

# Harshita Chopra

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

## RESEARCH INTERESTS

Machine Learning (ML), AI-assisted Decision Making, Agents, Personalization, User Modeling, Recommender Systems

## EDUCATION

- Ph.D. in Computer Science & Engineering** 2024 – present  
University of Washington, Seattle, WA, USA – GPA: 4.0/4.0
- B.Tech in Information Technology** 2018 – 2022  
GGS Indraprastha University, Delhi, India – GPA: 9.01/10.0

## EXPERIENCE

- Graduate Research Assistant** – University of Washington, Seattle Sep 2024 – present  
InfoSeeking Lab. Advisor: [Prof. Chirag Shah](#)
- Working on adaptive AI agents for sequential decision making under uncertainty; using ML and search algorithms.
  - Developing efficient methods for interactive information acquisition with LLMs.
- Research Associate II** – Adobe Research, India Jul 2022 – Aug 2024  
Data-driven Systems and Insights Group.
- Developed ML and optimization models for predictive user segmentation, reach maximization under marketing budget constraint, and simulation of user response in multi-touchpoint journeys. Worked on LLMs for digital marketing tasks.
  - Transferred models to Adobe products, mentored 2 graduate and 9 UG interns; published 2 full papers and filed 7 patents.
- Research Intern** – Adobe Research, India Summer 2021  
Big Data Intelligence Lab. Mentor: [Dr. Atanu R Sinha](#)
- Devised a user modeling approach that accounts for the latent effect of partially observed behavior by leveraging the click-stream logs unattributed to specific users on a firm's website. Implemented the idea end-to-end on *TensorFlow*.
  - Among the 8 out of 84 interns who were offered full-time position in the research team.
- Undergraduate Researcher** – University of California, Irvine Spring 2021  
Language and Learning Analytics Lab. Advisor: [Prof. Nia Dowell-Nixon](#)
- Introduced a framework to detect and track contextually coherent topics in student discourse over time. Modeled the nature of discourse under a topic as general vs. course-centric. Worked with Ph.D. students; published a full paper and poster.
- Undergraduate Researcher** – IIIT Delhi Fall 2020 – 21  
TavLab. Advisor: [Prof. Tavpritesh Sethi](#)
- Demonstrated that new knowledge can be captured by tracking the temporal evolution of association between entities in scientific literature. • Conducted a spatiotemporal analysis of the vaccine infodemic using NLP.
- Machine Learning Engineer** – Omdena Inc. Apr – Dec 2020
- Led and managed ML tasks within large cross-functional teams, working closely with the United Nations Development Program (UNDP) and the World Resources Institute (WRI), on real-impact AI4Good projects involving predictive analyses of the digital divide and modeling economic well-being using multi-band satellite imagery.

## PUBLICATIONS

Conference:

- 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". In the 32nd ACM International Conference on Information and Knowledge Management. [CIKM 2023](#). (acceptance rate 24%)
- 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". In the 32nd ACM International Conference on Information and Knowledge Management. [CIKM 2023](#). (acceptance rate 24%)
- 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". In the 16th International Conference on Educational Data Mining. [EDM 2023](#). (best paper award nominee)
- 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. In the 23rd International Conference on Artificial Intelligence in Education. [AIED 2022](#).

Journal:

\* equal contribution

- 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:

- [1] **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
- [2] **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
- [3] 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
- [4] 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
- [5] 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
- [6] 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
- [7] 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

##### 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**.

##### User Modeling with Unattributed Behavior Logs – Adobe Research [CIKM '23]

- Implemented an **actor-critic training** algorithm to find segments of users which are predictive of a target KPI, by encoding users' session-wise clickstream logs using a **hierarchical attention network**.
- Introduced the use of unattributed clickstream logs by proposing a new loss function that additionally incorporates the effect of partially observed browsing behavior on a firm's website, using **principles of boosting**.

##### Modeling Student Discourse via Semantic Topic Chains – UC Irvine [AIED '22, EDM '23]

- Designed a framework to **detect topics** in student discourse over time intervals and connected the semantically similar topics to create chains, using **word mover's distance** for finding dissimilarity between topic words.
- Modeled each topic's discourse as general or course-centric, and analyzed the evolving **course-centricity** of detected topic chains.

##### 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](#)

#### ACTIVITIES & ACHIEVEMENTS

- 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 – case study 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 hosted by Omdena with over 1000 registrations. 2020.
- Among top 10 students from the country selected for the Digital India Internship at NIC Headquarters, Winter 2019-20. My work on redesigning the interface of eTransport web portal 'Vahan Citizen Services' was deployed nationwide by Govt of India.

#### TECHNICAL SKILLS

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

#### 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