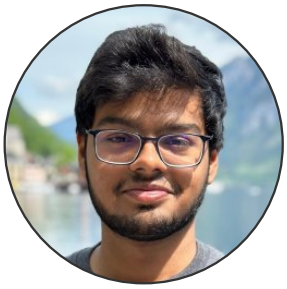

Ashmit Khandelwal

ashmit.contact@gmail.com | homepage | ashmitkx | ashmitkx | g-scholar

Education

Birla Institute of Technology and Science, Pilani — *India*

B.E. Computer Science, Minor in Data Science — *2020 - 2024*

GPA: **9.5/10.0**

Consistently awarded a merit-based scholarship each semester, awarded to the top 2% students.

Work Experience

Research Fellow - Microsoft Research — *Bangalore, India*

Advisors: Nagarajan Natarajan, Amit Sharma — *Jul 2024 - Present*

- Formally defined the task of **Deep Research** and introduced *LiveDRBench*, one the first benchmarks to evaluate deep research systems. Highlighted **performance gaps** in the deep research systems developed by OpenAI, Perplexity, and Gemini.
- Developing **adversarial methods** with LLMs to generate hard negatives for training robust verifiers on mathematical reasoning.
- Studied robustness of Direct Preference Optimization (DPO) under noisy labels produced by models ranging from trained verifiers to SLM judges and GPT-4.

Research Intern - Adobe — *Noida, India*

Advisors: Yaman Kumar (Adobe), Changyou Chen (SUNY Buffalo) — *May - Aug 2023*

- Simulating and optimizing for behavioural aspects of video/image content, such as likability and rewatchability, using a large language model. The first work to **embed content and the elicited human response** in the same space.
- Successfully **embedded vision** into a Llama LLM and fine-tuned it to understand the relationship between human behavior and video content. **Beat GPT-4**, showing that current SoTA models do not understand behavior.
- Automated scraping and processing of **terabytes of data** for multiple projects.

Publications

Large Content And Behavior Models To Understand, Simulate, And Optimize Content And Behavior — *Spotlight at ICLR*

A Khandelwal, A Agrawal, A Bhattacharyya, YK Singla, et al. — *2024*

Large Content and Behavior Models (LCBMs) reintroduce **behavior tokens** into LLM training data to simulate and explain human/audience behavior. These models generalize across content types and **adapt to diverse behavior domains**, demonstrated on the newly introduced Content Behavior Corpus (CBC).

Characterizing Deep Research: A Benchmark and Formal Definition — *Submitted to ICLR 2026*

A Java*, **A Khandelwal***, S Midigeshi*, et. al. | *alphabetical ordering, equal contribution — *2025*

Formally defined the **Deep Research (DR)** task as reasoning-intensive, high fan-out information search, distinguishing it from multi-hop search. Introduced **LiveDRBench**, a benchmark of 100 diverse and challenging tasks, revealing significant performance gaps in current DR systems – OpenAI, Perplexity, and Gemini.

Benchmarking VLMs' Reasoning About Persuasive Atypical Images — *WACV*

S Malakouti, A Aghazadeh, **A Khandelwal**, A Kovashka — *2025*

Benchmarked vision-language models on inferring the message in **rhetorical and atypical imagery used in advertisements**, through three novel tasks. Found that VLMs struggle with reasoning about atypicality, but simple methods enable **atypicality-aware verbalization** and improve rhetorical image understanding.

Research Explorations

Adversarial Generators for Training Robust Verifiers

Microsoft Research

Advisor: Nagarajan Natarajan

Ongoing

- Working with adversarial methods to generate **hard negatives** for mathematical reasoning, such as exploiting uncertainty in Process and Outcome Reward Models, and deceptive prompting.
- Leveraging these hard negatives to train more **robust verifiers**, improving their accuracy and resistance to misleading or tricky examples at **test time**.

Semi-Supervised Segmentation and VQA on Aerial Flood Images

BITS Pilani

Advisor: Sravan Danda

2022 - 2023

- Designed a **semi-supervised segmentation** and **graph-based VQA** system for the FloodNet challenge, combining CutMix and Cross Pseudo Supervision.
- The VQA module uses **geodesic dilation** and **morphological operations** on segmentation maps, followed by reasoning over **4-adjacency graphs** to count connected components.

Teaching Experience

Operating Systems

BITS Pilani

Teaching Assistant

Aug - Dec 2023

- Co-instructed for the Introduction to Machine Learning and Deep Learning course.
- Teaching the mathematical theory and providing python implementations of Machine Learning algorithms and Deep Learning models.

Object Oriented Programming

BITS Pilani

Teaching Assistant

Jan - May 2023

- Co-instructed for the Introduction to Machine Learning and Deep Learning course.
- Teaching the mathematical theory and providing python implementations of Machine Learning algorithms and Deep Learning models.

Introduction to ML and DL

BITS Pilani

Instructor | Student Run Effort

Nov 2022 - May 2023

- Co-instructed for the Introduction to Machine Learning and Deep Learning course.
- Teaching the mathematical theory and providing python implementations of Machine Learning algorithms and Deep Learning models.

Academic Honors

2020-24 **Merit Scholarship,** Awarded to top 2% students for academic excellance — *BITS Pilani*