Howard Yen

Website: howard-yen.github.io
Email: hyen@princeton.edu
LinkedIn: howard-yen
GitHub: github.com/howard-yen

Education

Princeton University — Princeton, NJ
M.S.E. in Computer Science, GPA: 4.00/4.00, Advisor: Danqi Chen — 2023–2024

Princeton University — Princeton, NJ
B.S.E. in Computer Science, Highest Honors (*summa cum laude*), GPA: 3.99/4.00 — 2019–2023
Thesis: "How to Answer a Question? Rethinking Open-Domain Question Answering with Multi-Type Questions", Advised by Danqi Chen.

- **Relevant Courses**: Advanced Topics: Understanding Large Language Models (graduate level), Advanced Topics: Systems and Machine Learning (graduate level), Advanced Computer Vision (graduate level), Advanced Topics: Embodied Natural Language Understanding (graduate level), Natural Language Processing

Publications

1. **Howard Yen**, Tianyu Gao, and Danqi Chen. "Infinite Context Extension for LLM via Cross Attention". Preprint, to be submitted to ACL 2024. Available upon request.
2. Ryan Liu, **Howard Yen**, Raja Marjieh, Thomas L. Griffiths, and Ranjay Krishna. "Optimizing Interpersonal Communication by Simulating Audiences with Large Language Models". Preprint, in submission to the Twelfth International Conference on Learning Representations (ICLR 2024). [Paper] [Code]
3. Tianyu Gao, **Howard Yen**, Jiatong Yu, and Danqi Chen. "Enabling Large Language Models to Generate Text with Citations". In Proc. of The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023). [Paper] [Code]
4. **Howard Yen**, Tianyu Gao, Jinhyuk Lee, and Danqi Chen. "MoQA: Benchmarking Multi-Type Open-Domain Question Answering". In Proc. of the 3rd Workshop on Dialogue and Conversational Question Answering (DialDoc @ ACL 2023). [Paper] [Code]

Experience

Princeton Natural Language Processing Group — Princeton, New Jersey
Research Assistant, Advised by Danqi Chen — Spring 2021–Current

- From Sept. 2021 - Dec. 2022, I led a research project that which involved benchmarking open-domain question answering systems on multi-type questions.
- From Jan. 2023 - May 2023, I assisted on a research project, which involved enabling large language models to generate text with citations.
- From July 2023 - Current, I am leading a research project to augment large language models with light-weight cross attention modules to enable their access to non-parametric knowledge.

Meta Reality Labs — Seattle, Washington
Software Engineering Intern — Summer 2022

- Improved automatic speech recognition generalization with semantic-aware speech augmentation
- Investigated the effect of augmentation techniques such as pauses, word duplication, and semantic-aware phrase replacement on training end-to-end automatic speech recognition and natural language understaning models. Our method achieved up to a 1% improvement on the Spoken Task Oriented Parsing (STOP) dataset.

Facebook AI Applied Research Menlo Park, California
Software Engineering Intern Summer 2021

- Generalization of gradient approximation algorithms on downstream tasks
- Analyzed the generalization ability of gradient approximation algorithms such as FetchSGD for CV and NLP downstream tasks. We achieved more than 80% reduction in communication costs with less than 5% performance drop on CIFAR10, CelebA, and Sent140.

Teaching

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

Scholarships and Awards

- 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

- Sierra.AI, "Enabling Large Language Models to Generate Text with Citations" 8/8/2023

Extracurricular Activities

- Vice Chair at Association for Computing Machinery (ACM) at Princeton University 2019–2023
  Practice problem-solving skills and algorithm and data structure optimization through competitions like ICPC. Host the annual Princeton Computer Science Contest: planning logistics, contacting sponsors, and writing problems.
- Design and Development Team Member at Research Innovation Design 2020–2022
  Conduct user interviews on the course selection process for college students to find ways to improve students' course selection experiences and develop a web app that integrates calendar and course reviews in ReactJS. Continuously iterate through designs using Figma to incorporate user feedback.