Jordan Boyd-Graber

jbg@umiacs.umd.edu • +1 (920) 524-9464 • http://umiacs.umd.edu/~jbg
Iribe 4146 • Computer Science • iSchool • College Park, MD

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 Erdös number is 2 (via Maria Klawe), and his Bacon number is 3 (by embarassing himself on *Jeopardy!*).

Positions Held

University of Maryland — College Park, MD

- **Associate Professor in Computer Science, UMIACS, and iSchool** — *2017–Present*
- **Assistant Professor in the Institute for Advanced Computer Studies** — *2011–2014*
- **Assistant Professor of Information Studies (iSchool)** — *2010–2014*

University of Colorado Boulder — Boulder, CO

- **Assistant Professor of Computer Science** — *2014–2017*
- **Associate Professor of Computer Science** — *2017*

Education

Princeton University — Princeton, NJ

- **Ph.D. in Computer Science** — *2004 – 2010*
  Advisor: David Blei; Thesis: Linguistic Extensions of Topic Models

California Institute of Technology — Pasadena, CA

- **B.S. in Computer Science and History (dual degree)** — *2000 – 2004*

Selected Publications

Note: Students I have advised are underlined.

1. 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.
2. Sander V Schulhoff, Jeremy Pinto, Anaum Khan, Louis-François Bouchard, Chenglei Si, Jordan Lee Boyd-Graber, Svetlina Anati, Valen Tagliabue, Anson Liu Kost, and Christopher R Carnahan. **Ignore This Title and HackAPrompt: Exposing Systemic Vulnerabilities of LLMs Through a Global Prompt Hacking Competition**. *Empirical Methods in Natural Language Processing*, 2023.
3. Shi Feng and **Jordan Boyd-Graber**. **Learning to Explain Selectively: A Case Study on Question Answering**. *Empirical Methods in Natural Language Processing*, 2022.
4. 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.
5. 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.
6. **Jordan Boyd-Graber**, Yuening Hu, and David Mimno. **Applications of Topic Models**. 2017.
7. Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, **Jordan Boyd-Graber**, and Hal Daumé III. **Feuding Families and Former Friends: Unsupervised Learning for Dynamic Fictional Relationships**. *North American Association for Computational Linguistics*, 2016.
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*).