



# Ashmit Khandelwal

✉ ashmit.contact@gmail.com | 🏠 ashmitkx.github.io | 📧 ashmitkx | 🌐 ashmitkx | 🎓 G-Scholar

## Education

### Birla Institute of Technology and Science, Pilani

India

B.E. COMPUTER SCIENCE WITH DISTINCTION, MINOR IN DATA SCIENCE

2020 - 24

GPA: **9.5/10.0**, Rank: **#10/262**

**Selected Courses:** Machine Learning (A), Artificial Intelligence (A-), Applied Statistical Methods (A-)

Consistently awarded a merit-based scholarship each semester, awarded to top 5% of students.

## Work Experience

### Research Fellow - Microsoft Research

Bangalore, India

**ADVISORS:** [NAGARAJAN NATARAJAN](#), [AMIT SHARMA](#)

Jul 2024 - Present

- Working on **probing and steering LLM search** and reasoning trajectories, to make them more reliable and transparent.
- Formally defined **Deep Research (DR)**, a previously under-defined class of LLMs for large-scale search and reasoning. Introduced *LiveDR-Bench*, one the first benchmarks to evaluate DR systems. Highlighted performance gaps in the DR systems from OpenAI, Perplexity, and Gemini.
- Studied **robustness of Direct Preference Optimization (DPO)** under noisy LLM labels produced by methods ranging prompt-based ones to trained verifiers.

### Research Intern - Adobe

Noida, India

**ADVISORS:** [YAMAN KUMAR \(ADOBE\)](#), [CHANGYOU CHEN \(SUNY BUFFALO\)](#)

Summer 2023

- Modeled human behavior** from YouTube videos to predict rewatchability, engagement metrics, and audience sentiment at scale across diverse content domains.
- Designed and **trained multimodal LLM** architectures that integrated video and textual representations, achieving significant performance gains and surpassing GPT-4 on multiple behavior prediction tasks.
- Built pipelines for video, audio, and text scraping and processing, automating the collection of **terabytes of web data**.

## Publications

\* = equal contribution

### Large Content And Behavior Models To Understand, Simulate, And Optimize Content And Behavior

Spotlight at ICLR

A KHANDLWAL\*, 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 in response to different forms of content. 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

SEA Workshop at NeurIPS

A JAVA\*, A KHANDLWAL\*, S MIDIGESHI\*, ET. AL.

2025

Deep Research (DR) systems lacked a formal definition despite the emergence of several proprietary and open-source models. This work established a concrete definition of DR, introduced LiveDRBench as one of the first DR benchmarks along with a recursive evaluation framework. Highlighted performance gaps in DR systems from OpenAI, Perplexity, and Gemini.

### Benchmarking VLMs' Reasoning About Persuasive Atypical Images

WACV

S MALAKOUTI\*, A AGHAZADEH\*, A KHANDLWAL, 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

---

## Measuring the Unpredictability of Human Society

Adobe

ADVISORS: [YAMAN KUMAR](#), [VEEKY BATHS](#)

Fall 2023

- Measured monthly Jensen–Shannon divergence in English Wikipedia visit and edit distributions, revealing a **steady rise in entropy** and growing randomness in collective information.
- **Simulated event propagation** on a graph-based SIR model to study how new events increase uncertainty and drive the observed rise in entropy.

## Semi-Supervised Segmentation and VQA on Aerial Flood Images

BITS Pilani

ADVISOR: [SRAVAN DANDA](#)

2022 - 23

- 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 | ~250 STUDENTS

Fall 2023

- Undergraduate course on operating system principles, including process management, memory management, and semaphore mechanisms.
- Responsibilities included lab sessions, lab test preparation, and student mentoring.

## Introduction to ML and DL

BITS Pilani

INSTRUCTOR | ~40 STUDENTS

Spring 2023

- Student run course on foundational ML/DL concepts and algorithms, and programming with python, numpy, and pytorch.
- Responsibilities included curriculum design, lecture delivery, assignment creation, and student mentoring.

## Object Oriented Programming

BITS Pilani

TEACHING ASSISTANT | ~250 STUDENTS

Fall 2022

- Undergraduate course introducing object-oriented programming concepts using Java.
- Responsibilities included lab sessions, lab test preparation, student mentoring.

# Academic Honors

---

2020-24 **Merit Scholarship**, Awarded to top 5% students for academic excellence

BITS Pilani