Ethan **Chau**

echau18@cs.uw.edu
Web: echau18.gitlab.io | LinkedIn: echau18 | GitHub: ethch18

Experience

Microsoft | Senior Applied Scientist

Jan. 2022 – Present | Redmond, WA
Senior Applied Scientist (Sept. 2023 – Present)
Applied Scientist 2 (Sept. 2022 – Aug. 2023)
Applied Scientist (Jan. 2022 – Aug. 2022)

- Building Bing's core **dense retrieval** model that encodes search queries into vectors to find relevant webpages, delivering consistent **+2-5% weighted recall** improvements in every release.
- Integrating state-of-the-art **NLP/deep learning** techniques in noisy-label scenarios, including for **distillation** and long-document modeling.
- Developing the critical integration between science and platform teams that maintains data freshness.
- Co-mentored four interns with MSR in experimenting and publishing [1][2][3].
- Coordinated across platform, hardware, and science teams to lead an organization-wide infrastructure migration that increased **training capacity 2x**.
- Greatly exceeded expectations in every annual performance review.

University of Washington | NLP Researcher

Mar. 2018 – Nov. 2021 | Seattle, WA | Advisor: Prof. Noah A. Smith

- Led the design of informed training methods for **low-resource** contextual word **representation**, with up to **39%** error reduction on Latin-script parsing evaluations [5] and a **57%** reduction on non-Latin part-of-speech tagging [4].

University of Washington | Teaching Assistant (ML, NLP)

Jan. 2019 – Mar. 2019 and Sept. 2020 – June 2021 | Seattle, WA

- Developed a new **PyTorch-** and AllenNLP-based text classification assignment.
- Published a comprehensive, officially featured tutorial for AllenNLP. (link)

Facebook | Software Engineering Intern (Machine Learning)

June 2020 – Sept. 2020 | Seattle, WA

- Drove the development of a configurable, E2E system for content **retrieval** and recommendation in **C++** and **Hack**, increasing engagement by **2.5%.**
- Designed efficient data-collection pipelines for **search index** construction.
- Contributed to the **PyTorch** open-source project.
- Achieved highest possible rating ("**Rockstar**") on final performance evaluation.

Facebook | Software Engineering Intern (Backend)

June 2019 – Sept. 2019 |Menlo Park, CA

- Architected a modular, interpretable, and configurable suggestion system in **Hack** with a **37%** higher success rate than previous versions.

Selected Publications

[1] G. Qin, C. Rosset, **E. C. Chau**, N. Rao, and B. Van Durme. Dodo: Dynamic Contextual Compression for Decoder-only LMs. In *Proc. Of ACL (to appear)*. (link)

[2] C. Xu, C. Rosset, **E. C. Chau**, et al. 2024. Automatic Pair Construction for Contrastive Post-training. In *Findings of ACL: NAACL*. (link)

[3] C. Rosset, H.-L. Chung, G. Qin, **E. C. Chau**, et al. 2024. Researchy Questions: A Dataset of Multi-Perspective, Decompositional Questions for LLM Web Agents. arXiv:2402.17896 [cs.CL]. (link)

[4] **E. C. Chau** and N. A. Smith. 2021. Specializing Multilingual Language Models: An Empirical Study. In *Workshop on Multilingual Representation Learning (EMNLP)*. (**Best Paper Honorable Mention**) (link)

[5] **E. C. Chau**, L. H. Lin, and N. A. Smith. 2020. Parsing with Multilingual BERT, a Small Corpus, and a Small Treebank. In *Findings of ACL: EMNLP*. (link)

Education

University of Washington

M.S., Computer Science
June 2021 | Seattle, WA | GPA: 4.00
Thesis: *Specializing Multilingual Language Models: An Empirical Study*
B.S. with Distinction,
Computer Science (Data Science)
B.A., Linguistics
June 2020 | Seattle, WA | GPA: 3.98
Thesis: *Towards Resource-Efficient Contextual Word Representations for Parsing*

ETH Zürich

Exchange Student, Computer Science
Sept. 2019 – Feb. 2020 | Zürich, CH

Coursework

Natural Language Processing
NLP Capstone
Advanced Machine Learning
Artificial Intelligence
Reliable and Interpretable AI
Generative Models • Graphical Models
Information Retrieval
Quantum Computing
Data Structures • Algorithms
Operating Systems • Security

Skills

Technologies

Proficient
Java, Git, NumPy, Python, PyTorch, SQL, Transformers
Familiar
Bash, C, C++, C#, Google Protobuf, Hack, HTML/CSS, JavaScript, Mercurial, React, Ruby, SCSS, Thrift

Languages

English (Native), Mandarin (Fluent), Cantonese (Proficient), Spanish (Intermediate), German (Beginner)

Service

Young Student Fellowship

High School Mentor
Sept. 2022 – Present