Brendan Kennedy

✉ bkennedy1704@gmail.com   🐦 @berendennedy

G Google Scholar

in https://www.linkedin.com/in/brendan-kennedy-nlp/

○ https://github.com/BrendanKennedy

Education

2016 – 2022	**Ph.D., Computer Science, University of Southern California** Los Angeles, California, USA Advisor: Morteza Dehghani, PhD Thesis: Balancing Prediction and Explanation in the Study of Language Usage and Speaker Attributes
2012 – 2016	**B.S. Computer Science, Gonzaga University** *Magna cum Laude* Minors: *Mathematics, Philosophy*

Experience

6/2022–*present*	**Machine Learning Engineer II,** Twitter Conducted large-scale experiments for content-based recommendation; developed a data pipeline for large-scale offline experimentation; contributed to model development for fine-tuning large-scale Transformer models.
6/2021–8/2021	**Machine Learning Engineering Intern,** Twitter Completed an independent project where I developed, experiment on, and helped push to production a new method for search retrieval that led to >= +1.25% social engagements, using the WikiData knowledge base to generate entity aliases
5/2019–5/2022	**Research Assistant,** Department of Psychology, USC Organized and directed compilation of new datasets for hate speech detection; conducted novel research into interpretable methods in NLP; conducted research in measuring and reducing bias in NLP systems
1/2018–8/2020	**Teaching Assistant,** Computer Science Department, USC Served as teaching assistant for introductory level programming courses, data structures and object-oriented programming, Artificial Intelligence, and Applied Machine Learning.
8/2016–12/2017	**Research Assistant,** Computer Science Department, USC (Information Sciences Institute) Performed data management and experiments with topic modeling for the use in event-centered forecasting using natural language indicators.
6/2015–9/2015	**Research Experience for Undergrads (REU),** DePaul University Conducted a focused research project using Association Rule Mining, focusing on categorizing potentially malignant nodules in lung CT scans.

Research Publications

Preprints, Under Review & Under Revision

1. **Kennedy**, **B.**, Golazizian, P., Trager, J., Atari, M., Hoover, J., Davani, A. M., & Dehghani, M. (2022). The (moral) language of hate. Under review at PNAS Nexus. 🔗 doi:10.31234/osf.io/eqp34

2. **Kennedy**, **B.**, Omrani, A., & Dehghani, M. (2022). Exemplar-based interpretations of speaker-language relationships. Under Revision.

3. Omrani, A., **Kennedy**, **B.**, Atari, M., & Dehghani, M. (2022). Social-group-agnostic word embedding debiasing via the stereotype content model. Under Review at EACL 2023.

4. **Kennedy**, **B.**, Reimer, N. K., & Dehghani, M. (2021). Explaining explainability: Interpretable machine learning for the behavioral sciences. 🔗 doi:10.31234/osf.io/9h6qr

5.   Mostafazadeh Davani, A., Atari, M., **Kennedy**, **B.**, & Dehghani, M. (2021a). Hate Speech Classifiers Learn Human-Like Social Stereotypes. Under Revision for Transactions of the ACL. arXiv: 2110.14839 [cs.CL]

Conference Proceedings

1.   Jin, X., Barbieri, F., **Kennedy**, **B.**, Davani, A. M., Neves, L., & Ren, X. (2021). On transferability of bias mitigation effects in language model fine-tuning. In *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 3770–3783).

2.   Mostafazadeh Davani, A., Omrani, A., **Kennedy**, **B.**, Atari, M., Ren, X., & Dehghani, M. (2021b). Fair Hate Speech Detection through Evaluation of Social Group Counterfactuals. In *The fifth workshop on online abuse and harms.*

3.   **Kennedy***, **B.**, Jin*, X., Mostafazadeh Davani, A., Dehghani, M., & Ren, X. (2020). Contextualizing hate speech classifiers with post-hoc explanation. In *Proceedings of the 2020 Annual Conference of the Association for Computational Linguistics* (pp. 5435–5442). Retrieved from
     🔗 https://www.aclweb.org/anthology/2020.acl-main.483/

4.   Mostafazadeh Davani, A., Atari, M., **Kennedy**, **B.**, Havaldar, S., & Dehghani, M. (2020). Hatred is in the Eye of the Annotator: Hate speech classifiers learn human-like social stereotypes. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

5.   Courtland, M., Mostafazadeh Davani, A., Reyes, M., Yeh, L., Leung, J., **Kennedy**, **B.**, … Zevin, J. (2019a). Modeling performance differences on cognitive tests using lstms and skip-thought vectors trained on reported media consumption. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)* (pp. 47–53). Retrieved from 🔗 https://www.aclweb.org/anthology/W19-2106.pdf

6.   Courtland, M., Mostafazadeh Davani, A., Reyes, M., Yeh, L., Leung, J., **Kennedy**, **B.**, … Zevin, J. (2019b). Subtle differences in language experience moderate performance on language-based cognitive tests. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society, Austin, Texas. Cognitive Science Society.* Retrieved from
     🔗 https://cogsci.mindmodeling.org/2019/papers/0280/0280.pdf

7.   Mostafazadeh Davani, A., Yeh, L., Atari, M., **Kennedy**, **B.**, Portillo-Wightman, G., Gonzalez, E., … Dehghani, M. (2019). Reporting the Unreported: Event extraction for analyzing the local representation of hate crimes. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCAI).*

8.   **Kennedy**, **B.**, Carrazza, M., Rasin, A., Furst, J., & Raicu, D. (2015). Applying association rule mining to semantic data in the lung image database consortium. In *Proceedings of the 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015* (pp. 463–471).

Journal Articles

1.   Atari, M., Davani, A. M., Kogon, D., **Kennedy**, **B.**, Saxena, N. A., Anderson, I., & Dehghani, M. (2022). Morally homogeneous networks and radicalism. *Social Psychological and Personality Science.* Retrieved from
     🔗 https://doi.org/10.1177/19485506211059329

2.   Atari, M., Reimer, N. K., Graham, J., Hoover, J., **Kennedy**, **B.**, Davani, A. M., … Dehghani, M. (2022). Pathogens are linked to human moral systems across time and space. *Current Research in Ecological and Social Psychology*, *3*, 100060. 🔗 doi:https://doi.org/10.1016/j.cresp.2022.100060

3.   Hossain, K., Harutyunyan, H., Ning, Y., **Kennedy**, **B.**, Ramakrishnan, N., & Galstyan, A. (2022). Identifying geopolitical event precursors using attention-based lstms. *Frontiers in Artificial Intelligence*, 236.

4.   **Kennedy**, **B.**, Atari, M., Mostafazadeh Davani, A., Yeh, L., Omrani, A., Kim, Y., … Dehghani, M. (2022). Introducing the Gab Hate Corpus: Defining and applying hate-based rhetoric to social media posts at scale. *Language Resources and Evaluation*, *1*, 79–108. Retrieved from 🔗 https://doi.org/10.1007/s10579-021-09569-x

5.   Reimer, N. K., Atari, M., Karimi-Malekabadi, F., Trager, J., **Kennedy**, **B.**, Graham, J., & Dehghani, M. (2022). Moral values predict county-level covid-19 vaccination rates in the united states. *American Psychologist*, *77*(6), 743.

6.   Hoover, J., Atari, M., Davani, A. M., **Kennedy**, **B.**, Portillo-Wightman, G., Yeh, L., & Dehghani, M. (2021). Investigating the role of group-based morality in extreme behavioral expressions of prejudice. *Nature Communications*, *12*(1), 1–13. Retrieved from 🔗 https://doi.org/10.1038/s41467-021-24786-2

7   **Kennedy**, **B.**, Atari, M., Mostafazadeh Davani, A., Hoover, J., Omrani, A., Graham, J., & Dehghani, M. (2021). Moral Concerns are Differentially Observable in Language. *Cognition*, *212*, 104696. Retrieved from
🔗 https://doi.org/10.1016/j.cognition.2021.104696

8   Hoover, J., Portillo-Wightman, G., Yeh, L., Havaldar, S., Mostafazadeh Davani, A., Lin, Y., … Dehghani, M. (2020). Moral Foundations Twitter Corpus: A Collection of 35k Tweets Annotated for Moral Sentiment. *Social Psychological and Personality Science*, *11*, 1057–1071. Retrieved from
🔗 https://doi.org/10.1177/1948550619876629

9   Garten, J., **Kennedy**, **B.**, Hoover, J., Sagae, K., & Dehghani, M. (2019). Incorporating Demographic Embeddings Into Language Understanding. *Cognitive Science*, *43*(1). 🔗 doi:10.1111/cogs.12701

10   Garten*, J., **Kennedy***, **B.**, Sagae, K., & Dehghani, M. (2019). Measuring the Importance of Context when Modeling Language Comprehension. *Behavior Research Methods*, *51*(2), 480–492. Retrieved from
🔗 https://doi.org/10.3758/s13428-019-01200-w

11   Hossain, K., Gao, S., **Kennedy**, **B.**, Galstyan, A., & Natarajan, P. (2018). Forecasting violent events in the Middle East and North Africa using the Hidden Markov Model and regularized autoregressive models. *Journal of Defense Modeling and Simulation*. Retrieved from 🔗 https://doi.org/10.1177/1548512918814698

Book Chapters

1   **Kennedy**, **B.**, Ashokkumar, A., Boyd, R., & Dehghani, M. (2022). Text analysis for psychology: Methods, principles, and practices. In M. Dehghani & R. Boyd (Eds.), *The Handbook of Language Analysis in Psychology*. Guilford Press. Retrieved from 🔗 https://doi.org/10.31234/osf.io/h2b8t

Skills

Programming	🔖	*Expert (5+ years experience)*: Python, R, SQL
		Capable: C++, Java, PHP, JavaScript
ML & NLP	🔖	*Tools*: Google Cloud Platform (Big Query, Dataflow), Kubeflow & Kubernetes
		Python libraries: pytorch, pandas, numpy, scikit-learn, nltk, spacy, gensim, tensorflow, keras, pytorch, pytorch lightning, transformers, optuna, matplotlib, seaborn
		Technical Expertise: Text classification, explainability & interpretability techniques, model evaluation, hyperparameter tuning, named entity recognition, relation extraction, language model fine-tuning, language model debiasing, word embedding training and evaluation, inferential statistics, Bayesian statistical methods, dataset compilation and curation, clustering heterogeneous data
R packages	🔖	tidyverse stats lme4 MASS
Miscellaneous Skills	🔖	LaTeX, Git, Vim, MS Office, Google Suite, Markdown, Beamer

Software and Data Resources

Neural Text Analysis Pipeline (NTAP)	🔖	Lead designer & developer of a Python package for high-level interface for applying modern NLP techniques for text analysis (https://pypi.org/project/ntap/).
Gab Hate Corpus (GHC)	🔖	Lead author for defining "hate-based rhetoric" and typology used for labeling 27,000 social media posts (https://osf.io/edua3/).

Teaching Assistantships

Applied Machine Learning	🔖	Grading, lectures, and content creation. Material: ISLR and Deep Learning with R (*Fall 2020*)
Data Structures & Object Oriented Design	🔖	Object-oriented principles, recursion, search trees, graph algorithms in C++ (*Summer 2020*)

Teaching Assistantships (continued)

Artificial Intelligence	▌	Search, knowledge representation, logic to computer science upper-classmen (*Spring 2019*)
Intro to Programming	▌	Programming concepts and object-oriented programming in C++ (*Spring 2018, Fall 2018*)

References

Available on Request