

Bhuvesh Kumar

Senior Research Scientist · Ph.D. in Computer Science

✉ kumarbhuvesh@gmail.com | 🌐 www.bhuveshkumar.com | 📄 bhuvesh-kumar | 🏠 Google Scholar

Experience

Snap Inc.

SENIOR RESEARCH SCIENTIST

Bellevue, WA

Sept. 2024 - Present

- Co-created the internal Python/PyTorch library powering sequential user modeling, Semantic-ID-based generative recommendation, and content representations, and now serve as its sole maintainer and technical lead; actively used by all 5 research scientists on the team, 10+ research-intern projects, and 20+ machine-learning engineers across Content, Lens, and Ads.
- Designed and implemented core components for LLM/VLM-derived multimodal Semantic-ID representations and Semantic-ID-based generative recommendation, spanning continued pre-training (CPT), supervised fine-tuning (SFT), Ray-based reinforcement-learning post-training, and fast inference with TensorRT-LLM.
- Designed and implemented core components of large-scale TorchRec-based sequential recommendation models with sharded embedding tables.
- Through this platform, enabled multiple recommendation launches across Content, Ads, Lens, and Notifications, including cross-domain user-representation models such as the Universal User Embedding Model, whose representations are now used by most personalization models at Snap.
- Co-created GRID by open-sourcing a subset of the platform; GRID has 700+ GitHub stars, and its accompanying paper received the CIKM 2025 Best Paper Award.

TikTok

MACHINE LEARNING SCIENTIST

Bellevue, WA

Jan. 2023 - Sept. 2024

- Developed large-scale sequential user models and personalization algorithms for TikTok U.S. Core Recommendation.
- Introduced and owned a new final re-ranking stage that moved TikTok U.S. recommendation from point-wise item scoring to list/session-level Transformer sequence modeling, increasing total watch time by 2.4%.
- Designed a personalized mixing stage for organic, cold-start, and commercial content; coordinated with horizontal teams to preserve cold-start and commercial-content distribution goals while optimizing long-term user retention.
- Led the migration of TikTok U.S. Friends, Hashtag, and Social recommendation services from a single data center to a disaster-tolerant multi-data-center design.

Amazon

APPLIED SCIENTIST INTERN

Seattle, WA

May 2021 - Aug. 2021

- Introduced a group-fairness notion for multi-armed bandits and developed a no-regret algorithm balancing fairness and performance.

University of Washington

VISITING RESEARCH SCIENTIST

Seattle, WA

Jan. 2021 - Apr. 2021

- Designed incentive-compatible, differentially private auction mechanisms for revenue maximization in repeated auctions.

Meta

RESEARCH INTERN, CORE DATA SCIENCE

Menlo Park, CA

May 2020 - Nov. 2020

- Developed optimal-spend-plan estimation and automatic bid-pacing algorithms for budget-constrained advertising campaigns.

Meta

MACHINE LEARNING ENGINEERING INTERN

Menlo Park, CA

May 2019 - July 2019

- Improved Dynamic Product Ads ranking and increased ad revenue while reducing model memory by 60 TB (86%).

Georgia Institute of Technology

GRADUATE RESEARCH ASSISTANT

Atlanta, GA

Aug. 2017 - Dec. 2022

- Developed robust learning algorithms with provable guarantees using online learning, algorithmic game theory, and differential privacy.

Education

Georgia Institute of Technology

PH.D. IN COMPUTER SCIENCE (ADVISORS: DR. JACOB ABERNETHY AND DR. JAMIE MORGENSTERN)

Atlanta, Georgia

2017 - 2022

Georgia Institute of Technology

M.S. IN COMPUTER SCIENCE (SPECIALIZATION IN MACHINE LEARNING, GPA: 4.0/4.0)

Atlanta, Georgia

2017 - 2022

Indian Institute of Technology, Kanpur

B.TECH IN COMPUTER SCIENCE AND ENGINEERING (GPA: 9.7/10)

Kanpur, India

2013 - 2017

Selected Recent Publications

- **[8] Semantic IDs for Recommender Systems at Snapchat: Use Cases, Technical Challenges, and Design Choices**
SIGIR 2026
C. M. Ju, T. Zhao, L. Neves, L. Collins, B. Kumar, et al.
- **[7] LLM-Based Generative Retrieval for Snapchat Content Recommendation**
2026 Preprint
L. Collins, J. Ren, D. Loveland, B. Kumar, et al.
- **[6] Exploiting ID-Text Complementarity via Ensembling for Sequential Recommendation**
RecSys 2026
L. Collins, B. Kumar, C. M. Ju, T. Zhao, D. Loveland, L. Neves, N. Shah
- **[5] Training-Free LLM-Based Recommendation with Post-LLM Item Refinement Using Collaborative Signals**
CIKM 2026
K. Kim, S. Kim, G. Lee, S. Kang, S. Kim, L. Collins, B. Kumar, et al.
- **[4] Sequential Data Augmentation for Generative Recommendation**
WSDM 2026
G. Lee, B. Kumar, C. M. Ju, T. Zhao, K. Shin, N. Shah, L. Collins
- **[3] Learning Universal User Representations Leveraging Cross-Domain User Intent at Snapchat**
SIGIR 2025
C. M. Ju, L. Neves, B. Kumar, L. Collins, T. Zhao, et al.
- **[2] Generative Recommendation with Semantic IDs: A Practitioner's Handbook**
CIKM 2025
C. M. Ju, L. Collins, L. Neves, B. Kumar, L. Y. Wang, T. Zhao, N. Shah
Best Paper Award; co-creator of GRID (700+ stars on GitHub)
- **[1] CoSearch: Joint Training of Reasoning and Document Ranking via Reinforcement Learning for Agentic Search**
COLM 2026
H. Zeng, L. Collins, B. Kumar, N. Shah, H. Zamani

Complete publication list: [Google Scholar](#).

Technical Skills

Programming & ML Python, C++, C; PyTorch, TorchRec, TensorFlow, verl
Distributed ML & Data Ray, vLLM, TensorRT-LLM; GCP (GKE, Vertex AI); SQL, Spark, Flink, Hive

Awards

2025 **Best Paper Award**, ACM CIKM
2019 **NeurIPS Travel Award**, Neural Information Processing Systems
2017 **Chair's Fellowship**, Georgia Institute of Technology
2014-16 **Academic Excellence Award (Dean's List)**, IIT Kanpur
2011 **KVPY Fellowship**, Government of India
2009 **NTSE Scholarship**, Government of India

Academic Service

Reviewer / Program Committee: ACL (2026); ICLR (2025, 2026); NeurIPS (2020, 2021, 2025); COLT (2020, 2021, 2023); SODA (2021, 2023); ICML (2022); JMLR (2021); ALT (2018, 2020).