

## Education

2021– **Ph.D**, *Carnegie Mellon University*, Pittsburgh  
School of Computer Science, Language Technologies Institute  
**Advisor**: Graham Neubig  
**cGPA**: 4.03/4.33  
**Relevant Coursework**: Machine Learning, Advanced Natural Language Processing, Multimodal Machine Learning, Machine Learning with Large Datasets

2016–2021 **B.Sc**, *University of Toronto*, Toronto  
Computer Science Specialist & Cognitive Science Major & Statistics Minor

## Experience

2021– **Research Assistant**, *Neulab*, Carnegie Mellon University

Worked on multiple topics in natural language processing, including training language models, curating datasets, interpretability of models, and model evaluation. Experienced with working with models in huggingface, fairseq, or pytorch for diverse applications like question answering, machine translation, and reasoning.

June 2023–August 2023 **Visiting Researcher**, *Language and Intelligence Group (LINGO)*, Massachusetts Institute of Technology

2023 Worked on a novel form of prompt engineering for reasoning, by leveraging large language models to induce rules from in-context examples and applying rules to improve performance on compositional and reasoning tasks.

2018–2019 **Associate Engineer**, *Wattpad*, Toronto

Contributed large web features and new data science experiments through a year-long internship (through undergrad internship program)

- Independently rebuilt major parts of frontend web app serving 65+ million monthly users, including Home page, tag (explore) pages, and onboarding
- Implemented data science experiments in content recommendations and rankings affecting millions of users

2018–2021 **Research Assistant**, *Cognitive Lexicon Laboratory*, University of Toronto

Worked on computational linguistics research, especially efficient communication.

2020–2021 **Research Assistant**, *Intelligent Adaptive Interventions Laboratory*, University of Toronto

Worked on human-computer interaction research, especially adaptive interventions with bandits.

June 2020–Aug 2020 **Research Assistant**, *Computational Psycholinguistics Laboratory*, MIT

2020–Aug 2020 Worked on examining gender bias with language models.

## Awards

- (2023) Best Resource Paper, ACL 2023
- (2023 – 2024) NEC Fellowship
- (2023 – 2026) NSERC PGS-D Fellowship
- (2021 – 2022) CMU ASA Presidential Fellowship
- (2021) NSERC USRA

## Skills

- Languages: Python, Javascript/Typescript, R, L<sup>A</sup>T<sub>E</sub>X
- Frameworks/tools: Pytorch, Numpy/Pandas, Deepspeed, Slurm, React + Redux, Pyspark, anything else as needed
- General skills: Model training and fine-tuning, interpretability of language models, benchmark/dataset construction, technical writing

## Publications

### Journal Articles

- [1] Patrick Fernandes, Aman Madaan, **Emmy Liu**, et al. “Bridging the Gap: A Survey on Integrating (Human) Feedback for Natural Language Generation”. In: *Transactions of the Association for Computational Linguistics* (2023).
- [13] **Emmy Liu** and Yang Xu. “Regularity in the Temporal Dynamics of Publishing Behaviour across Scientists”. In: *PLoS One* (2022).
- [18] Yang Xu\*, **Emmy Liu**\*, and Terry Regier. “Numeral Systems across Languages Support Efficient Communication: From Approximate Numerosity to Recursion”. In: *Open Mind* 4 (2020), pp. 57–70.

### Conference Articles

- [2] Patrick Fernandes\*, Kayo Yin\*, **Emmy Liu**, and Graham Neubig. “When Does Translation Require Context? A Data-driven, Multilingual Exploration (**Best Resource Paper**)”. In: Toronto, Canada: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023.
- [3] Anubha Kabra\*, **Emmy Liu**\*, Simran Khanuja\*, Alham Fikri Aji, Genta Indra Winata, Samuel Cahyawijaya, Anuoluwapo Aremu, Perez Ogayo, and Graham Neubig. “Multilingual and Multi-cultural Figurative Language Understanding”. In: Toronto, Canada: Findings of the Association for Computational Linguistics: ACL 2023, 2023.
- [4] Andy Liu, Hao Zhu, **Emmy Liu**, Yonatan Bisk, and Graham Neubig. “Computational Language Acquisition with Theory of Mind”. In: Kigali, Rwanda: Proceedings of the Eleventh International Conference on Learning Representations (ICLR), 2023.
- [5] Emmy Liu, Aditi Chaudhary, and Graham Neubig. “Crossing the Threshold: Idiomatic Machine Translation through Retrieval Augmentation and Loss Weighting”. In: Singapore: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

- [7] Mohi Reza, Angela Zavaleta Bernuy, **Emmy Liu**, Tong Li, Calista Barber, Zhongyuan Liang, and Joseph Jay Williams. “Exam Eustress: Designing a Brief Online Intervention for Helping Students Identify Positive Aspects of Stress”. In: Hamburg, Germany: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), 2023.
- [8] Lindia Tjuatja, **Emmy Liu**, Lori Levin, and Graham Neubig. “Syntax and Semantics Meet in the ”Middle“: Probing the Syntax-Semantics Interface of LMs Through Agentivity”. In: Toronto, Canada: Proceedings of the 12th Joint Conference on Lexical and Computational Semantics, 2023.
- [10] **Emmy Liu**, Chenxuan Cui, Kenneth Zheng, and Graham Neubig. “Testing the Ability of Language Models to Interpret Figurative Language”. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. Seattle, United States: Association for Computational Linguistics, July 2022, pp. 4437–4452. DOI: 10.18653/v1/2022.naacl-main.330. URL: <https://aclanthology.org/2022.naacl-main.330>. (**Oral**).
- [11] **Emmy Liu** and Graham Neubig. “Are Representations Built from the Ground Up? An Empirical Examination of Local Composition in Language Models”. In: Abu Dhabi, UAE: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), Dec. 2022.
- [15] Aida Ramezani\*, **Emmy Liu**\*, Jr. Renato Ferreira Pinto, Spike W.S. Lee, and Yang Xu. “The Emergence of Moral Foundations in Child Language Development”. In: vol. 44. Toronto, Canada: Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci), July 2022. (**Oral**).
- [17] **Emmy Liu** and Yang Xu. “Chaining and the Process of Scientific Innovation”. In: vol. 42. Toronto, Canada: Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci), 2020, pp. 1639–1646.
- [20] **Emmy Liu** and Yang Xu. “Rapid information gain explains cross-linguistic tendencies in numeral ordering”. In: vol. 41. Montreal, Canada: Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci), 2019, pp. 2166–2173.

#### Workshops & Posters

- [12] **Emmy Liu**, Michael Henry Tessler, Nicole Dubosh, Katherine Hiller, and Roger Levy. “Assessing Group-level Gender Bias in Professional Evaluations: The Case of Medical Student End-of-Shift Feedback”. In: *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*. Seattle, United States: Association for Computational Linguistics, 2022, pp. 86–93.
- [14] Sedrick Scott Keh, Rohit K. Bharadwaj, **Emmy Liu**, Simone Tedeschi, Varun Gangal, and Roberto Navigli. “EUREKA: EUphemism Recognition Enhanced through kNN-based methods and Augmentation (**First place in shared task**)”. In: *The Third Workshop on Figurative Language Processing @ EMNLP 2022*. Abu Dhabi, UAE, 2022.
- [16] Nayan Saxena, Pan Chen, and **Emmy Liu**. “Statistical Consequences of Duelling Bandits”. In: *EDM21 Workshop on Reinforcement Learning for Education*. Virtual, 2021. (**Spotlight**).
- [19] **Emmy Liu** and Yang Xu. “Ordering in numerals supports rapid information processing”. In: *32nd CUNY Conference on Human Sentence Processing*. Boulder, United States, 2019.

## Manuscripts & In Submission

- [6] Aida Ramezani, **Emmy Liu**, Spike W.S. Lee, and Yang Xu. “Quantifying the emergence of moral foundational lexicon in child language development”. in submission. 2023.
- [9] Yuchen Zhou, **Emmy Liu**, Graham Neubig, and Leila Wehbe. “Discovering Divergences between Language Models and Human Brains”. in submission. 2023.

## Service

- (2023) 4th Workshop on Figurative Language Processing @ NAACL 2024, **Workshop Organizer**
- (2022 – ) LTI Mentoring Program, **Mentor**
- (2021 – ) SCS PhD Dean’s Advisory Committee, **Social Working Group**: Organizing social events for PhD students
- (2021) CMU Graduate Application Support, **Mentor**: Mentoring applicants from underrepresented groups
- (2019 – 2021) **Chief Editor**, Review of Undergraduate Computer Science: Reviewing papers for undergrad CS journal and holding research events
- **Reviewer/Program Committee**, ACL (2023), EMNLP (2022, 2023), CogSci (2020, 2021), {TL4NLP, MemARI} Workshop @ Neurips (2022)

## Teaching and Mentoring

### Teaching

- Fall 2023 CMU 11-667 (Large Language Models Methods and Applications), Teaching Assistant
- Fall 2022 CMU 11-711 (Advanced NLP), Teaching Assistant

### Research Mentorship

- **Anubha Kabra**
  - Multi-lingual and Multi-cultural Figurative Language Understanding (ACL Findings 2023)
  - Code Calibrates: Investigating the Edge of Program-Aided Language Models for Reasoning (draft)
- (Course Project) **Liangze Li, Shih-Lun Wu, Yi-Hui Chou**
  - Listener Model for the PhotoBook Referential Game with CLIP Scores as Implicit Reference Chain (ACL 2023)
- **Andy Liu**
  - Computational Language Acquisition with Theory of Mind (ICLR 2023)
- **Nayan Saxena, Pan Chen**
  - Statistical Consequences of Duelling Bandits (RL4ED 2021)