Peter West

Canadian & American Citizen
2109 N 39 St., Seattle, WA 98103
+1 (206) 790 8030

homes.cs.washington.edu/~pawest
pawest@cs.washington.edu

EDUCATION

Doctor of Philosophy, University of Washington 2017 - Present
Paul G. Allen School of Computer Science & Engineering
Research Areas: Natural Language Processing, Machine Learning
Advisor: Yejin Choi

Bachelor of Science, University of British Columbia 2013 - 2017
Department of Computer Science
Major: Honors Computer Science

PUBLICATIONS & PREPRINTS

The Generative AI Paradox: ”What It Can Create, It May Not Understand”
In Sumbission. 2023
Peter West*, Ximing Lu*, Nouha Dziri*, Faeze Brahman*, Linjie Li*, Jena D. Hwang, Liwei Jiang, Jillian Fisher, Abhilasha Ravichander, Khyathi Chandu, Benjamin Newman, Pang Wei Koh, Allyson Ettinger, Yejin Choi

Value Kaleidoscope: Engaging AI with Pluralistic Human Values, Rights, and Duties
In Sumbission. 2023
Taylor Sorensen, Liwei Jiang, Jena Hwang, Sydney Levine, Valentina Pyatkin, **Peter West**, Nouha Dziri, Ximing Lu, Kavel Rao, Chandra Bhagavatula, Maarten Sap, John Tasioulas, Yejin Choi

Impossible Distillation: from Low-Quality Model to High-Quality Dataset & Model for Summarization and Paraphrasing
In Sumbission. 2023
Jaehun Jung, **Peter West**, Liwei Jiang, Faeze Brahman, Ximing Lu, Jillian Fisher, Taylor Sorensen, Yejin Choi

Generative Models as a Complex Systems Science: How can we make sense of large language model behavior?
Self-published. 2023
Ari Holtzman, **Peter West**, Luke Zettlemoyer

Localized Symbolic Knowledge Distillation for Visual Commonsense Models
Neurips 2023
Jae Sung Park, Jack Hessel, Khyathi Chandu, Paul Pu Liang, Ximing Lu, **Peter West**, Youngjae Yu, Qiuyuan Huang, Jianfeng Gao, Ali Farhadi, Yejin Choi

Faith and Fate: Limits of Transformers on Compositionality
Neurips 2023 (Spotlight)
Nouha Dziri, Ximing Lu, Melanie Sclar, Xiang Lorraine Li, Liwei Jiang, Bill Yuchen Lin, **Peter West**, Chandra Bhagavatula, Ronan Le Bras, Jena D. Hwang, Soumya Sanyal, Sean Welleck, Xiang Ren, Allyson Ettinger, Zaid Harchaoui, Yejin Choi

NovaCOMET: Open Commonsense Foundation Models with Symbolic Knowledge Distillation
Findings of EMNLP 2023
Peter West, Ronan Le Bras, Taylor Sorensen, Bill Yuchen Lin, Liwei Jiang, Ximing Lu, Khyathi Chandu, Jack Hessel, Ashutosh Baheti, Chandra Bhagavatula, Yejin Choi

SODA: Million-scale Dialogue Distillation with Social Commonsense Contextualization
EMNLP 2023 (Outstanding Paper Award for Dialogue and Interactive Systems)
Hyunwoo Kim, Jack Hessel, Liwei Jiang, **Peter West**, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, Yejin Choi

Inference-Time Policy Adapters (IPA): Tailoring Extreme-Scale LMs without Fine-tuning
EMNLP 2023
Ximing Lu, Faeze Brahman, **Peter West**, Jaehun Jung, Khyathi Chandu, Abhilasha Ravichander, Lianhui Qin, Prithviraj Ammanabrolu, Liwei Jiang, Sahana Ramnath, Nouha Dziri, Jillian Fisher, Bill Yuchen Lin, Skyler Hallinan, Xiang Ren, Sean Welleck, Yejin Choi

We're Afraid Language Models Aren't Modeling Ambiguity
EMNLP 2023
Alisa Liu, Zhaofeng Wu, Julian Michael, Alane Suhr, **Peter West**, Alexander Koller, Swabha Swayamdipta, Noah A Smith, Yejin Choi

Minding Language Models'(Lack of) Theory of Mind: A Plug-and-Play Multi-Character Belief Tracker
ACL 2023 (Outstanding Paper Award)
Melanie Sclar, Sachin Kumar, **Peter West**, Alane Suhr, Yejin Choi, Yulia Tsvetkov

I2D2: Inductive Knowledge Distillation with NeuroLogic and Self-Imitation
ACL 2023
Chandra Bhagavatula, Jena D. Hwang, Doug Downey, Ronan Le Bras, Ximing Lu, Lianhui Qin, Keisuke Sakaguchi, Swabha Swayamdipta, **Peter West**, Yejin Choi

Generating Sequences by Learning to Self-Correct
ICLR 2023
Sean Welleck*, Ximing Lu*, **Peter West****, Faeze Brahman**, Tianxiao Shen, Daniel Khashabi, Yejin Choi

Quark: Controllable Text Generation with Reinforced Unlearning
Neurips 2022
Ximing Lu, Sean Welleck, Jack Hessel, Liwei Jiang, Lianhui Qin, **Peter West**, Prithviraj Ammanabrolu, Yejin Choi

Referee: Reference-Free Sentence Summarization with Sharper Controllability through Symbolic Knowledge Distillation
EMNLP 2022
Melanie Sclar, **Peter West**, Sachin Kumar, Yulia Tsvetkov, Yejin Choi

Symbolic knowledge distillation: from general language models to commonsense models
NAACL 2022
Peter West, Chandra Bhagavatula, Jack Hessel, Jena D Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck, Yejin Choi

NeuroLogic A*esque Decoding: Constrained Text Generation with Lookahead Heuristics
NAACL 2022 (Best Method Paper Award)
Ximing Lu, Sean Welleck**, Peter West**, Liwei Jiang, Jungo Kasai, Daniel Khashabi, Ronan Le Bras, Lianhui Qin, Youngjae Yu, Rowan Zellers, Noah A Smith, Yejin Choi

Probing Factually Grounded Content Transfer with Factual Ablation
Findings of ACL 2022
Peter West, Chris Quirk, Michel Galley, Yejin choi

Generated Knowledge Prompting for Commonsense Reasoning
ACL 2022
Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, **Peter West**, Ronan Le Bras, Yejin Choi, Hannaneh Hajishirzi

Symbolic Brittleness in Sequence Models: on Systematic Generalization in Symbolic Mathematics
AAAI 2021
Sean Welleck, **Peter West**, Jize Cao, Yejin Choi

Surface Form Competition: Why the Highest Probability Answer Isn't Always Right
EMNLP 2021
Ari Holtzman*, **Peter West***, Vered Shwartz, Yejin Choi, and Luke Zettlemoyer

Reflective Decoding: Unsupervised Paraphrasing and Abductive Reasoning
ACL 2021
Peter West, Ximing Lu, Ari Holtzman, Chandra Bhagavatula, Jena Hwang, and Yejin Choi

Neurologic Decoding: (Un)supervised Neural Text Generation with Predicate Logic Constraints
NAACL 2021
Ximing Lu, **Peter West**, Rowan Zellers, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi

Unsupervised Commonsense Question Answering with Self-Talk
EMNLP 2020
Vered Shwartz, **Peter West**, Ronan Le Bras, Chandra Bhagavatula, Yejin Choi

Back to the Future: Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning
EMNLP 2020
Lianhui Qin, Vered Shwartz, **Peter West**, Chandra Bhagavatula, Jena Hwang, Ronan Le Bras, Antoine Bosselut, Yejin Choi

BottleSum: Self-Supervised and Unsupervised Sentence Summarization using the Information Bottleneck Principle
EMNLP 2019
Peter West, Ari Holtzman, Jan Buys, Yejin Choi

∗ co-first author
∗∗ co-second author

PRESENTATIONS

Impossible Distillation (co-presented with Jaehun Jung) 2023
Microsoft Research Algorithms Group

Making NLP Less Supervised with Natural Language Generators 2022
University of British Columbia NLP Group

Making NLP Less Supervised with Natural Language Generators 2022
University of Pittsburgh NLP Group

Symbolic Knowledge Distillation: from General Language Models to Commonsense Models 2022
NAACL Oral Presentation

BottleSum: Unsupervised & Self-supervised Sentence Summarization w/ *Information Bottleneck Principle* 2020
AISC ML Explained

BottleSum: Unsupervised & Self-supervised Sentence Summarization w/ *Information Bottleneck Principle* 2019
EMNLP Oral Presentation

AWARDS & HONORS

EMNLP Outstanding Paper Award (Dialogue and Interactive Systems) 2023

ACL Outstanding Paper Award 2023

NAACL Best Methods Paper Award 2022

NSERC Postgraduate Scholarship-PGSD ($63,000 CAD over 3 years) 2018
at University of Washington

NSERC Undergraduate Student Research Award 2016
University of British Columbia

NSERC Undergraduate Student Research Award 2014, 2015
University of Alberta

Trek Excellence Scholarship 2016, 2015, 2014
University of British Columbia

Charles and Jane Banks Scholarship 2014, 2015, 2016
University of British Columbia

TEACHING

Teaching Assistant

Natural Language Processing (Masters) 2021
University of Washington

Natural Language Processing (Undergraduate) 2020
University of Washington

Intro to Algorithms (Undergraduate) 2017
University of Washington

Functional and Logic Programming (Undergraduate) 2016
University of British Columbia

Models of Computation (Undergraduate) 2016
University of British Columbia

PREVIOUS POSITIONS

Research Scientist Intern (multiple) 2019 - 2023
Allen Institute For Artificial Intelligence Seattle, WA
Hosts: Ronan Le Bras, Chandra Bhagavatula, Yejin Choi (Mosaic Team)
Projects: Reflective Decoding for Unsupervised Paraphrasing and Abductive Reasoning; Symbolic Knowledge Distillation; NovaCOMET: Open Commonsense Models

Research Scientist Intern 2020
Microsoft Research Seattle, WA
Hosts: Chris Quirk, Michel Galley
Project: Probing Factually Grounded Content Transfer with Factual Ablation

Undergraduate Researcher 2016
University of British Columbia Vancouver, BC, Canada
Hosts: Sara Mostafavi
Project: Multi-view Discriminative Gene Identification for Multi-class Setting

Undergraduate Researcher 2014, 2015
University of Alberta Edmonton, AB, Canada
Hosts: Al Meldrum
Project: Protein biosensing with fluorescent microcapillaries

PROGRAM COMMITTEE & REVIEWING

Program Committee/Reviewer (Conference):

ACL Roling Review	2021-2023
North American Association for Computational Linguistics (NAACL)	2019-2023
Association for Computational Linguistics (ACL)	2021-2023
Empirical Methods in Natural Language Processing (EMNLP)	2020-2023
International Conference on Learning Representations (ICLR)	2022-2023
Neural Information Processing Systems (NeurIPS)	2023
AAAI Conference on Artificial Intelligence (AAAI)	2021-2023

Program Committee/Reviewer (Workshop)

Blackbox NLP	2021 - 2023

SERVICE

Pre-Application Mentorship Service 2021, 2022, 2023
University of Washington

QueerInAI Grad App Program 2021, 2022

NLP Seminar Organizer 2022
University of Washington

NLP Retreat Organizer 2021
University of Washington