Nathan Ng

nathanng@mit.edu
https://nng555.github.io

Research Interests

My general research interest is in developing methods to learn efficient and robust representations of discrete-valued sequence data (specifically natural language). I'm also interested in understanding and finding ways to fix their pathologies once trained. My research broadly covers topics in representation learning, interpretability, and out-of-domain robustness/generalization.

Education

Massachusetts Institute of Technology — Cambridge, MA
Visiting Scholar — *Sept 2021 – June 2024*
Advisor: Prof. Marzyeh Ghassemi

University of Toronto — Toronto, Ontario
Ph.D. Machine Learning — *Sept 2019 – June 2024*
Advisor: Prof. Marzyeh Ghassemi

University of California San Diego — San Diego, California
BS Computer Science (Summa Cum Laude) — *Sep 2014 – Jun 2018*
Advisor: Prof. Zachary Lipton and Prof. Julian McAuley

Professional Experience

Prescient Design — New York, New York
Research Intern (Kyunghyun Cho) — *Summer 2022*
Blind Biological Sequence Denoising with Self-Supervised Set Learning

Meta — New York, New York (Virtual)
Research Intern (Naman Goyal) — *Summer 2021*
Growing Switch Transformers for Multilinguality

Google — Mountain View, California (Virtual)
Research Intern (Qi Guo) — *Summer 2020*
Improving Dialogue Breakdown Detection with Semi-Supervised Learning

Meta (Full Time) — Menlo Park, California
Research Engineer (Michael Auli) — *Sep 2018 – Sep 2019*

Meta — Menlo Park, California
Software Engineering Intern — *Summer 2016 / Summer 2017*

Qualcomm — San Diego, California
Software Engineering Intern — *Summer 2015*

Refereed Publications

1. **N. Ng**, R. Grosse, and M. Ghassemi. "Measuring Stochastic Data Complexity with Boltzmann Influence Functions". In: *ICML (In Review)*. 2024.
2. **N. Ng**, J. W. Park, J. H. Lee, R. Kelly, S. Ra, and K. Cho. "Blind Biological Sequence Denoising with Self-Supervised Set Learning". In: *TMLR*. 2024.
3. K. O'Brien, **N. Ng**, I. Puri, J. Mendez, H. Palangi, Y. Kim, M. Ghassemi, and T. Hartvigsen. "Improving Black-box Robustness with In-Context Rewriting". In: *TMLR (In Review)*. 2024.
4. **N. Ng**, N. Hulkund, K. Cho, and M. Ghassemi. "Predicting Out-of-Domain Generalization with Neighborhood Invariance". In: *TMLR*. 2023.
5. J. Bae, **N. Ng**, A. Lo, M. Ghassemi, and R. Grosse. "If Influence Functions are the Question, What is the Answer?" In: *Proc. of NeurIPS*. 2022.

6. **N. Ng**, K. Cho, and M. Ghassemi. "SSMBA: Self-Supervised Manifold Based Data Augmentation for Improving Out-of-Domain Robustness". In: *Proc. of EMNLP*. 2020.

7. T. Lau, **N. Ng**, J. Gingold, N. Desai, J. McAuley, and Z. C. Lipton. "Embryo staging with weakly-supervised region selection and dynamically-decoded predictions". In: *Proc. of Machine Learning for Healthcare*. 2019.

8. **N. Ng**, K. Yee, A. Baevski, M. Ott, M. Auli, and S. Edunov. "Facebook FAIR's WMT19 News Translation Task Submission". In: *Proc. of WMT*. 2019.

9. K. Yee, **N. Ng**, Y. Dauphin, and M. Auli. "Simple and Effective Noisy Channel Modeling for Neural Machine Translation". In: *Proc. of EMNLP*. 2019.

10. **N. Ng**, R. Gabriel, J. McAuley, C. Elkan, and Z. Lipton. "Predicting surgery duration with neural heteroscedastic regression". In: *Proc. of Machine Learning for Healthcare*. 2017.

Workshop Publications

1. **N. Ng**, N. Thangarajan, J. Pan, M. Ghassemi, and Q. Guo. "Improving Dialogue Breakdown Detection with Semi-Supervised Learning". In: *Proc. of Workshop on Human in the Loop Dialogue Systems at NeurIPS*. 2020. Oral.

2. M. Ott, S. Edunov, A. Baevski, A. Fan, S. Gross, **N. Ng**, D. Grangier, and M. Auli. "fairseq: A fast, extensible toolkit for sequence modeling". In: *Proc. of NAACL-HLT: Demonstrations*. 2019.

3. **N. Ng**, J. McAuley, Z. Lipton, and N. Desai. "Predicting Embryo Morphokinetics in Videos with Late Fusion Nets & Dynamic Decoders". In: *Proc. of ICLR Workshops*. 2018.

Shared Tasks

1st in Dialogue Breakdown Detection Challenge English task	*2020*
1st in WMT News Translation English ↔ German task	*2019*
1st in WMT News Translation English ↔ Russian task	*2019*

Professional Activities

Chief Organizer

Workshop on Robustness in Sequence Modeling at NeurIPS	*2022*

Reviewer

ICML	*2024*
NeurIPS	*2022, 2023*
ICLR	*2023*
Machine Learning for Healthcare	*2020*

Teaching and Talks

University of Toronto	Teaching Assistant
CSC 2515: Introduction to Machine Learning (Graduate Level)	*Fall 2020*
CSC 2541: Topics in Machine Learning: Machine Learning for Health	*Winter 2020*
CSC 311: Introduction to Machine Learning	*Fall 2019*

Meta	Internal Lecturer
Special Topics in Deep Learning: NLP and Translation	*Feb 2019, Sep 2019*

University of California, San Diego	Teaching Assistant
CSE 101: Design and Analysis of Algorithms	*Winter 2018*
CSE 158: Web Mining and Recommender Systems	*Fall 2017*
CSE 21: Mathematics for Algorithms and Systems	*Winter 2017*
CSE 11: Introduction to Object-Oriented Programming	*Fall 2015*

Honors and Awards

- Jacobs Scholarship, University of California San Diego *2014*
- Regents Scholarship, University of California San Diego *2014*