

Liwei Jiang

Paul G. Allen School of Computer Science & Engineering
University of Washington
Box 352350
Seattle, WA 98195

+1 207-313-6030
lwjiang@cs.washington.edu
<https://liweijiang.me>
[@liweijianglw](https://twitter.com/liweijianglw)

RESEARCH FOCUS

Machine ethics, commonsense moral reasoning, AI applications for social good, and trustworthy human-machine interactions

EDUCATION

University of Washington, Seattle, WA

09/2019–current Ph.D. in Computer Science & Engineering
09/2019–05/2022 M.S. in Computer Science & Engineering
Advisor: Yejin Choi

Colby College, Waterville, ME (top 1%, Dean's List 15–19)

09/2015–01/2019 B.A. in Computer Science, *summa cum laude*, GPA 4.08
B.A. in Mathematics, *summa cum laude*, GPA 4.13

PROFESSIONAL EXPERIENCES

Allen Institute for Artificial Intelligence (AI2)

06/2020–current Research Intern at the Mosaic Team, *with Yejin Choi*

University of Washington, Computer Science & Engineering

04/2020–current Research Assistant, *with Yejin Choi*
Commonsense moral reasoning, value alignment, applications for social good

09/2019–08/2020 Research Assistant, *with James Fogarty*
Effective and easy-to-use self-tracking tools for migraine patients

Stanford University, Computer Science

03/2019–09/2019 Research Intern, *with James Landay*
EnglishBot, a language learning chatbot conversing with students interactively [P.10]
BookBuddy, an interactive chatbot tutor for children to practice reading skills [W.1]

06/2018–11/2018 Research Intern, *with James Landay*
QuizBot, a dialogue-based conversational agent for factual knowledge learning [P.11]

06/2017–09/2017 CSLI Summer Research Intern, *with James Landay*
A rule-based chatbot tutor for answering children’s questions, as an early prototype of the [Smart Primer](#) project

Colby College, Computer Science

09/2017–01/2019 Undergraduate Research Assistant, *with Bruce Maxwell*
Honors thesis project on *Assistive Robot Guide for Visually Impaired Users*

05/2016–08/2016 Undergraduate Research Assistant, *with Ying Li*
Navigation with mobile phones without connectivity (e.g., GPS or Wi-Fi)

PUBLICATIONS

*, † denote equal contribution

Manuscripts and Pre-prints

2022 P.1 **Liwei Jiang**, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jenny Liang, Jesse Dodge, Keisuke Sakaguchi, Maxwell Forbes, Jon Borchardt, Saadia Gabriel, Yulia Tsvetkov, Oren Etzioni, Maarten Sap, Regina Rini, and Yejin Choi. [Can machines learn morality? The Delphi Experiment](#). In *arXiv preprint arXiv:2110.07574*, arXiv 2022

Peer-reviewed Conference and Journal Publications

2023 P.2 Valentina Pyatkin, Jena D. Hwang, Vivek Srikumar, Ximing Lu, **Liwei Jiang**, Yejin Choi, and Chandra Bhagavatula. Reinforced clarification question generation with defeasibility rewards for disambiguating social and moral situations, 2022

P.3 Hyunwoo Kim, Jack Hessel, **Liwei Jiang**, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, and Yejin Choi. Soda: Million-scale dialogue distillation with social commonsense contextualization, 2022

2022 P.4 Ximing Lu, Sean Welleck, **Liwei Jiang**, Jack Hessel, Lianhui Qin, Peter West, Prithviraj Ammanabrolu, and Yejin Choi. [Quark: Controllable Text Generation with Reinforced Unlearning](#). In *Advances in Neural Information Processing Systems*, NeurIPS 2022

P.5 Hyunwoo Kim*, Youngjae Yu*, **Liwei Jiang**, Ximing Lu, Daniel Khashabi, Gunhee Kim, Yejin Choi, and Maarten Sap. [ProsocialDialog: A Prosocial Backbone for Conversational Agents](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2022

P.6 Prithviraj Ammanabrolu, **Liwei Jiang**, Maarten Sap, Hannaneh Hajishirzi, and Yejin Choi. [Aligning to Social Norms and Values in Interactive Narratives](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL 2022

- P.7 Ximing Lu, Sean Welleck*, Peter West*, **Liwei Jiang**†, Jungo Kasai†, Daniel Khashabi†, Ronan Le Bras†, Lianhui Qin†, Youngjae Yu†, Rowan Zellers†, Noah A. Smith, and Yejin Choi. [NeuroLogic A*esque Decoding: Constrained Text Generation with Lookahead Heuristics](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL 2022 🏆 **Best Paper Award**
- P.8 Peter West, Chandra Bhagavatula*, Jack Hessel*, Jena D. Hwang*, **Liwei Jiang***, Ronan Le Bras*, Ximing Lu*, Sean Welleck*, and Yejin Choi. [Symbolic Knowledge Distillation: from General Language Models to Commonsense Models](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL 2022
- 2021 P.9 **Liwei Jiang**, Antoine Bosselut, Chandra Bhagavatula, and Yejin Choi. [“I’m Not Mad”: Commonsense Implications of Negation and Contradiction](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL 2021
- P.10 Sherry Ruan*, **Liwei Jiang***, Qianyao Xu*, Zhiyuan Liu, Glenn M. Davis, Emma Brunskill, and James A. Landay. [EnglishBot: An AI-Powered Conversational System for Second Language Learning](#). In *26th International Conference on Intelligent User Interfaces*, IUI 2021
- 2019 P.11 Sherry Ruan, **Liwei Jiang**, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L. Murnane, Emma Brunskill, and James A. Landay. [QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge](#). In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI 2019

Posters, Extended Abstracts, Workshop Papers and Technical Reports

- 2019 W.1 Sherry Ruan, Angelica Willis, Qianyao Xu, Glenn M. Davis, **Liwei Jiang**, Emma Brunskill, and James A. Landay. [BookBuddy: Turning Digital Materials Into Interactive Foreign Language Lessons Through a Voice Chatbot](#). In *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, L@S WIP 2019

HONORS AND AWARDS

- | | |
|------------------|--|
| 2019–2020 | Anne Dinning - Michael Wolf Endowed Regental Fellowship
<i>University of Washington, Paul G. Allen School First-Year Ph.D. Fellowship</i> |
| 2018 | Member of the Phi Beta Kappa Society
<i>Colby College, elected as a member of Phi Beta Kappa with junior standing</i> |
| 2016, 2017, 2018 | Julius Seelye Bixler Scholar
<i>Colby College, top-ranking students as determined by the academic record, three-time recipient</i> |
| 2018 | Honorable Mention of Interdisciplinary Contest in Modeling (ICM)
<i>20th annual Interdisciplinary Contest in Modeling (ICM)</i> |

- 2017 Phi Beta Kappa Undergraduate Scholastic Achievement Award
Colby College, top two students in the sophomore and junior classes
- 2016 Phi Beta Kappa Summer Research Scholar
Colby College, summer research stipend

INVITED TALKS

- 2023 **Toward Interpretable and Interactive Socially & Ethically Informed AI**
Guest lecturer, LAW E 553 Technology Law And Public Policy Seminar at UW (2023.3)
- 2022 **Toward Socially Aware & Ethically Informed AI**
UW NLP Retreat (2022.9)
Toward Socially Aware & Ethically Informed AI
Co-guest lecturer w/ Saadia Gabriel, The Downtown School, Seattle (2022.9)
Toward Ethically Informed & Socially Aware AI
Guest lecturer, HONORS 222 B at UW (2022.5)
- 2021 **Delphi: Toward Machine Ethics and Norms**
Allen Institute for Artificial Intelligence, All-AI2 Meeting (2021.10)

TEACHING EXPERIENCES

Teaching Assistant

- 01/2023–03/2023 **CSE599 D1 Exploration on Language, Knowledge, and Reasoning**, UW
- 09/2016–01/2019 **CS151 Introduction to Computational Thinking**, Colby College
CS231 Data Structure & Algorithm, Colby College
CS251 Data Analysis & Visualization, Colby College
Graded programming projects and homework, held TA office hours and tutored sessions weekly
- 09/2018–01/2019 **MA311 Ordinary Differential Equation**, Colby College
Held TA office hours and graded problem sets weekly for 30 students

Guest Lecturer

- 03/2023 **Toward Interpretable and Interactive Socially & Ethically Informed AI**
In LAW E 553 Technology Law And Public Policy Seminar, University of Washington
- 09/2022 **Toward Socially Aware & Ethically Informed AI**
In Ethics and Citizenship, w/ Saadia Gabriel, The Downtown School, Seattle
- 05/2022 **Toward Ethically Informed & Socially Aware AI**
In HONORS 222 B: Artificial Intelligence Meets Society, University of Washington

MENTORING EXPERIENCES

- 01/2022–present **Kavel Rao** (Undergraduate student at UW '19)
Explainable defeasible moral reasoning
- 01/2022–05/2022 **Jillian Fisher** (Junior PhD student at UW CSE)
On Model Revision: Content Re-mapping and Avoidance

- 09/2021–02/2023 **Sravani Nanduri** (Undergraduate student at UW '21)
 Co-supervised with Maarten Sap & Tongshuang (Sherry) Wu
 Online hate speech moderation with explanations
- 12/2021–03/2022 **Nuria Alina Chandra** (Undergraduate student at UW '19)

PROFESSIONAL SERVICE

Paper Reviewing

- NLP **EMNLP 2022, ACL 2021, AAAI 2023**
 HCI **CHI 2020 LBW 2020**

Community Service

- 2023 **Area Chair, Reviewer**, UW Allen School PhD Admissions
Coordinate between PhD students/postdocs reading for PhD admissions and advising staff/admissions committee.
Liaison, UW Allen School Faculty Recruiting
Keep students informed about faculty recruiting and coordinate with student hosts to carry out responsibilities.
Co-chairs, UW Allen School Prospective Student Committee
Organize visit days for prospective PhD students at UW Allen School.
- 2022 **Volunteer**, UW NLP Retreat
Area Chair, Reviewer, UW Allen School PhD Admissions
Coordinate between PhD students/postdocs reading for PhD admissions and advising staff/admissions committee.
Liaison, UW Allen School Faculty Recruiting
Keep students informed about faculty recruiting and coordinate with student hosts to carry out responsibilities.
Student Representative, UW Allen School Diversity and Inclusion Committee
Co-chairs, UW Allen School Prospective Student Committee
Organize visit days for prospective PhD students at UW Allen School.
Volunteer Coordinator, NAACL 2022
- 2021 **Mentor**, UW Allen School Pre-Application Mentorship Service (PAMS)
A program supporting potential CS PhD applicants, with 80% from underrepresented communities.
Liaison, Reviewer, UW Allen School PhD Admissions
- 2020 **Co-organizer**, UW Allen School Pre-Application Mentoring Service (PAMS)

MEDIA COVERAGE

- 2022 [Can Computers Learn Common Sense?](#)
 The New Yorker, 04/2022
- 2021 [Can a Machine Learn Morality?](#)
 The New York Times, 12/2021
- 2021 [This Program Can Give AI a Sense of Ethics–Sometimes](#)
 Wired, 12/2021
- 2021 [Machines Learn Good From Commonsense Norm Bank](#)
 IEEE Spectrum, 11/2021

Updated February 2023