

Yanbang Wang

Department of Computer Science
Cornell University
Bill & Melinda Gates Hall 344
Ithaca, NY 14850

+1 650-283-9267
ywangdr@cs.cornell.edu
<https://www.cs.cornell.edu/~ywangdr>
Google Scholar

SUMMARY

Machine-learning PhD researcher specializing in **LLMs for code generation**. At Google, led and first-authored **Gemini-SQL2**, Google's state-of-the-art LLM for SQL generation, ranking **#1 on the BIRD Bench**, the de-facto standard benchmark for text-to-SQL generation. Working full-stack across **post-training, agentic harness, and evaluation**. First-authored publications at NeurIPS (2), ICML (1), ICLR (2), and WebConf (2) on **LLM reasoning, retrieval, and verification** over structured data, large-scale recommendation, and deep learning on graphs.

EDUCATION

2025 - Present	Visiting Ph.D. Student Stanford University, Stanford, CA
2021 - Present	Ph.D. in Computer Science Cornell University, Ithaca, NY <i>Advisor:</i> Jon Kleinberg
2019 - 2021	M.S. in Computer Science Stanford University, Stanford, CA <i>Advisor:</i> Jure Leskovec
2015 - 2019	B.S. in Computer Science, Mathematics Hong Kong University of Science and Technology, Hong Kong, Hong Kong

PROFESSIONAL EXPERIENCE

2025 – 2026	Google Research, Research Intern - Led and first-authored Gemini-SQL2, Google's state-of-the-art text-to-SQL coding LLM – currently #1 on the industry-standard BIRD Bench (single-model track). (Google Announcement, Post 2, Post 3, Post 4) - Post-trained Gemini 3.1 Pro: SFT, RLVR, Param-efficient adaptor tuning, data synthesis. - Worked full-stack on the inference harness, including context retrieval, code analysis, prompt engineering, execution feedback, self-refinement loop, test-time scaling, validation, and eval pipeline.
2025	Meta AI, Research Intern - Redesigned the Transformer backbone for re-ranking for Facebook videos personalization and recommendation. - Successfully shipped – improving personalized video experience of all Facebook users (~3.1 billion).
2024	Meta AI, Research Intern - Researched and developed provably optimal negative sampling for contrastive learning. - Published as first-author an ICML paper [C.1].
2023	Microsoft Research, Research Intern - Researched and developed temporal graph-learning models for retrieving and detecting anomalies from network communication log, for Microsoft Defender and Microsoft 365. - Published as first-author a WebConf paper [C.5] and a patent [PA.1].

PUBLICATIONS & PATENTS

(Citations: 1452, H-Index: 11)

PREPRINTS & MANUSCRIPTS

- [P.1] **Yanbang Wang**, Qitian Wu, Sami Abu-el-Haija, Mohammadreza Pourreza, Michael Galkin, Hadi Hemmati, Hailong Li, Yeounoh Chung, Fatma Ozcan, Bryan Perozzi, Vahab Mirrokni. [Gemini-SQL2: Model, Harness, and System Design](#). *Under Review*.
- [P.2] **Yanbang Wang**, Sami Abu-el-Haija, Mohammadreza Pourreza, Michael Galkin, Hadi Hemmati, Yeounoh Chung, Fatma Ozcan, Bryan Perozzi. Graph-Language Models as Text-to-SQL Verifier. *Under Review*.
- [P.3] Sami Abu-el-Haija, Fatma Ozcan, **Yanbang Wang**, Brandon Mayer, Bryan Perozzi. Bidirectional Graph–Code Reasoning: Generating and Validating Queries through Structural Representations. *Under Review*.

PEER-REVIEWED CONFERENCE PUBLICATIONS

- [C.1] **Yanbang Wang**, Jon Kleinberg, Yanhong Wu. [Negative Sampling From the Ground Up](#). Proceedings of the 43rd International Conference on Machine Learning (**ICML**), 2026.
- [C.2] **Yanbang Wang**, Hejie Cui, Jon Kleinberg. [Microstructures and Accuracy of Graph Recall by Large Language Models](#). Advances in Neural Information Processing Systems (**NeurIPS**) 38, 2025.
- [C.3] **Yanbang Wang**, Jon Kleinberg. [From Graphs to Hypergraphs: Hypergraph Projection and its Reconstruction](#). Proceedings of the 12th International Conference on Learning Representations (**ICLR**), 2024.
- [C.4] **Yanbang Wang**, Jon Kleinberg. [On the Relationship Between Relevance and Conflict in Online Social Link Recommendations](#). Advances in Neural Information Processing Systems (**NeurIPS**) 37, 2023.
- [C.5] **Yanbang Wang**, Karl Hallgren, Jonathan Larson. [A Graph-based Framework for Reducing False Positives in Authentication Alerts in Security Systems](#). Companion Proceedings of the International World Wide Web Conference (**WebConf**), 2023 .
- [C.6] Haoteng Yin, Muhan Zhang, **Yanbang Wang**, Jianguo Wang, Pan Li. [Algorithm and system co-design for efficient subgraph-based graph representation learning](#). Proceedings of the VLDB Endowment (**VLDB**), 2022.
- [C.7] **Yanbang Wang**, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, Pan Li. [Inductive Representation Learning in Temporal Networks via Causal Anonymous Walks](#). Proceedings of the 9th International Conference on Learning Representations (**ICLR**), 2021.
- [C.8] **Yanbang Wang***, Pan Li*, Chongyang Bai, Jure Leskovec. [TEDIC: Neural Modeling of Behavioral Patterns in Dynamic Social Interaction Networks](#). Proceedings of the International World Wide Web Conference (**WebConf**), 2021 .
- [C.9] Pan Li, **Yanbang Wang**, Hongwei Wang, Jure Leskovec. [Distance Encoding: Design Provably More Powerful Neural Networks for Graph Representation Learning](#). Advances in Neural Information Processing Systems (**NeurIPS**) 34, 2020.

PEER-REVIEWED JOURNAL PUBLICATIONS

- [J.1] Haipeng Zeng, Xinhuan Shu, **Yanbang Wang**, Yong Wang, Liguang Zhang, Ting-Chuen Pong, Huamin Qu. [EmotionCues: Emotion-Oriented Visual Summarization of Classroom Videos](#). IEEE Transactions on Visualization and Computer Graphics (**TVCG**), 2020.

WORKSHOPS & EXTENDED ABSTRACTS

- [W.1] **Yanbang Wang**, Hejie Cui, Jon Kleinberg. Network Recall by Large Language Models. International Conference on Computational Social Science (**IC2S2**), 2024. *Oral*.
- [W.2] **Yanbang Wang**, Jon Kleinberg. On the Relationship Between Relevance and Conflict in Online Social Link Recommendations. International Conference on Computational Social Science (**IC2S2**), 2024.
- [W.3] Haoteng Yin, **Yanbang Wang**, Pan Li. Revisiting Graph Neural Networks and Distance Encoding in a Practical View. Deep Learning on Graphs: Method and Applications (**AAAI-DLG**), 2021.
- [W.4] **Yanbang Wang**, Pan Li, Chongyang Bai, VS Subrahmanian, Jure Leskovec. Generic Representation Learning for Dynamic Social Interaction. 19th International Workshop on Mining and Learning with Graphs (**KDD-MLG**), 2020.
- [W.5] **Yanbang Wang**, Bijia Chen, Cameron Campbell, A Network-based Method for Estimating Potential for Career Advancement from Incomplete Data. Social Science and History Association Annual Meeting (**SSHA**), 2020.

PATENTS

- [PA.1] Karl Hallgren, **Yanbang Wang**, Jonathan Larson. [Network Authentication Evaluation](#). United States Patent and Trademark Office, Patent Application No. US20250337760A1, 30 Oct. 2025.

TECHNICAL SKILLS

Languages	Python, C++
ML / LLM	PyTorch, JAX, Numpy, PEFT, SFT, RLVR
System / Tools	Google-internal tools & workflows all suite (XManager, Fig/Piper, CNS, Critique, Cider, Jetski/Gemini Coder etc.), Google Cloud Platforms, Vertex AI Model Endpoint Deployment

ACADEMIC SERVICE

GENERAL CHAIR

- 2025 [Learning on Graphs Conference](#) (LoG 2025)
- 2026 [Learning on Graphs Conference](#) (LoG 2026)

PROGRAM COMMITTEES / REVIEWER

(ML: Machine Learning; DM: Data Mining; CSS: Computational Social Science)

ML	NeurIPS 2021- 2026, ICLR 2022 - 2025, ICML 2023 - 2026, LoG 2022 - 2026
DM	KDD 2023, SDM 2024, Journal of Big Data
CSS	IC2S2 2024

SELECTED HONORS AND AWARDS

- 2026 Thinking Machine - Tinker Research Grant
- 2025 Vessel AI Research Grant
- 2024 Microsoft Accelerating Foundation Models Research Grant
- 2024 ICLR Student Travel Award
- 2023 NeurIPS Student Travel Award
- 2021 Stanford Graduate with Distinction in Research Honor
- 2019 HKUST Academic Achievement Medal (highest undergraduate honor, top 1%)
- 2019 HKUST Outstanding Graduate

TALKS AND PRESENTATIONS

INVITED TALKS

2026	Google Research, GenAI
2025	Cornell Theory Seminar
2025	Meta AI
2023	Microsoft Research
2021	Tsinghua University, AI-Time Forum
2021	Tianjin University, College of Intelligence and Computing
2020	UIUC, Data Mining Group (DMG)
2018	MIT CSAIL, Any-scale Learning For All Group (ALFA)