Arpandeep Khatua

202-710-8808 | arpandeepk@gmail.com | github.com/akhatua2 | arpandeep.com

Education

Stanford University Sep 2024 – Jun 2026

Master of Science, Computer Science *GPA: 3.9/4.00*

- Coursework: Conversational Virtual Assistants, Human Centered NLP, Reinforcement Learning, Deep RL

University of Illinois at Urbana-Champaign Aug 2019 – Dec 2022

BS Computer Engineering (Minor, Mathematics), Highest Honors, Bronze Tablet *GPA: 4.00/4.00*

- Senior Thesis: Generating Large Real World and Synthetic Graph Datasets for GNN Applications
- Coursework: Natural Language Processing, Machine Learning, Deep Learning, Algorithms, Parallel Programming, Data Structures, Discrete Structures, Database Systems, Computer Systems, Number Theory

Industry Experience

Meta Inc. Jan 2023 – Sep 2024

Software Engineer IC4 (AI-Driven Groups) *Menlo Park, CA*

- Worked on Groups Admin Moderation to reduce admin work and streamline user generated content.
- Created automated weekly summary post generation system for groups, synthesizing group activity and key discussions. Built onboarding flows increasing daily active group creations by 3%.
- Full-stack development using PHP, JavaScript, GraphQL, and React for web and mobile platforms.

Meta Inc. May 2022 – Aug 2022

Software Engineering Intern (Live Video Community) *New York, NY*

- Created infinite scroll comments for FB Live resulting in 1M hours increase in total live watch time per week. Revamped FB Live Polls and built overview tab with stat-sig increases via A/B testing over 200M+ users.
- Simplified E2E testing framework and documentation for internal languages used by 50+ teams. Received highest intern rating (GE - greatly exceeds expectations).

Facebook Inc. May 2021 – Aug 2021

Software Engineering Intern (Commerce Shop Rankings) *Menlo Park, CA*

- Built personalized product retrieval transformer model based on user history and viewing habits. Model achieved 35% win over production retrieval model and 30% accuracy increase over online ranking model via A/B testing over 30M users.
- Optimized data collection pipeline throughput by 60× (6000%), enabling real-time queries. Created new API for querying PyTorch models, now used as internal tool across teams.

Publications

1. **Khatua, A.**, Zhu, H., Tran, P., Prabhudesai, A., Yu, X., Sadrieh, F., Lieberwirth, J. K., Fu, Y., Ryan, M. J., Pei, J., & Yang, D. The Curse of Coordination: Why Current AI Cannot Be Your Teammates. *in progress, targeting ICML*
2. **Khatua, A.***, Wu, S.*, Choi, E.*, Wang, H., He-Yueva, J., Weerasooriya, C., Wei, W., Yang, D., Leskovec, J., & Zou, J.* HumanLM: Building Digital Humans from Large Language Models. *in progress, targeting ICML*
3. Durante, Z., Singh, S.*, **Khatua, A.***, Agarwal, S., Tan, R., Lee, Y. J., Gao, J., Adeli, E., & Fei-Fei, L. VideoWeave: A Data-Centric Approach for Efficient Video Understanding. *NeurIPS 2025 Workshop (Oral)*
4. Semnani, S. J., Burapacheep, J., **Khatua, A.**, Atchariyachanvanit, T., Wang, Z., & Lam, M. S. Detecting Corpus-Level Knowledge Inconsistencies in Wikipedia with Large Language Models. *EMNLP 2025*
5. Kugo, N., Li, X., Li, Z., Gupta, A., **Khatua, A.**, Jain, N., Patel, C., Kyuragi, Y., Ishii, Y., Tanabiki, M., Kozuka, K., & Adeli, E. VideoMultiAgents: A Multi-Agent Framework for Video Question Answering. *arXiv:2504.20091, 2025*
6. **Khatua, A.**, Mailthody, V., Taleka, B., Song, X., Ma, T., Bigaj, P. & Hwu W. IGB: Addressing The Gaps In Labeling, Features, Heterogeneity, and Size of Public Graph Datasets for Deep Learning Research. *KDD 2023*

Research Experience

Stanford Artificial Intelligence Laboratory Sep 2024 – Present

Graduate Researcher, Advisors: Prof. Diyi Yang, Prof. Jure Leskovec, Prof. Monica Lam *Stanford, CA*

- **Cotomata: Multi-Agent Collaboration Benchmark.** Designed multi-agent interaction protocols and failure analysis pipeline for evaluating LLM coordination in version-controlled programming environments. Models drop from 70-75% accuracy (single-agent) to 10-20% (multi-agent); 60%+ failures from mismatched shared-state assumptions.
- **HumanLM: Hierarchical RL User Simulation.** Built hierarchical RL modules for user simulation, designing stance and style reward functions and training the model across multi-level generations. Improved alignment with real user responses, raising LLM-judge similarity by 36% over SFT baselines.
- **Multilingual SWE-smith.** Extended SWE-smith by automating test environments construction for JS/TS, Java, Rust, and C++. Scaled procedural bug generation to any repo, producing mid-training data for improving Multi-SWE-Bench performance.
- **VideoWeave (NeurIPS 2025 Workshop).** Developed data-centric approach for efficient video understanding by splicing short captioned videos into synthetic long-context training samples. Achieved 3% absolute improvement on VideoMME under identical compute constraints without architectural modifications.
- **Wikipedia Corpus-Level Inconsistency Detection (EMNLP 2025).** Built LLM-based system using claim extraction and cross-article verification to detect internal contradictions across Wikipedia. Proved 3%+ of claims are internally inconsistent through large-scale corpus analysis.
- **VideoMultiAgents: Multi-Agent Video QA Framework. (Panasonic Research)** Built multi-agent system with role specialization (visual perception, temporal reasoning, answer synthesis) for video question answering, improving performance through collaborative reasoning and information aggregation improving zero-shot performance by up to 6% over prior SOTA.

IMPACT Lab, University of Illinois Sep 2021 – Dec 2022

Undergraduate Researcher, Advisors: Prof. Wen-Mei Hwu, Dr. Vikram Sharma Mailthody *Champaign, IL*

- **IGB: Illinois Graph Benchmark (KDD 2023).** Created largest academic GNN dataset: 260M nodes, 4B edges, 220M labels (162× larger than prior datasets). Adopted as MLPerf industry standard for GNN benchmarking.

FORWARD Data Lab, University of Illinois Jan 2022 – Dec 2022

Undergraduate Researcher, Advisor: Prof. Kevin Chen-Chuan Chang *Champaign, IL*

- Built data-driven search engine with ranking and retrieval models achieving 97% accuracy. Automated Wikipedia-style article generation by analyzing structural relationships across articles.

Selected Honors & Awards

Scholarships & Fellowships: O. Thomas and Martha S. Purl Scholarship ($3,400, awarded to 2 of 2,000 students), NCSA Student Pushing Innovation Fellow ($6,000), C3SR-URAI Scholar ($1,000), Illinois Geometry Lab Scholar (×2)
Academic Honors: Edmund J. James Scholar, Dean's List (2019–2022), Highest Honors, Bronze Tablet
Leadership: HKN Honor Society Corporate Director, ECE PULSE Corporate Committee, PURE Vice President
Teaching: CS 224V (Conversational Virtual Assistants, Stanford), ECE 391 (Computer Systems), ECE 313 (Probability), ECE 210 (Signals), ECE 110 & Honors Lab
Invited Talks & Presentations: CLAIRE (Corpus-Level Article Inconsistency REsolver), Wikimedia Research, 2025; IGB Benchmark, Amazon AI, 2023; IGB Benchmark, NVIDIA Research, 2023
Additional Publications & Posters: **Khatua, A.**, Agarwal, A., & Chang, K. Generating High-Level Article Structure based on Topic using Two-stage seq2seq Model; **Khatua, A.***, Li, X.*, Garitsis, E., & Hildebrand, A. Large Scale Statistical Analysis of the Randomness of π Continued Fraction Digits; Hazim, A.*, **Khatua, A.***, Stratton, A.*, Yang, P., & Diesner, J. Detection of Displacement and Needs Related to the Russia–Ukraine Conflict (*IGL Open House 2022 Poster*); Jin, A.*, **Khatua, A.***, Li, X.*, Singh, S.*, Zhang, Z.*, Garitsis, E., & Hildebrand, A. Analysis of the Continued Fraction Digits of π (*MAA MathFest 2022 Poster*); **Khatua, A.**, Worthey, G., Capitanu, B., & Downie, J. Detection of Page Objects from the HathiTrust Digital Library (*Engineering Open House 2022 Poster*)