

Jordan Boyd-Graber

Email: jordan.ying@ntu.edu.sg

Contact number: (+65) 9148 1138

Office S3-01c-103, CCDS, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

Web: <https://boydgraber.org>

Summary

Jordan Boyd-Graber's research focus is in applying machine learning to problems that help computers better work with or understand humans. His research applies statistical models to natural language problems in ways that interact with humans, learn from humans, or help researchers understand humans.

Jordan is an expert in the application of topic models, automatic tools that discover structure and meaning in large, multilingual datasets. His work has been supported by NSF,

DARPA, IARPA, and ARL.

Three of his students have gone on to tenure track positions at NYU, U Mass Amherst, and Ursinus.

His awards include a 2017 NSF CAREER, the Karen Spärk Jones prize; "best of" awards at NIPS, CoNLL, EMNLP, and NAACL; and a Computing Innovation Fellowship (declined). His Erdős number is 2 (via Maria Klawe), and his Bacon number is 3 (by embarrassing himself on *Jeopardy!*).

Positions Held

University of Maryland

Full Professor in Computer Science, UMIACS, and Information Science

Associate Professor

Assistant Professor of Information Studies (iSchool)

COLLEGE PARK, MD

2024–Present

2017–2024

2010–2014

University of Colorado Boulder

Assistant Professor of Computer Science

Associate Professor of Computer Science

BOULDER, CO

2014–2017

2017

Education

Princeton University

Ph.D. in Computer Science

Advisor: David Blei; Thesis: Linguistic Extensions of Topic Models

PRINCETON, NJ

2004 – 2010

California Institute of Technology

B.S. in Computer Science and History (dual degree)

PASADENA, CA

2000 – 2004

Selected Publications

Note: Students I have advised are underlined.

1. Yoo Yeon Sung, Eve Fleisig, Ishani Mondal, and Jordan Lee Boyd-Graber. **ADVSCORE: A Metric for the Evaluation and Creation of Adversarial Benchmarks**. *North American Association for Computational Linguistics*, Preprint.
2. Chenglei Si, Navita Goyal, Tongshuang Wu, Chen Zhao, Shi Feng, Hal Daumé III, and Jordan Boyd-Graber. **Large Language Models Help Humans Verify Truthfulness—Except When They Are Convincingly Wrong**. *North American Association for Computational Linguistics*, 2024.
3. Sander V Schulhoff, Jeremy Pinto, Ansum Khan, Louis-François Bouchard, Chenglei Si, Jordan Lee Boyd-Graber, Svetlana Anati, Valen Tagliabue, Anson Liu Kost, and Christopher R Carnahan. **Ignore This Title and Hack-A-Prompt: Exposing Systemic Vulnerabilities of LLMs Through a Global Prompt Hacking Competition**. *Empirical Methods in Natural Language Processing*, 2023.
4. Shi Feng and Jordan Boyd-Graber. **Learning to Explain Selectively: A Case Study on Question Answering**. *Empirical Methods in Natural Language Processing*, 2022.
5. Alexander Hoyle, Pranav Goel, Denis Peskov, Andrew Hian-Cheong, Jordan Boyd-Graber, and Philip Resnik. **Is Automated Topic Model Evaluation Broken?: The Incoherence of Coherence**. *Neural Information Processing Systems*, 2021.
6. Eric Wallace, Pedro Rodriguez, Shi Feng, Ikuya Yamada, and Jordan Boyd-Graber. **Trick Me If You Can: Human-in-the-loop Generation of Adversarial Question Answering Examples**. *Transactions of the Association of Computational Linguistics*, 2019.
7. Jordan Boyd-Graber, Yuening Hu, and David Mimno. **Applications of Topic Models**. 2017.
8. He He, Jordan Boyd-Graber, Kevin Kwok, and Hal Daumé III. **Opponent Modeling in Deep Reinforcement Learning**. *International Conference on Machine Learning*, 2016.
9. Alvin Grissom II, Jordan Boyd-Graber, He He, John Morgan, and Hal Daumé III. **Don't Until the Final Verb Wait: Reinforcement Learning for Simultaneous Machine Translation**. *Empirical Methods in Natural Language Processing*, 2014.
10. Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M. Blei. **Reading Tea Leaves: How Humans Interpret Topic Models**. *Neural Information Processing Systems*, 2009.

Natural languages: English (*native*), German (*working*), and Mandarin Chinese (*beginner*).