

Ximing Lu

PH.D. STUDENT · UNIVERSITY OF WASHINGTON

1414 NE 42nd St, Seattle, WA 98105

✉ lux32@cs.washington.edu | 🏠 gloriaximinglu.github.io | 🐦 @GXiming

Education

University of Washington

Seattle

PH.D. COMPUTER SCIENCE

2023 - current

- Advisor: Yejin Choi

University of Washington

Seattle

B.S. COMPUTER SCIENCE (SUMMA CUM LAUDE)

2017 - 2023

- Advisor: Yejin Choi

Professional Experience

2021-2024 **Pre-Doctoral Young Investigator**, The Allen Institute for AI, Mosaic

2020-2021 **Research Intern**, The Allen Institute for AI, Mosaic

2019 **Research Assistant**, University of Washington, Xlab

Publications

* denotes equal contribution

PUBLISHED

A Roadmap to Pluralistic Alignment

T Sorensen, J Moore, J Fisher, M Gordon, N Mireshghallah, CM Rytting, A Ye, L Jiang, **X Lu**, N Dziri, T Althoff, Y Choi
International Conference on Machine Learning (ICML) 2024

Impossible Distillation: From Low-Quality Model to High-Quality Dataset & Model for Summarization and Paraphrasing

J Jung, P West, L Jiang, F Brahman, **X Lu**, J Fisher, T Sorensen, Y Choi
North American Chapter of the Association for Computational Linguistics (NAACL) 2024

JAMDEC: Unsupervised Authorship Obfuscation using Constrained Decoding over Small Language Models

J Fisher, **X Lu**, J Jung, L Jiang, Z Harchaoui, Y Choi
North American Chapter of the Association for Computational Linguistics (NAACL) 2024

Phenomenal Yet Puzzling: Testing Inductive Reasoning Capabilities of Language Models with Hypothesis Refinement

L Qiu, L Jiang, **X Lu**, M Sclar, V Pyatkin, C Bhagavatula, B Wang, Y Kim, Y Choi, N Dziri, X Ren
International Conference on Learning Representations (ICLR) 2024, **Oral (top 1.2%)**

THE GENERATIVE AI PARADOX: “What It Can Create, It May Not Understand”

P West*, **X Lu***, N Dziri*, F Brahman*, L Li*, JD Hwang, L Jiang, J Fisher, A Ravichander, K Chandu, B Newman, P Koh, A Ettinger, Y Choi
International Conference on Learning Representations (ICLR) 2024

The Unlocking Spell on Base LLMs: Rethinking Alignment via In-Context Learning

BY Lin, A Ravichander, **X Lu**, N Dziri, M Sclar, K Chandu, C Bhagavatula, Y Choi
International Conference on Learning Representations (ICLR) 2024

Tailoring Self-Rationalizers with Multi-Reward Distillation

S Ramnath, B Joshi, S Hallinan, **X Lu**, LH Li, A Chan, J Hessel, Y Choi, X Ren
International Conference on Learning Representations (ICLR) 2024

Improving Language Models with Advantage-Based Offline Policy Gradients

A Baheti, **X Lu**, F Brahman, RL Bras, M Sap, M Riedl
International Conference on Learning Representations (ICLR) 2024

Value Kaleidoscope: Engaging AI with Pluralistic Human Values, Rights, and Duties
T Sorensen, L Jiang, JD Hwang, S Levine, V Pyatkin, P West, N Dziri, **X Lu**, K Rao, C Bhagavatula, M Sap, J Tasioulas, Y Choi
AAAI Conference on Artificial Intelligence (AAAI) 2024

Faith and Fate: Limits of Transformers on Compositionality
N Dziri*, **X Lu***, M Sclar*, X Lorraine Li, L Jiang, BY Lin, S Welleck, P West, C Bhagavatula, RL Bras, JD Hwang, S Sanyal, X Ren, A Ettinger, Z Harchaoui, Y Choi
Conference on Neural Information Processing Systems (NeurIPS) 2023, **Spotlight**

Localized Symbolic Knowledge Distillation for Visual Commonsense Models
J Park, J Hessel, K Chandu, P Pu Liang, **X Lu**, P West, Y Yu, Q Huang, J Gao, A Farhadi, Y Choi
Conference on Neural Information Processing Systems (NeurIPS)

Soda: Million-scale Dialogue Distillation with Social Commonsense Contextualization
H Kim, J Hessel, L Jiang, P West, **X Lu**, Y Yu, P Zhou, RL Bras, M Alikhani, G Kim, M Sap, Y Choi
Empirical Methods in Natural Language Processing (EMNLP) 2023, **Outstanding Paper Award**

Inference-time Policy Adapters (IPA): Tailoring Extreme-Scale LMs Without Fine-Tuning
X Lu, F Brahman, P West, J Jung, K Chandu, A Ravichander, P Ammanabrolu, L Jiang, S Ramnath, N Dziri, J Fisher, B Lin, S Hallinan, L Qin, X Ren, S Welleck, Y Choi
Empirical Methods in Natural Language Processing (EMNLP) 2023

NovaCOMET: Open Commonsense Foundation Models with Symbolic Knowledge Distillation
P West, RL Bras, T Sorensen, BY Lin, L Jiang, **X Lu**, K Chandu, J Hessel, A Baheti, C Bhagavatula, Y Choi
Findings of ACL: EMNLP 2023

STEER: Unified Style Transfer with Expert Reinforcement
S Hallinan, F Brahman, **X Lu**, J Jung, S Welleck, Y Choi
Findings of ACL: EMNLP 2023

ClarifyDelphi: Reinforced Clarification Questions with Defeasibility Rewards for Social and Moral Situations
V Pyatkin, JD Hwang, V Srikumar, **X Lu**, L Jiang, Y Choi, C Bhagavatula
Association for Computational Linguistics (ACL) 2023

I2D2: Inductive Knowledge Distillation with Neurologic and Self-Imitation
C Bhagavatula, JD Hwang, D Downey, RL Bras, **X Lu**, K Sakaguchi, S Swayamdipta, P West, Y Choi
Association for Computational Linguistics (ACL) 2023

Multimodal Knowledge Alignment with Reinforcement Learning
Y Yu, J Chung, H Yun, J Hessel, J Park, **X Lu**, R Zellers, P Ammanabrolu, RL Bras, G Kim, Y Choi
Conference on Computer Vision and Pattern Recognition (CVPR) 2023

Generating Sequences by Learning to Self-Correct
S Welleck*, **X Lu***, P West, F Brahman, T Shen, D Khashabi, Y Choi
International Conference on Learning Representations (ICLR) 2023

Quark: Controllable Text Generation with Reinforced Unlearning.
X Lu, S Welleck, L Jiang, J Hessel, L Qin, P West, P Ammanabrolu, Y Choi.
Conference on Neural Information Processing Systems (NeurIPS) 2022, **Oral**

NaturalProver: Grounded Mathematical Proof Generation with Language Models
S Welleck, J Liu, **X Lu**, H Hajishirzi, Y Choi
Conference on Neural Information Processing Systems (NeurIPS) 2022

Rainier: Reinforced Knowledge Introspector for Commonsense Question Answering
J Liu, S Hallinan, **X Lu**, P He, S Welleck, H Hajishirzi, Y Choi
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2022

ProsocialDialog: A Prosocial Backbone for Conversational Agent
H Kim, Y Yu, L Jiang, **X Lu**, D Khashabi, G Kim, Y Choi and M Sap
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2022

Twist Decoding: Diverse Generators Guide Each Other
J Kasai, K Sakaguchi, RL Bras, H Peng, **X Lu**, D Radev, Y Choi, NA Smith
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2022

MERLOT Reserve: Multimodal Neural Script Knowledge through Vision and Language and Sound
R Zellers, J Lu, **X Lu**, Y Yu, Y Zhao, M Salehi, A Kusupati, J Hessel, A Farhadi, Y Choi
Conference on Computer Vision and Pattern Recognition (CVPR) 2022, **Oral**

Connecting the Dots between Audio and Text without Parallel Data through Visual Knowledge Transfer
Y Zhao, J Hessel, Y Yu, **X Lu**, R Zellers, Y Choi
North American Chapter of the Association for Computational Linguistics (NAACL) 2022, **Oral**

NeuroLogic A*esque Decoding: Constrained Text Generation with Lookahead Heuristics
X Lu, S Welleck, P West, L Jiang, J Kasai, D Khashabi, RL Bras, L Qin, Y Yu, R Zellers, NA Smith, Y Choi
North American Chapter of the Association for Computational Linguistics (NAACL) 2022, **Best Paper Award**

Symbolic Knowledge Distillation: from General Language Models to Commonsense Models
P West, C Bhagavatula*, J Hessel*, JD Hwang*, L Jiang*, RL Bras*, **X Lu***, S Welleck*, Y Choi
North American Chapter of the Association for Computational Linguistics (NAACL) 2022, **Oral**

Generated knowledge prompting for commonsense reasoning
J Liu, A Liu, **X Lu**, S Welleck, P West, RL Bras, Y Choi, H Hajishirzi
Association for Computational Linguistics (ACL) 2022

MERLOT: Multimodal Neural Script Knowledge Models
R Zellers*, **X Lu***, J Hessel*, Y Yu, J Park, J Cao, A Farhadi, Y Choi
Conference on Neural Information Processing Systems (NeurIPS) 2021, **Oral (top 1%)**

Analyzing Commonsense Emergence in Few-shot Knowledge Models
J Da, RL Bras, **X Lu**, Y Choi, A Bosselut
Automated Knowledge Base Construction (AKBC) 2021

DExperts: On-the-Fly Controlled Text Generation with Experts and Anti-Experts
A Liu, M Sap, **X Lu**, S Swayamdipta, C Bhagavatula, NA Smith, Y Choi
Association for Computational Linguistics (ACL) 2021, **Oral**

Reflective Decoding: Beyond Unidirectional Generation with Off-the-shelf Language Models
P West, **X Lu**, A Holtzman, C Bhagavatula, JD Hwang, Y Choi
Association for Computational Linguistics (ACL) 2021

On-the-Fly Attention Modulation for Neural Generation
Y Dong, C Bhagavatula, **X Lu**, JD Hwang, A Bosselut, JCK Cheung, Y Choi
Findings of the Association for Computational Linguistics (ACL) 2021

Neurologic decoding: (un)supervised neural text generation with predicate logic constraints
X Lu, P West, R Zellers, RL Bras, C Bhagavatula, Y Choi
North American Chapter of the Association for Computational Linguistics (NAACL) 2021

End-to-End diagnosis of breast biopsy images with transformers
S Mehta*, **X Lu***, W Wu, D Weaver, H Hajishirzi, JG Elmore, LG Shapiro
Medical Image Analysis 79, 102466

Applications of the ESPNet Architecture in Medical Imaging
S Mehta, N Nuechterlein, E Mercan, B Li, S Nofallah, W Wu, **X Lu**, A Caspi, M Rastegari, J Elmore, H Hajishirzi, L Shapiro
State of the Art in Neural Networks and their Applications, 117-131

Analysis of Regions of Interest and Distractor Regions in Breast Biopsy Images
X Lu, S Mehta, TT Brunyé, DL Weaver, JG Elmore, LG Shapiro
International Conference on Biomedical and Health Informatics (BHI) 2021

IN REVIEW

Information-Theoretic Distillation for Reference-less Summarization
J Jung, **X Lu**, L Jiang, F Brahman, P West, PW Koh, Y Choi

In Search of the Long-Tail: Systematic Generation of Long-Tail Knowledge via Logical Rule Guided Search
H Li, Y Ning, Z Liao, S Wang, XL Li, **X Lu**, W Zhao, F Brahman, Y Choi, X Ren

Awards & Scholarships

- 2023 **Outstanding Paper Award**, Empirical Methods in Natural Language Processing Conference (EMNLP)
Best Senior Thesis Award, Paul G. Allen School of Computer Science & Engineering
- 2022 **Best Paper Award**, North American Chapter of the Association for Computational Linguistics (NAACL)
- 2020 **Outstanding Undergraduate Researcher Award Runners-Up**, Computing Research Association
Lisa Simonyi Prize, Paul G. Allen School of Computer Science & Engineering
Levinson Emerging Scholars Award, University of Washington
Mary Gates Research Scholarship, University of Washington
- 2019 **Denton, Denice Dee Scholars Endowment**, Paul G. Allen School of Computer Science & Engineering
Dean's List, University of Washington
- 2018 **Second Prize of UW Datathon**, Citadel Investment Group, LLC
Conference Travel Award, University of Washington

Teaching Experience

- Winter, 2021 **CSE P 517: Natural Language Processing**, Teaching Assistant *University of Washington*
- Winter, 2024 **CSE 447/517: Natural Language Processing**, Teaching Assistant *University of Washington*

Service

- 2024 Area Chair for conference NAACL
Reviewer for conference ACL, Neurips and COLM
- 2023 Area Chair for conference EACL
Reviewer for conference ACL, ICML, EMNLP, Neurips, AAAI and CVPR
Reviewer for workshop InterNLP and CSRR
Reviewer for Journal of Artificial Intelligence Research
- 2020 Mentor in UW Dream Project
Member of Google's Women Techmakers Program
Member of Association for Computing Machinery's Council on Women in Computing
- 2019 Mentor in ACM Big/Little Mentorship Program
Volunteer at UW CSE Computing Open House 2019