Ethan C. Chau

E-mail: echau18@cs.uw.edu | Web: echau18.gitlab.io | LinkedIn: echau18

Education

University of Washington, Seattle, WA

M.S., Computer Science **Sept. 2020 - June 2021**
Thesis: *Specializing Multilingual Language Models: An Empirical Study*
Advisor: Noah A. Smith
GPA: 4.00
Selected Coursework: Generative Models, Graphical Models, Quantum Computing, Convex Optimization

B.S. with Distinction, Computer Science (Data Science) **Sept. 2016 - June 2020**

B.A., Linguistics **Sept. 2016 - June 2020**
Thesis: *Towards Resource-Efficient Contextual Word Representations for Parsing*
Advisor: Noah A. Smith
GPA: 3.98
Selected Coursework: Natural Language Processing, NLP Capstone, Machine Learning, Artificial Intelligence, Algorithms, Data Structures & Parallelism, Operating Systems, Computer Security

Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

Exchange Student, Computer Science **Sept. 2019 - Feb. 2020**
Selected Coursework: Advanced Machine Learning (graduate), Information Retrieval, Reliable & Interpretable AI (graduate)

Academic Experience

Paul G. Allen School of Computer Science & Engineering, Seattle, WA

NLP Researcher **Mar. 2018 - Nov. 2021**
I investigated methods for enhancing the performance of language systems for which data is scarce and conventional methods are ineffective.

- Led the design of informed model training methods for low-resource contextual word representation, with up to 39% error reduction on Latin-script dependency parsing evaluations [1] and a 57% reduction on non-Latin part-of-speech tagging [2]
- Investigated efficient methods for machine translation, data-driven approaches to modeling lexical borrowing, and transfer learning across semi-similar domains
- Advisor: Professor Noah A. Smith

Teaching Assistant **Jan. 2019 - Mar. 2019 and Sept. 2020 - Jun. 2021**
CSE 446/546 (Machine Learning), Spring 2021 and Fall 2020
CSE 447/517 (Natural Language Processing), Winter 2021
CSE 447/547M (Natural Language Processing), Winter 2019

- Developed and administered a new PyTorch- and AllenNLP-based text classification assignment
- Planned and led a weekly discussion section and office hours for 35 students
- Published a comprehensive, officially featured tutorial for AllenNLP (link)

Industry Experience

Microsoft, Redmond, WA

Senior Applied Scientist **Sept. 2023 - Present**
Applied Scientist 2 **Sept. 2022 - Aug. 2023**
Applied Scientist **Jan. 2022 - Aug. 2022**
I conduct applied research on dense retrieval for web search and develop methods and models for improving retrieval quality. I also work closely with engineering partners to ensure high-quality

experimentation environments for the whole team.

- Building Bing's core dense retrieval model that encodes queries into vectors for web search, delivering consistent +2-6% weighted recall improvements in every release through state-of-the-art NLP/deep learning techniques
- Shipped Bing's first multi-vector retriever with +6% weighted recall over single-vector models and 96% less storage than naïve multi-vector models
- Co-mentored four interns with MSR in experimenting and publishing [3] [4] [5]
- Drove end-to-end studies using LLMs to enhance inverted-index retrieval
- Coordinated across platform, hardware, and science teams to lead an organization-wide infrastructure migration that doubled training capacity

Facebook, Seattle, WA and Menlo Park, CA

Software Engineering Intern (Machine Learning) **June 2020 - Sept. 2020**
As an individual contributor to the Search-Discovery team, I developed an end-to-end system for recommending content while fine-tuning the machine learning algorithms used to rank suggestions.

- 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
- Investigated ML algorithms for content representation and achieved a 75% average relative error reduction
- Contributed to the PyTorch open-source project
- Achieved highest possible rating ("Rockstar") in final performance evaluation

Software Engineering Intern (Backend) **June 2019 - Sept. 2019**
I built a high-performance suggestion system to underpin a suite of internal tools and designed a machine learning pipeline to improve its accuracy.

- Architected a modular, interpretable, and configurable suggestion system in Hack with a 37% higher success rate than previous versions, while maintaining an average latency of 0.5 seconds
- Engineered machine learning features that attained 80% accuracy and F1 on the task
- Developed a multi-stage model training pipeline tailored to the application

Indeed.com, Seattle, WA

Software Engineering Intern (Backend/Data) **June 2018 - Aug. 2018**
I augmented a candidate screening platform with a natural language processing framework and developed a system that reliably integrated a machine learning model into the platform.

- Integrated an NLP library into a backend suggestion service, yielding 6x the number of acceptances and reaching 2x as many users as the next best suggestion source
- Architected and developed a Java-Protobuf service for machine learning evaluation that decreased required requests by 5x, increased logging coverage by 15x, and enabled model onboarding in 5 lines of code
- Deployed a Jenkins- and Mesos-based continuous integration system

Software Engineering Intern (Full Stack) **June 2017 - Sept. 2017**
I implemented a hotly requested feature for the My Indeed product, which involved building a brand new web interface and restructuring how user data was stored and accessed.

- Build and stylized a custom React frontend to minimize server load
- Implemented an extensible Java data storage and retrieval model
- Developed a set of REST API controllers with a test coverage of over 90%

Publications

[1] **Ethan C. Chau**, Lucy H. Lin, and Noah A. Smith. Parsing with Multilingual BERT, a Small Corpus, and a Small Treebank. In *Findings of ACL: EMNLP*, 2020. (link).

[2] **Ethan C. Chau** and Noah A. Smith. Specializing Multilingual Language Models: An Empirical Study. In *Workshop on Multilingual Representation Learning*, 2021. **(Best Paper Honorable Mention).** (link).

[3] Guanghui Qin, Corby Rosset, **Ethan C. Chau**, Nikhil Rao, and Benjamin Van Durme. Dodo: Dynamic Contextual Compression for Decoder-only LMs. In *Proc. of ACL*, 2024. (link).

[4] Canwen Xu, Corby Rosset, **Ethan C. Chau**, Luciano Corro, Shweti Majahan, Julian McAuley, Jennifer Neville, Ahmed Awadallah, and Nikhil Rao. Automatic Pair Construction for Constructive Post-training. In *Findings of ACL: NAACL*, 2024. (link).

[5] Corbin Rosset, Ho-Lam Chung, Guanghui Qin, **Ethan C. Chau**, Zhuo Feng, Ahmed Awadallah, Jennifer Neville, and Nikhil Rao. Researchy Questions: A Dataset of Multi-Perspective, Decompositional Questions for Deep Research. In *Proc. of SIGIR*, 2025. (link).

Leadership and Community Service

Young Student Fellowship, Redmond, WA

High School Mentor **Sept. 2022 - Present**

As a High School Mentor, I oversee small- and large-scale programming for a group of high schoolers while providing one-on-one and small group mentorship to them.

- Facilitating curricula and leading discussions for a weekly small group session attended by 12 students
- Mentoring multiple students in the organization's band
- Developing large-group programs for semi-monthly general meetings
- Organizing activities, scheduling, and student leadership for an annual weekend-long retreat attended by 80 people
- Coordinating with other department leaders to ensure continuity between programs

Common Ground Fellowship, Seattle, WA

Discussion Leader **June 2020 - Aug. 2021**

Core Team Leader **Aug. 2017 - Aug. 2018**

Common Ground Fellowship is a registered student organization at UW. As part of an 8-person leadership team, I oversaw programming for the entire academic year. I founded a weekly presentation and discussion group with an average attendance of 15 people.

- Coordinated, facilitated, and researched presentation materials for a weekly discussion group attended by 15 people
- Planned agendas and prepared study materials for weekly general meetings attended by over 40 people
- Organized weekly transportation logistics between two college campuses
- Coordinated and oversaw workshops, activities, and scheduling for a weekend-long retreat attended by 50 people

Evangelical Chinese Church of Seattle, Redmond, WA and New Taipei City, Taiwan

Taiwan Missions Youth Leader **Jan. 2016 - Sept. 2018**

The Taiwan Missions program sends American youth to a rural village in Taiwan to teach English to and provide after-school enrichment for junior high school students. As a fluent Mandarin speaker, I served as one of the head teachers for my class.

- Planned and taught daily English classes to groups of 20+ under-resourced youth
- Tutored troubled students 1-on-1, markedly increasing content retention rates
- Facilitated daily re-evaluations about content and methods with fellow teachers

AWANA Youth Leader **Sept. 2013 - May 2016**

The AWANA program engages school-aged children in activities to develop their reading skills and

build character. As the most senior youth leader, I directly led a group of 20 students while also overseeing the entire 80-person age group.

- Guided elementary-age students through literacy-building activities
- Partnered with program directors to prepare session materials
- Supervised and mentored incoming youth leaders

Honors and Awards

Best Paper Honorable Mention, Workshop on Multilingual Representation Learning (2021)
- Listed among the top three papers out of 19 accepted works, with a 38% overall acceptance rate

Winner, Outstanding Computer Science Senior Award (2020)
- Awarded to four graduating students with the most distinguished records in coursework, research, and extracurricular activities

Nominee, UW College of Arts and Sciences Dean's Award, Computer Science (2020)
- Nominated based on excellence in Computer Science coursework and research

Nominee, University of Washington President's Medal (2020)
- Nominated based on distinguished academic record (*Summa Cum Lauda*, GPA > 3.97)

Phi Beta Kappa (2018)
- Elected into a national honor society based on academic excellence (GPA > 3.83)

3rd Place, Google Games Seattle (2017)
- Placed 3rd out of nearly 50 teams in a regional collegiate problem solving competition

Leo W. and Alberta V. Thomas Utz Scholarship (2016)
- Awarded to 1 student annually for excellence in mathematics

Alpha Delta Kappa Scholarship (2016)
- Awarded to 2-3 students annually for excellence in academic endeavors and community service

Kiwanis Club of Issaquah Leadership Scholarship (2016)
- Awarded to 1 student annually for excellence in community service, leadership activities, and academic performance

Theta Xi Franklin Scholarship (2016)
- Awarded to 3 students annually for excellence in intellectual curiosity and dedication to community

Selected Personal Projects

Where's My Bus?
- Technologies used: Python, PyTorch

Full-Stack Web App **Mar. 2018 - Present**

Designed and deployed a lightweight transit tracking website for the Seattle Area.
Available at https://echau18.gitlab.io/obaview

- Technologies used: Python, Bottle, Azure App Service, React, Brunch

echau18.gitlab.io

Personal Portfolio **Sept. 2017 - Present**

Built a modular personal portfolio template based on a lightweight data entry format, allowing code-free website generation.
Available at https://echau18.gitlab.io

- Technologies used: React, Brunch, SCSS

EVC Charge

Windows Metro App **May 2015 - Aug. 2016**

Designed, built, and published an app to locate electric vehicle charging stations.

- Technologies used: C#, XAML

Skills

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

Programming Languages
Proficient: Java, JavaScript, Python
Familiar: Bash Shell, C/C++, C#, Google Protobuf, Hack, HTML/CSS, Racket, Ruby, SCSS, SQL, Thrift

Technologies
AllenNLP, Git, Mercurial, NumPy, PyTorch, React.js, Spring Framework

Interests

Academic
Natural Language Processing, Machine Learning, Low-Resource NLP, Machine Translation, Adversarial Robustness, Syntactic Parsing

Extracurricular
Piano, Violin, Guitar, Tennis, Ultimate Frisbee