

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

RESEARCH FOCUS

My research focuses on cultural analytics and digital humanity approaches that mine and analyze moral norms, social biases, pluralistic human values, and cultures through massive human-curated and web data, with assistance from language models. I apply theoretical frameworks motivated by cognitive science, moral philosophy, social psychology, among other humanity fields. I also work on designing applications for social good, e.g., healthcare and education, and for understanding human commonsense.

EDUCATION

University of Washington, Seattle, WA

- 09/2019–current Ph.D. in Computer Science & Engineering
Focusing on computational social science, cultural analytics, digital humanity
- 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

University of Washington, Computer Science & Engineering

- 04/2020–current Research Assistant, *with Yejin Choi*
Computational social and moral norms [P.1, P.4, P.5, P.8, P.9, P.10],
Pluralistic human values [P.2],
Social biases [P.7],
Commonsense reasoning [P.11, P.6],
Multi-cultural norms
- 09/2019–08/2020 Research Assistant, *with James Fogarty*
Effective and easy-to-use self-tracking tools for migraine patients [P.3]

Allen Institute, Seattle, Washington, US

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

Stanford University, Computer Science

- 03/2019–09/2019 Research Intern, *with James Landay*
EnglishBot, a language learning chatbot conversing with students interactively [P.12]
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.13]
- 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

PUBLICATIONS

*, †denote equal contribution

Selected Manuscripts and Pre-prints

- 2022 P.1 **Liwei Jiang**, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jenny Liang, Sydney Levine, Jesse Dodge, Keisuke Sakaguchi, Maxwell Forbes, Taylor Sorensen, Jon Borchardt, Jack Hessel, 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*, In-submission 2022
- 2023 P.2 Taylor Sorensen, **Liwei Jiang**, Jena Hwang, Sydney Levine, Valentina Pyatkin, Peter West, Nouha Dziri, Ximing Lu, Kavel Rao, Chandra Bhagavatula, Maarten Sap, John Tasioulas, and Yejin Choi. [Value Kaleidoscope: Engaging Machines with Pluralistic Human Values, Rights, and Duties](#). In-submission 2023
- P.3 Yasaman S. Sefidgar, Carla L. Castillo, Shaan Chopra, **Liwei Jiang**, Tae Jones, Anant Mittal, Hyeyoung Ryu, Jessica Schroeder, Allison Cole, Natalia Murinova, Sean A. Munson, and James Fogarty. [MigraineTracker: Examining Patient Experiences with Goal-Directed Self-Tracking for a Chronic Health Condition](#). In-submission 2023

Selected Peer-reviewed Conference and Journal Publications

- 2023 P.4 Kavel Rao*, **Liwei Jiang***, Valentina Pyatkin, Yuling Gu, Niket Tandon, Nouha Dziri, Faeze Brahman, and Yejin Choi. [What Makes it Ok to Set a Fire? Iterative Self-distillation of Contexts and Rationales for Disambiguating Defeasible Social and Moral Situations](#). Findings of EMNLP 2023
- P.5 Seungju Han, Junhyeok Kim, Jack Hessel, **Liwei Jiang**, Jiwan Chung, Yejin Son, Yejin Choi, and Youngjae Yu. [Reading Books is Great, But Not if You Are Driving! Visually Grounded Reasoning about Defeasible Commonsense Norms](#). EMNLP 2023
- P.6 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](#). EMNLP 2023

- P.7 Yiming Zhang, Sravani Nanduri, **Liwei Jiang**, Tongshuang Wu, and Maarten Sap. [BiasX: "Thinking Slow" in Toxic Content Moderation with Explanations of Implied Social Biases](#). EMNLP 2023
- P.8 Valentina Pyatkin, Jena D. Hwang, Vivek Srikumar, Ximing Lu, **Liwei Jiang**, Yejin Choi, and Chandra Bhagavatula. [ClarifyDelphi: Clarification Questions with Defeasibility Rewards for Social and Moral Situations](#). ACL 2023
- 2022 P.9 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.10 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
- 2021 P.11 **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.12 Sherry Ruan*, **Liwei Jiang***, Qianqiao Xu*, Zhiyuan Liu, Glenn M. Davis, Emma Brunskill, and James A. Landay. [EnglishBot: A Conversational System for Second Language Learning](#). In *26th International Conference on Intelligent User Interfaces*, IUI 2021
- 2019 P.13 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, Qianqiao 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
University of Washington, Paul G. Allen School First-Year Ph.D. Fellowship

- 2018 **Member of the Phi Beta Kappa Society**
Colby College, elected as a member of Phi Beta Kappa with junior standing
- 2016, 2017, 2018 **Julius Seelye Bixler Scholar**
Colby College, top-ranking students as determined by the academic record, three-time recipient
- 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 **Can we teach machines human ethics and values?**
 Co-speaker w/ Valentina Pyatkin and Taylor Sorensen, The Downtown School, Seattle (2023.9)
Toward Interpretable, Interactive, Informative Machine Moral Reasoning
 Discussant, Allen Institute Mosaic Morality Series (2023.2)
- 2021 **Delphi: Toward Machine Ethics and Norms**
 Allen Institute, All Team 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

- 09/2023 **Can we teach machines human ethics and values?**
In Ethics and Citizenship, w/ Valentina Pyatkin and Taylor Sorensen, The Downtown School, Seattle

MENTORING EXPERIENCES

- 01/2022–present **Kavel Rao** (Undergraduate student at UW CSE)
 Explainable defeasible moral reasoning [P.4]
- 03/2023–present **Kelly Chiu** (Master student at UW Linguistics)
 Automatic mining of multi-cultural norms (on-going)
- 09/2021–02/2023 **Sravani Nanduri** (Undergraduate student at UW CSE)
 Co-supervised with Maarten Sap & Tongshuang (Sherry) Wu
 Online hate speech moderation with explanations [P.7]
- 12/2021–03/2022 **Nuria Alina Chandra** (Undergraduate student at UW CSE)

PROFESSIONAL SERVICE

Paper Reviewing

NLP EMNLP 2022, ACL 2021
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 [Machines Learn Good From Commonsense Norm Bank](#)
IEEE Spectrum, 11/2021

Updated October 2023