Ian Stewart

Researcher in natural language processing.

✉ ibs@dartmouth.edu
🖥 ianbstewart.github.io

I develop natural language processing methods that address human and societal needs, e.g. making models more robust to variation in user input.

Education

2015–2020 **Georgia Institute of Technology**, *Ph.D. in Human-Centered Computing*.
Thesis: "The laws of LOL: computational approaches to sociolinguistic variation in online discussions."
Advisors: Jacob Eisenstein and Diyi Yang.

2010–2014 **Dartmouth College**, *A.B. Linguistics (cum laude), minor in Computer Science*.
Thesis: "African American English syntax in Twitter."
Advisors: James Stanford and Sravana Reddy.

Skills

Programming   Python (proficient), R (intermediate), Java (intermediate)

Libraries   huggingface, sklearn, statsmodels, torch, gensim, nltk, Beautiful Soup, matplotlib, pandas

Natural languages   English (native), French (advanced), Japanese (beginner), Spanish (beginner), Māori (beginner)

Work experience

Oct 2022 – present **Data Scientist III**, *Pacific Northwest National Laboratory*.
- Develop machine learning methods for robustness and interpretability in the national security domain.
- Collaborate with user experience researchers to implement effective and engaging human-AI systems.
- Adapt AI methods to relevant mission goals within DOE.

Sep 2020 – Sept 2022 **Research Fellow**, University of Michigan.
- Worked with Rada Mihalcea in the Language and Information Technologies Lab at University of Michigan.
- Developed text generation models to adapt to different social needs.
- Implemented crowdworker surveys to evaluate chat system output.
- Investigated emergent social bias in NLP systems, including machine translation and word embeddings.
- Mentored undergraduate and graduate students in research projects related to machine learning.

Aug 2015 – Aug 2020 **Graduate Research Assistant**, Georgia Institute of Technology.
- Developed computational methods to describe and predict language variation in online discussions.
- Applied linguistic analysis to understand social dynamics in real-life scenarios, e.g. breaking news events.
- Adapted statistical models such as causal inference and survival analysis to test theories of sociolinguistic variation.

May 2018 – Aug 2018 **PhD Fellow**, Max Planck Institute for Demographic Research.
- Leveraged Facebook advertising API to assess immigrant behavior in the US.
- Collected, organized and analyzed large-scale data with statistical models to test hypotheses on immigrant assimilation.
- Engaged with interdisciplinary research team to set up survey for follow-up analysis.

Teaching

- Sept 2021 - May 2022: NLP+CSS 201 Tutorial Series (co-organizer, with Katherine Keith)
- Jan - May 2020: Natural language processing (Teaching Assistant)
- Sep - Dec 2015: Introduction to artificial intelligence (Teaching Assistant)

Awards

Sept 2022   **Research Grant**, *Amazon*.
Computing credits for a proposed project on personalized text generation.

Sept 2021   **Research Grant**, Summer Institutes in Computational Social Science.
Funding for an open-source tutorial series about advanced natural language processing methods for computational social science.

March 2017   **Honorable Mention**, National Science Foundation Graduate Research Fellowship.
Runner-up in annual Graduate Research Fellowship competition. Award includes increased access to supercomputing resources.

August 2016   **Tuition Reimbursement Award**, Pacific Northwest National Laboratory.
Awarded to PhD interns in the National Security Internship Program who demonstrated outstanding performance in their summer research project. Covers one semester of graduate school tuition.

August 2015   **Great Promise Award**, Charles River Analytics.
Awarded to an intern in each company division who shows significant promise as a software engineer.

May 2014   **Academic Achievement Award**, Dartmouth College Linguistics Department.
Awarded to a graduating student who has demonstrated considerable commitment to research in linguistics.

In preparation.

Ian Stewart, Sameera Horawalavithana, Brendan Kennedy, Sai Munikoti, Karl Pazdernik. Surprisingly fragile: assessing and addressing instability in multimodal foundation models.

Sai Munikoti, Ian Stewart, Sameera Horawalavithana, Henry Kvinge, Tegan Emerson, Sandra E Thompson, Karl Pazdernik. Generalist Multimodal AI: A Review of Architectures, Challenges and Opportunities.

Publications

2025   Rounak Meyur, Hung D Phan, Koby B Hayashi, Ian Stewart, Shivam Sharma, Sarthak Chaturvedi, Mike Parker, Daniel M. Nally, Sadie A. Montgomery, Karl Pazdernik, Ali Jannesari, Mahantesh M Halappanavar, Sai Munikoti, Sameera Horawalavithana, Anurag Acharya . Benchmarking LLMs for Environmental Review and Permitting. KDD SocLLM Workshop.

2024   Sameera Horawalavithana, Sai Munikoti, Ian Stewart, Henry Kvinge. SCITUNE: Aligning Large Language Models with Scientific Multimodal Instructions. NLP4Science Workshop.

2024   I. Stewart. R. Mihalcea. Whose wife is it anyway? Assessing bias against same-gender relationships in machine translation. 5th Gender Bias in NLP Workshop (ACL).

2023   I. Stewart, K. Keith. Democratizing Machine Learning for Interdisciplinary Scholars: Report on Organizing the NLP+CSS Online Tutorial Series. KONVENS Teach4NLP Workshop.

2023   L. Biester, I. Stewart, L. Hirschfield, R. Mihalcea, S. Pozzi. Lexical Measurement of Teaching Qualities. American Society of Engineering Education.

2023   I. Stewart, C. Welch, R. Mihalcea, K. Resnicow, L. An, J. Pennebaker. Expressive Interviewing Agents to Support Health-Related Behavior Change: A Study of COVID-19 Behaviors. Journal of Medical Internet Research: Formative Research.

2023   A. Gunal, I. Stewart, R. Mihalcea, V. Pérez-Rosas. Understanding the Role of Questions in Mental Health Support-Seeking Forums. AAAI Workshop on Health Intelligence.

2022   I. Stewart, R. Mihalcea. How well do you know your audience? Socially-aware question generation. SIGDIAL.

2022   S. Castro, R. Wang, P. Huang, I. Stewart, O. Ignat, N. Liu, J. Stroud, R. Mihalcea. FIBER: Fill-in-the-Blanks as a Challenging Video Understanding Evaluation Framework. ACL.

2022   G. Kambhatla, I. Stewart, R. Mihalcea. Surfacing Racial Stereotypes through Identity Portrayal. Fairness, Accountability, and Transparency Conference.

2021 M. Dong, X, Xu, Y. Zhang, I. Stewart, R. Mihalcea. Room to Grow: Understanding Personal Characteristics Behind Self Improvement Using Social Media. SocialNLP Workshop.

2021 I. Stewart, D. Yang, J. Eisenstein. Tuiteamos o pongamos un tuit? Investigating the Social Constraints of Loanword Integration in Spanish Social Media. Society for Computation in Linguistics.

2020 I. Stewart, D. Yang, J. Eisenstein. Characterizing Collective Attention via Descriptor Context: A Case Study of Public Discussions of Crisis Events. ICWSM.

2019 I. Stewart, R. Flores, T. Riffe, I. Weber, E. Zagheni. Rock, rap, or reggaeton? Assessing Mexican immigrants' cultural assimilation using Facebook data. WebConf.

2018 I. Stewart, J. Eisenstein. Making "fetch" happen: the influence of social and linguistic context on nonstandard word growth and decline. EMNLP.

2018 I. Stewart, Y. Pinter, J. Eisenstein. Si o no, que penses? Catalonian independence and linguistic identity on social media. NAACL.

2017 I. Stewart, S. Chancellor, M. De Choudhury, J. Eisenstein. #anorexia, #anarexia, #anarexyia: Characterizing Online Community Practices with Orthographic Variation. IEEE Big Data, SocialNLP Workshop.

2017 F. Hohman, S. Soni, I. Stewart, J. Stasko. A Viz of Ice and Fire: Exploring Entertainment Video Using Color and Dialogue. Workshop on Visualization for the Digital Humanities.

2017 I. Stewart, D. Arendt, E. Bell, S. Volkova. Measuring, Predicting and Visualizing Short-Term Change in Word Representation and Usage in VKontakte Social Network. ICWSM.

2015 E. Stickgold, B. Skarin, I. Stewart, C. Lofdahl. Extending generative models of large scale networks. 24th Conference on Behavior Representation in Modeling and Simulation (BRiMS).

2014 I. Stewart. Now we stronger than ever: African American syntax on Twitter. EACL.

Non-archival presentations

2022 Y. Shen, I. Stewart, R. Mihalcea. Biased Attention is Not What We Need: A Cross-Cultural Analysis on Racial Attention. IC2S2.

2021 I. Stewart. Assessing same-gender relationship bias in machine translation. Text as Data Conference.

2019 I. Stewart. What natural language processing should do for LGBTQ people. Queer in AI Workshop at NeurIPS. Vancouver, Canada.

2017 I. Stewart, J. Eisenstein. Social and Distributional Predictors of the Success of Lexical Innovations in Online Writing. New Ways of Analyzing Variation (NWAV). Madison, WI.

2017 I. Stewart, J. Eisenstein. #thighgap to #thyghgapp: Incrementation of Orthographic Variation on Instagram. Diversity and Variation in Language Conference (DiVar1). Atlanta, GA.

2015 I. Stewart. We some young kings: Communities, age, and African American English online. 2015 Annual Meeting of the American Dialect Society. Portland, Oregon.

Invited talks

Slides and materials available online.

2025 "Designing Robust and Domain-Specific Multimodal Foundation Models." UW Center for Statistics and the Social Sciences.

2022 "Linguistic Bias as a Window into Social Attitudes." SocialNLP Workshop (at NAACL).

2022 "People first: toward socially-aware NLP-powered writing tools." Pacific Northwest National Laboratory.

2021 "Topic modeling 101." Information Retrieval class at University of Michigan.

2020	"Exploratory text analysis for computational social science." CORE Congress 2020.
2020	"Through the language glass: What NLP can reveal about sociolinguistic variation." NLP@GT talk.
2019	"Language change tutorial." Presented at IC2S2 in Amsterdam.
2019	"Why do you talk like that? Minority languages and politics." Guest lecture for "Applications of Linguistics" course at Georgia Tech.
2018	"Is it social or linguistic? Examining internal factors in language change." Workshop on New Methods in Computational Sociolinguistics. Leiden, Netherlands.

Reviewing

* indicates outstanding reviewer.

2025	AAAI, NAACL, EMNLP, IEEE Transactions on Artificial Intelligence.
2024	ACL, NLP+CSS Workshop, Human-Centered LLM Workshop, EMNLP.
2023	ICWSM, IC2S2, Cross-Cultural NLP Workshop.
2022	ICWSM, Language Change workshop, ACL Rolling Review, EMNLP.
2021	CHI, ICWSM, EACL, NAACL, EMNLP, W-NUT.
2020	ICWSM, ACL, EMNLP*, W-NUT.
2019	ICWSM, NAACL*, ACL, Language Change workshop, W-NUT.
2018	ICWSM, NAACL, EMNLP, Workshop on Noisy User-generated Text (W-NUT).

Service

2021, 2024	NSF grant reviewer.
2022-2023	ICWSM Senior Program Committee.
2021-2022	Co-organizer for computational social science tutorial series ("NLP+CSS 201"): `https://nlp-css-201-tutorials.github.io/nlp-css-201-tutorials/`.
2021-2022	Judge for Undergraduate Research Opportunity Program Symposium (Michigan).
2019	Co-leader for Human-Centered Computing seminar (weekly discussion for PhD students).
2018-2019	Graduate Student Council Travel Fund Committee.

Mentoring

2025	Ian Meyer: undergraduate student (University of Nebraska); mathematics of AI.
2024	Ian Thomas: PhD Student (University of North Carolina); generative AI explanability.
2023	Seth Ireland: PhD Student (Colorado State University); deep learning model interpretability.
2022	Bontu Fufa: Undergraduate (Addis Ababa University); classification of emotion in text.
2021-22	Aman Choudhury: Undergraduate (University of Michigan); analysis of news media.
2021-22	Aylin Gunal: Undergraduate (University of Michigan); analysis of mental health writing.
2021	Dillan Pribak: Undergraduate (University of Michigan); text generation.
2021-22	Yiting Shen: Master's student (University of Michigan); analysis of bias in text.
2021-22	Gauri Kambhatla: Master's student (University of Michigan); detection of stereotypes in text.
2018	Xiaochuang Han: Undergraduate (Georgia Institute of Technology); deep learning for entity resolution.

Interests

- Music making
- Language learning

- Cookie baking
- Letter kerning