

EDUCATION

Princeton University Ph.D. in Computer Science, Advisor: Danqi Chen	Princeton, NJ 2024–Current
Princeton University M.S.E. in Computer Science, GPA: 4.00/4.00 Thesis: “Long-Context Language Modeling with Parallel Context Encoding”, Advised by Danqi Chen.	Princeton, NJ 2023–2024
Princeton University B.S.E. in Computer Science, Highest Honors (<i>summa cum laude</i>), GPA: 3.99/4.00 Thesis: “How to Answer a Question? Rethinking Open-Domain Question Answering with Multi-Type Questions”, Advised by Danqi Chen.	Princeton, NJ 2019–2023

PUBLICATIONS

* denotes equal contribution.

1. Tianyu Gao*, Alexander Wettig*, **Howard Yen**, and Danqi Chen. “How to Train Long-Context Language Models (Effectively)”. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025. [Paper] [Code]
2. **Howard Yen**, Tianyu Gao, Minmin Hou, Ke Ding, Daniel Fleischer, Peter Izsak, Moshe Wasserblat, and Danqi Chen. “HELMET: How to Evaluate Long-Context Language Models Effectively and Thoroughly”. In *Proceedings of The International Conference on Learning Representations (ICLR)*, 2025. [Paper] [Code] [Site]
3. Hongjin Su*, **Howard Yen***, Mengzhou Xia*, Weijia Shi, Niklas Muennighoff, Han-yu Wang, Haisu Liu, Quan Shi, Zachary S. Siegel, Michael Tang, Ruoxi Sun, Jinsung Yoon, Sercan O. Arik, Danqi Chen, Tao Yu. “BRIGHT: A Realistic and Challenging Benchmark for Reasoning-Intensive Retrieval”. In *Proceedings of The International Conference on Learning Representations (ICLR)*, 2025. **Spotlight, top 5.1%**. [Paper] [Code] [Site]
4. Kenneth Enevoldsen, Isaac Chung, Imene Kerboua, . . . , **Howard Yen**, . . . , Orion Weller, Siva Reddy, Niklas Muennighoff. “MMTEB: Massive Multilingual Text Embedding Benchmark”. In *Proceedings of The International Conference on Learning Representations (ICLR)*, 2025. [Paper] [Code]
5. Ryan Liu, **Howard Yen**, Raja Marjeh, Thomas L. Griffiths, and Ranjay Krishna. “Optimizing Interpersonal Communication by Simulating Audiences with Large Language Models”. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2025. [Paper] [Code]
6. Xi Ye, Fangcong Yin, Yinghui He, Joie Zhang, **Howard Yen**, Tianyu Gao, Greg Durrett, and Danqi Chen. “LongProc: Benchmarking Long-Context Language Models on Long Procedural Generation”. In *Proceedings of the Second Conference on Language Modeling*, 2025. [Paper] [Code] [Site]
7. Wuwei Zhang, Fangcong Yin, **Howard Yen**, Danqi Chen, Xi Ye. “Query-Focused Retrieval Heads Improve Long-Context Reasoning and Re-ranking”. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025. [Paper] [Code]
8. Jacqueline He, **Howard Yen***, Margaret Li*, Shuyue Stella Li, Zhiyuan Zeng, Weijia Shi, Yulia Tsvetkov, Danqi Chen, Pang Wei Koh, Luke Zettlemoyer. “Precise Information Control in Long-Form Text Generation”. *Preprint*, 2025. [Paper] [Code]

9. **Howard Yen**, Tianyu Gao, and Danqi Chen. “Long-Context Language Modeling with Parallel Context Encoding”. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024. [Paper] [Code]
10. Tianyu Gao, **Howard Yen**, Jiatong Yu, and Danqi Chen. “Enabling Large Language Models to Generate Text with Citations”. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023. [Paper] [Code]
11. **Howard Yen**, Tianyu Gao, Jinhyuk Lee, and Danqi Chen. “MoQA: Benchmarking Multi-Type Open-Domain Question Answering”. In *Proceedings of the 3rd Workshop on Dialogue and Conversational Question Answering (DialDoc) @ ACL*, 2023. [Paper] [Code]

INDUSTRY EXPERIENCE

Samaya AI PhD ML Research Intern	Mountain View, California Summer 2025
Meta Reality Labs Software Engineering Intern	Seattle, Washington Summer 2022
Facebook AI Applied Research Software Engineering Intern	Menlo Park, California Summer 2021

TEACHING

- **Graduate Teaching Assistant** at Princeton University
Natural Language Processing (COS484) Spring 2024
- **Graduate Teaching Assistant** at Princeton University
Introduction to Machine Learning (COS324) Fall 2023
- **Research Instructor** at Princeton University
Princeton AI4ALL Summer Camp Summer 2023
- **Undergraduate Course Assistant** at Princeton University
Natural Language Processing (COS484) Spring 2022, Spring 2023
- **Undergraduate Course Assistant** at Princeton University
Algorithms and Data Structures (COS226) Spring 2020 – Fall 2022

HONORS AND AWARDS

- The William A. Dippel '50 *55 Graduate Fellowship 2025
- Tau Beta Pi 2023
- Sigma Xi 2023
- Sigma Xi Book Award 2023
- Phi Beta Kappa 2022–2023
- Outstanding Student Teaching Award 2023
- International Collegiate Programming Contest (ICPC) North America Finalist 2021
- Shapiro Prize for Academic Excellence 2021
- Citadel Terminal Live 2nd Place 2020
- North Dallas Toyota Scholarship 2019

INVITED TALK

- NVIDIA, “HELMET: How to Evaluate Long-Context Language Models Effectively and Thoroughly” 2024/10
- Sierra.AI, “Enabling Large Language Models to Generate Text with Citations” 2023/08

PROFESSIONAL SERVICES

Organizer

- ICML 2025 Workshop on Long Context Foundation Models (ICFM).

Reviewer

- Conference on Language Modeling (COLM): 2025.
- International Conference on Learning Representations (ICLR): 2025.
- ACL Rolling Review (ARR): February 2025.
- ICML 2024 Workshop on Long Context Foundation Models (ICFM).