Harshita Chopra

hchopra3@cs.washington.edu Website Google-scholar LinkedIn

Research Interests

Machine Learning (ML), Reasoning and Planning in Large Language Models (LLM), Personalization, User Modeling

Education

Ph.D. in Computer Science & Engineering 2024 – present
University of Washington, Seattle, US – *Advisor*: Prof. Chirag Shah

B.Tech. in Information Technology 2018 – 2022
Indraprastha University, Delhi, India – *GPA*: 9.01/10.0

Experience

Microsoft Research Summer 2025
Research Intern – Mentor: Dr. Krishna Chintalapudi, Dr. Ryen White, Dr. Suman Nath Redmond, US

- Personalization and memory retrieval in tool-calling AI agents.

Adobe Research Jul 2022 – Aug 2024
Research Associate II Bangalore, India

- Developed **ML and optimization** models for predictive user segmentation and reach maximization
- **Trained** and **evaluated LLMs** for simulating online user behavior as a sequence of browsed pages and transaction events.
- Shipped work across 2 **products**, mentored 2 graduate and 9 UG interns; published 3 full **papers** and 8 **patents**.

Adobe Research Summer 2021
Research Intern – Mentor: Dr. Atanu R Sinha Bangalore, India

- Modeled **partially observed** user behavior on websites using actor-critic **RL** and **Boosting** techniques.
- Recognized among top 8 out of 84 interns; offered a full-time role in the research team.

University of California, Irvine Feb 2021 – 2022
Undergraduate Researcher – Advisor: Prof. Nia Dowell-Nixon Remote

- Worked at the Language and Learning Analytics Lab. Introduced a framework to detect and track the **temporal flow of topics** in academic discussion forums and quantify the course-centricity of topics.

IIIT Delhi Oct 2020 – 2021
Undergraduate Researcher – Advisor: Prof. Tavpritesh Sethi Delhi, India

- Demonstrated that new knowledge can be captured by tracking the **temporal evolution** of association between **entities in scientific literature** using NLP and community detection on entity networks.

Omdena Inc. Apr – Dec 2020
Machine Learning Engineer Remote

- Led end-to-end **AI4Good** projects with the United Nations Development Program (**UNDP**) and the World Resources Institute (**WRI**); developed predictive models of the Digital Divide and economic well-being using multi-modal data.

Publications

Conference:

[1] **H Chopra**, C Shah. "Feedback-Aware Monte Carlo Tree Search for Efficient Information Seeking in Goal-Oriented Conversations". **NeurIPS 2025** ☆ **Spotlight** [*top* 3%] | **ICLR 2025** Workshop on Reasoning & Planning for LLMs ***(Oral)***

[2] S Sundaresan, **H Chopra**, A R Sinha, K Goswami, N S Naidu, R Karan, N Anushka. "Subjective Behaviors and Preferences in LLM: Language of Browsing." **EMNLP 2025** ***(Oral)***. [*acceptance rate 22%*]

[3] **H Chopra**, A R Sinha, S Choudhary, R A Rossi, P Indela, V P Parwatala, S Paul, and A Maiti. "Delivery Optimized Discovery in Behavioral User Segmentation under Budget Constraint". **CIKM 2023** ***(Oral)***. [*acceptance rate 24%*]

[4] A R Sinha*, **H Chopra***, A Maiti, A Ganesh, S Kapoor, S Myana, and S Mahapatra. "The Role of Unattributed Behavior Logs in Predictive User Segmentation". **CIKM 2023** ***(Oral)***. [*acceptance rate 24%*]

[5] **H Chopra**, Y Lin, M A Samadi, J G Cavazos, R Yu, S Jaquay, and N Nixon. "Semantic Topic Chains for Modeling Temporality of Themes in Online Student Discussion Forums". **EDM 2023** (*Oral, Best paper award nominee*)

[6] **H Chopra**, Y Lin, M A Samadi, J G Cavazos, R Yu, S Jaquay, and N Nixon. "Modeling Student Discourse in Online Discussion Forums Using Semantic Similarity Based Topic Chains". Extended Abstract. **AIED 2022** *(Poster)*

Journal:

[1] R Pal, **H Chopra**, R Awasthi, H Bandhey, A Nagori, and T Sethi. "Predicting Emerging Themes in Rapidly Expanding COVID-19 Literature With Unsupervised Word Embeddings and Machine Learning: Evidence-Based Study". Journal of Medical Internet Research. JMIR 2022 (*impact factor 7.2, ranked Q1 in Medical Informatics*)

[2] **H Chopra***, A Vashishtha*, R Pal, Ashima, A Tyagi, and T Sethi. "Mining Trends of COVID-19 Vaccine Beliefs on Twitter With Lexical Embeddings: Longitudinal Observational Study". JMIR Infodemiology 2023.

Patents: * *equal contribution*

[1] S N Sunderesan, A R Sinha, **H Chopra**, K Goswami, R Karan, N S Naidu, N Anushka. "Heterogeneous LLMs for Subjective Behaviors". [Filed] US Patent Application.

[2] **H Chopra**, A R Sinha, S Mahapatra. "Utilizing Digital Page Sequence Tokens With Large Language Models to Generate Digital Content Predictions". [Filed] US Patent Application No.: 18/829,774

[3] **H Chopra**, S Choudhary, A R Sinha, S Surange-Dev, V Holtcamp, S Nair, Z Courtois, S Bhat. "Campaign Journey User Response Computer Simulation". [Filed] US Patent Application No.: 18/777,311

[4] V Porwal, **H Chopra**, A R Sinha, S K Modanwal, C N Reddy, Z Niaz. "Clustering Users According to Causal Relationships Among User Data". [Filed] US Patent Application No.: 18/609,625

[5] S Choudhary, A R Sinha, **H Chopra**, R A Rossi, V P Parwatala, P Indela, S Paul, S Guo. "Segment Discovery and Channel Delivery". [Filed] US Patent Application No.: 18/543,666

[6] A R Sinha, R A Rossi, S Choudhary, **H Chopra**, P Indela, V P Parwatala, S Paul, S Mahapatra, A Maiti. "Delivery Aware Audience Segmentation". [Filed] US Patent Application No.: 18/451,590

[7] A Maiti, A R Sinha, **H Chopra**, S Kapoor, A Ganesh, S Myana, S Mahapatra. "Generating Segments of Users Based on Unobserved Behaviors". [Filed] US Patent Application No.: 17/660,544

[8] A R Sinha, A Maiti, A Ganesh, **H Chopra**, S Myana, S Kapoor, S Mahapatra. "Systems and Methods for Content Customization". [Filed] US Patent Application No.: 17/813,622

Selected Projects

Feedback-Aware Planning for Information-Seeking under Uncertainty – UW, Seattle [NeurIPS '25]

- Developed a novel framework that uses **LLMs** for **seeking missing information** by applying **Monte Carlo Tree Search** to select questions that reduce uncertainty, as a part of **inference-time planning**.
- Proposed a **hierarchical feedback** mechanism to enable learning from past successful questioning trajectories.

Personalized Language Models for Predicting Online User Behavior – Adobe Research [EMNLP '25]

- Developed HeTLM (**Heterogeneity-aware Training** of LMs), a cluster-based training framework for **small language models** that learns diverse user browsing behaviors.
- Achieved significantly higher accuracy in **predicting next-page visits** and **purchase outcomes** compared to larger LMs.

Delivery Aware Discovery of Behavior-based User Segments – Adobe Research [CIKM '23]

- Developed a **joint stochastic optimization** model for effective discovery of user segments based on browsing behavior and matching them with marketing/media channels that maximize reach, under a given **budget constraint**.

Detection of Emerging Themes in Scientific Literature – IIIT Delhi [JMIR '22]

- Detected and predicted disease-centric themes in COVID-19 literature using **temporal link prediction** and word embeddings to track the **evolving semantic similarity** among entities. Demo: EvidenceFlow

Recognition & Impact

- Received the **Microsoft Research Redmond Azure Credit Award** of $4500 for Research Collaboration post internship.
- Recipient of the **ICLR 2025 Student Registration Grant** for attending the conference and workshops.
- **Department Rank 1** in the final semester of undergraduate degree – Scored GPA 10/10.
- **GHC 2021 Scholar** – Won the student scholarship to attend the Grace Hopper Celebration.
- **Vice Chairperson – IEEE** Women in Engineering (2020-21), Student Branch Executive Committee of MSIT.

- **Invited Talk** at IE Business School, Madrid – On Collaborative AI projects and Ethics in the Information Age. 2020.
- **Invited Speaker** on a Panel of Experts from NASA and Harvard University – "Building Artificial Intelligence through Collaborative Innovation", a virtual event with over 1000 registrations, hosted by Omdena Inc. in 2020.
- Among **top 10 students** from the country selected for the **Digital India** Internship at NIC Headquarters, Winter 2019-20. My redesigned interface of eTransport web portal 'Vahan Citizen Services' was **deployed nationwide** by Govt of India.

Technical Skills

- Programming languages: Python, SQL, C/C++, Bash
- Packages & Frameworks: PyTorch, TensorFlow, Keras, scikit-learn, SciPy, Git, PySpark, LaTeX

Academics

- **Ph.D. Coursework**: CSE 573 (Artificial Intelligence), CSE 546 (Machine Learning), IMT 526 (Building and Applying LLMs), CSE 599 (Social Reinforcement Learning)
- **Teaching Assistant**: CSE 447 (Natural Language Processing) with Prof. Noah Smith

Volunteer Experience

- **Data Science Researcher** – PathCheck Foundation, MIT
  Brainstormed effective and affordable COVID-19 mitigation strategies. Our submission "Privacy-preserving Crowdsourcing for Citizen Engagement in Pandemics" was awarded as one of the '*Highly Commended Solutions*' at The Trinity Challenge 2021
- **Data Analyst** – Red Dot Foundation | Safecity
  Developed dashboards based on inferential statistics on past 8 years of crowd-sourced harassment reports. 2020