Simultaneous Optimization of Multiple Agents." *Accepted for filing (patent being filed).*

[P.3] Gautam Prasad, ***Avni Mittal***, Yugal Sachdev. "A System to Navigate Product Complexities through Automatic Visual Guide Generation." *Submitted for review for filing.*

Research Experience

- **Multi-Agent System for Predictive Analysis of Low-Resource Languages**
  Supervisors: Dr. Sandipan Dandapat and Dr. Monojit Choudhury July 2025 – Present
  - Designed a multi-agent system based on directed acyclic graph (DAG) guided by an expert knowledge graph to predict model performance for the task language without fine-tuning or full benchmarking; agents retrieve multilingual evidence, extract relevant features and train lightweight regressors for prediction.
  - Created benchmark scenarios across six tasks and five multilingual settings to evaluate predictive accuracy and reasoning quality under diverse low-resource conditions.
  - Work accepted at **IJCNLP-AACL 2025 (System Demonstration Track)**; patent filing in progress [**C.2**, **P.1**].

- **Faithfulness and Deceptive Reasoning in Mixture-of-Experts Models**
  Supervised Research at SPAR AI (Safety Alignment Research) Sept. 2025 – Present
  - Studying faithfulness and deceptive reasoning in Mixture-of-Experts (MoE) models by analyzing expert-level token routing and evaluating how disabling high-deception experts affects reasoning faithfulness and task accuracy.
  - Developing a unified faithfulness metric combining correctness and CoT reasoning alignment with token level importance based decomposition, and extending it through **Thought Anchors** for cross-domain reasoning evaluation.

- **Medical Image Segmentation using Neural Cellular Automata and Diffusion**
  Supervisors: Dr. Arnav Bhavsar and Dr. Anirban Mukhopadhyay Aug. 2023 – Present
  - Conducted bachelor's thesis research on enhancing the **M3D-NCA** framework for medical image segmentation through region-of-interest detection and Vision Transformer–based Cellular Automata (ViTCA), achieving ~40% faster inference with comparable accuracy.
  - Proposed the integration of diffusion models with Neural Cellular Automata for improved segmentation; published as **MedSegDiffNCA** at **IEEE CBMS 2025 [C.4]**.
  - Currently extending this work through **Grapher-NCA**, incorporating graph-based global reasoning into NCA architectures for improved segmentation accuracy and convergence (*see [C.1]*).
- **Multimodal Engagement Prediction in Classroom Environments**
  Supervisor: Dr. Rajiv Ratn Shah, MIDAS Lab IIIT Delhi, Research Intern Sept. 2023 – Aug. 2024
  - Developed novel pipeline for multimodal engagement prediction in classroom environments using **Neural Cellular Automata (NCA)** based framework integrating local-global attention and behavioral facial features, achieving improved efficiency and performance over existing baselines.
  - Curated large-scale multimodal datasets (**ExCEDA**, **ECLIPSE**) for prolonged engagement and behavioral analysis.
  - Introduced a **Context-Guided Vision Transformer (ViT)** approach for modeling temporal shifts in engagement, capturing transitions across four affective states—boredom, engagement, frustration, and confusion.
  - Research outcomes accepted at **ECAI 2024** and **IEEE MIPR 2024** (*see [C.4], [C.5]*).
- **Explainable AI Methods and Robustness Analysis**
  Student Research Assistant, TU Darmstadt Dec. 2022 – Mar. 2023
  - Conducted a comprehensive literature review of explainable AI (XAI) methods and implemented approaches including Integrated Gradients, Grad-CAM, and SmoothGrad.
  - Evaluated robustness of XAI methods under adversarial attacks (DeepFool, Carlini–Wagner L2) by analyzing changes in focus maps and to measure sensitivity of XAI visualizations under perturbations.
- **C-Arm Orientation Tracking for Surgical Workflow Optimization** *Dr. Anirban Mukhopadhyay, TU Darmstadt)*
  Dec. 2022 – Mar. 2023
  - Developed real-time methods for 3D C-arm orientation estimation from 2D images using ML and analytical geometry approaches; complete pipeline implemented and preliminary work submitted to **MICCAI-AE CAI 2023**.
- **Human Behavior Modeling and Yoga Pose Classification**
  Research Intern, with Dr. Varun Dutt and Dr. Arnav Bhavsar, IIT Mandi iHub & HCL Dec. 2021 – Mar. 2022
  - Developed LSTM and LRCN-based pose classification method for real-time gamified environments.
  - Modeled human gameplay behavior using LSTM-based next-action prediction to analyze strategic decision-making and behavioral patterns in reinforcement learning settings for Unity-RL agents.
  - Work recognized among top 10 nationwide teams at the *ACM-W Hackathon 2021*.

Industry Experience

- **Microsoft** *Jan. 2024 – Present*
  Data Scientist (AI Research) Hyderabad, India
  Microsoft Security Platform - Research in multi-agent systems, AI alignment, and compliance automation.
  - Developed the **Security Analysis Agent**, evolving from chained workflows to autonomous agentic designs with context engineering and self-correcting reasoning, achieving **89%** precision and earning the *AI Innovation Award*. Currently working to improve security reasoning(planning-coding) depth through RL post-training
  - Proposed and implemented the **Multi-Agent Prompt Tuner**, a scalable framework for coordinated prompt optimization across interdependent agents using hierarchical prompt design, textual gradient-based feedback, and dynamic reward generation to enable simultaneous tuning while maintaining task coherence (*patent filed*, see [P.2]).
  - Designed the **Swarm Pinboard**, a multi-agent orchestration framework *inspired by swarm intelligence*, enabling autonomous collaboration across a graph of intelligent nodes with reasoning capabilities for complex workflows.
  - Built a **Security Copilot Skill** for natural language–driven data-table search and API generation, deployed to over 2,000 enterprise tenants. Optimized payload generation and API mapper modules; conducted extensive accuracy testing. Improved RAG performance from **60% to 91%**.
  - Developed a multi-agent system for visual guide generation for enterprise workflows (*patent submitted*, see [P.3]).
  - Proposed a policy-driven value taxonomy integrating human ethics and compliance objectives for assessing LLM organizational alignment, resulting in ACL SICon 2025 publication (*see [C.2]*).
- **Intuit India** *May 2023 – Jul. 2023*
  Software Engineer Intern Bangalore, India
  - Redesigned the internal *TestEasy* platform from a monolithic to a microservices-based architecture, integrating downstream document, consent and identity pipelines.
  - Implemented and deployed a production-ready test account cloning feature with unit and integration tests.

- **Fellowship.AI** *Oct. 2022 – Dec. 2022*
  Data Science Fellow Remote
  - Reproduced and modularized the **Total Relighting** pipeline for portrait relighting and background replacement.
  - Integrated open-source models (Semantic Guided Human Matting, GracoNet Object Placement, Harmonizer) into an interactive real-time demo with seamless pipeline orchestration for end-to-end processing.

Standardized Tests

TOEFL (iBT): 119/120 — Listening: 30, Reading: 29, Speaking: 30, Writing: 30
GRE General Test: 326/340 — Quant: 168, Verbal: 158, Analytical Writing: 4.0

Awards and Talks

- **Invited Talk**, "How Agentic Systems Can Be Developed for Organizational Information Processing Using RAG." at IIT Hyderabad for Digital Transformation series *Sept. 2025*
- **AI Innovation Award**, Microsoft Security Platform - For Security Analysis Agent *2025*
- **Academic Excellence Award (CSE)**, IIT Mandi - Highest SGPA in 5th and 6th semesters. *2023, 2024*
- **Top 5 Rank**, Inter-IIT Tech Meet - Representing IIT Mandi. Also achieved Top 10 Rank in 2022. *2023*
- **Top 10 Finalist**, ACM-W India Hackathons *2021, 2022*
- **Winner**, LLM Hack 2024, Atlas Astrathon, Vision Arcadia, Build-a-Bot, and Hack36 organized at IIT Mandi *2024*

Teaching and Leadership Experience

- **CS672 Advanced Deep Learning** Feb.-May 2025
  Research group lead, guided two UG student groups on GNNs and NCA, Advised by Dr. Arnav Bhavsar *SCEE, IIT Mandi*
- **CS202 Paradigms of Programming** Feb.-May 2024
  Teaching assistant, advised by Dr. Varunkumar Jayapaul *SCEE, IIT Mandi*
- **IC272 Data Science III** Aug.-Dec. 2023
  Teaching assistant, advised by Dr. Dilip A. D. *SCEE, IIT Mandi*

Skills and Relevant Courses

- **Relevant Courses:** Advanced Deep Learning, Computer Vision, Learning Robots, Probabilistic Graphical Models, Text Analytics, Information and Database systems, Computer Orgnanization, Data Structures and Algorithms
- **Programming Languages:** Python, C++, Java, JavaScript, C, Shell Scripting, LaTeX
- **Frameworks & Tools:** LangChain, AutoGen, LangGraph, FastAPI, Docker, Git, Azure, PyTorch, TensorFlow

References

- **Dr. Monojit Choudhury**
  Professor of Natural Language Processing
  Mohamed bin Zayed University of Artificial Intelligence.
  Email: monojit.choudhury@mbzuai.ac.ae
- **Dr. Sandipan Dandapat**
  Associate Professor, Department of Computer Science and Engineering
  Indian Institute of Technology Hyderabad
  Email: sdandapat@cse.iith.ac.in
- **Dr. Rajiv Ratn Shah**
  Professor, MIDAS Lab
  Indraprastha Institute of Information Technology Delhi
  Email: rajivratn@iiitd.ac.in