

# Joe Hoover, PhD

+1 (571) 485-9240 • [joe Hoover88@gmail.com](mailto:joe Hoover88@gmail.com) • [joe-hoover.com](http://joe-hoover.com)  
[/in/JoeEHoover/](https://in/JoeEHoover/)

## Profile

---

I help teams and organizations research, develop, and deploy AI systems. I've founded and lead high-performing NLP teams in both early-startup and enterprise contexts. As a leader and individual contributor, I've shipped a wide range of ML systems, with a primary focus on putting large language models in production. I thrive under pressure and am driven by a love for building innovative technology that solves hard problems.

## Experience

---

### Associate Director of Machine Learning

(Remote) New York, NY

*Engineering Solutions, S&P Global*

2022-2023

- Directed a full-stack applied NLP/ML/Search team in an innovative SaaS product unit
- Lead development of generative AI road map
- Mentored team members through legacy code audits, refactors, architecture proposals, and prototyping of information extraction and semantic search services
- Coordinated with executive leadership, Product, customer stakeholders, and other technical teams to establish product direction and specify technical road maps
- Collaborated with other ML teams to architect and implement MLOps platform

### Sr. Applied Scientist

(Remote) Cincinnati, OH

*AI Team, Mantium*

2021-2022

- Lead 4 person AI Team and collaborated with C-Suite and Product to help set direction and priorities
- Lead research on Large Language Models and applied NLP with a primary focus on extracting complex, multi-modal information from structured documents (e.g. PDFs); Question-Answer systems; and efficient methods for training and deploying large neural networks
- Oversaw productionization and deployment of large-scale (11B parameter) NLP models and pipelines
- Mentored junior and senior team members
- Lead AI Team hiring committee and on-boarding

### Computational Social Scientist

(Remote) Herndon, VA

*Machine Intelligence Research and Applications, Ntrepid*

2020-2021

- Developed a named-entity directed sentiment system for modeling social media users' sentiment toward entities over time. Involved fine-tuning PyTorch Transformer models, linking entities via semantic similarity, and modeling user preferences with a Bayesian autoregressive Beta model
- Owned end-to-end NLP/ML development, from prototyping to production deployment
- Functioned as in-house expert on psychological measurement and group dynamics
- Communicated with internal stakeholders to specify, develop, scope, and deliver NLP and ML products
- Presented technical innovations and capabilities to non-technical audiences, particularly through the development of dashboards

### Postdoctoral Research Fellow

Evanston, IL

*Kellogg School of Management, Northwestern University*

2019-2020

- Developed hierarchical, semantic role labeling annotation schema to support quantifying expressions of dehumanization in natural language texts
- Assembled and managed a 5-person team of annotators, selected, deployed, and maintained annotation software and conducted QA to ensure annotation quality
- Integrated ML, NLP, and experimental and observational research methods to study human behavior and online social systems
- Delivered reports, statistical analyses, and data visualizations

### Graduate Research Fellow

Los Angeles, CA

*Computational Social Sciences Lab, University of Southern California*

2016,2017,2019

- Developed novel uses of embeddings, deep neural networks, and statistical models for the study of human psychology and behavior
- Used LSTMs and time-series modeling to investigate the relationship between online moral outrage and violent protest (funded by Army Research Lab)
- Developed and published a 35k corpus of Tweets annotated for moral sentiment
- Contributed to development of TACIT, an open-source text analysis, crawling, and interpretation tool

### Graduate Teaching Fellow

Los Angeles, CA

Department of Psychology, University of Southern California

2018

- TAd Introduction to Statistics and Experimental Methods

### Data Science Consultant

Remote

Independent Contractor

March, 2017-2018

- Developed Structural Topic Model to model language changes in corpus of Federal documents
- Designed system for social-media based market segmentation for Vans

### Data Science Intern

Los Angeles, CA

Ranker

June, 2017-August, 2017

- Developed time-series models (ARIMAX) for dynamic Google AdSense ad pricing
- Developed Slackbot information recall system for embeddings-based similarity queries

### Graduate Research Fellow/TA

Eugene, OR

University of Oregon

Sept, 2014-May, 2015

- Developed hierarchical structural equation models
- Helped develop nuanced understanding of academic intervention targeting Native American students
- TAd Graduate Statistics in Education Department

## Education

### University of Southern California

Los Angeles, CA

PhD, Psychology, Specialization in Computational Social Science

2015-2019

### University of Oregon, OR

Eugene, OR

MS, Psychology

2013-2015

### Central Connecticut State University

New Britain, CT

BA, English

2008-2010

## Skills & Experience

**Model development:** PyTorch, Transformers, PyTorch Lightning, DeepSpeed, SageMaker Distributed Model Parallel, NumPyro, Stan

**MLOps:** Docker, CI/CD, ONNX, TensorRT, Triton, AWS SageMaker, Kubeflow, Hydra, Ray

**Databases:** SQL, PostgreSQL, Elasticsearch, DynamoDB, MongoDB

**General SWE:** Agile, testing, architecture review, code review, git, SOLID

## Open Source Software

- **DDR** Python implementation of Distributed Dictionary Representations, developed in Garten, Hoover, Boghrati, Iskiwitch, & Dehghani (2017) and Hoover, Johnson, Bohgrhati, Graham, & Dehghani (2018)
- **ymProcessing** R package for geolocating survey data collected via YourMorals.org

## Selected Publications

### Journal Articles

Kennedy, B., Atari, M., Davani, A., ..., **Hoover, J.**, ..., & Dehghani, M. (2022). Introducing the gab hate corpus: Defining and applying hate-based rhetoric to social media posts at scale. *Language Resources and Evaluation*.

**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*.

- Hoover, J., & Dehghani, M.** (2019). The big, the bad, and the ugly: Geographic estimation with flawed psychological data. *Psychological Methods*.
- Hoover, J., Portillo-Wightman, G., Yeh, L., & et al.** (2019). Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*.
- Hoover, J., Johnson, K. M., Bohgrhati, R., Graham, J., & Dehghani, M.** (2018). Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra Psychology*.
- Hoover\*, J., Mooijman\*, M., Lin, Y., Ji, H., & Dehghani, M.** (2018). Moralization in social networks and the emergence of violent protests. *Nature Human Behavior*.
- Garten, J., Hoover, J., M, J. K., Boghrati, R., Iskiwitch, C., & Dehghani, M.** (2017). Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis. *Behavior Research Methods*.
- Hoover, J., Horton, K., & Dehghani, M.** (2017a). The meanings of morality: Investigating the psychometric properties of distributed representations of latent moral concepts. *In the proceedings of 38th Annual Conference of the Cognitive Science Society (CogSci) Philadelphia, PA*.
- Iliev, R., Hoover, J., Dehghani, M., & Axelrod, R.** (2016). Linguistic positivity in historical texts reflects dynamic environmental and psychological factors. *Proceedings of the National Academy of Sciences*, 113(49), E7871–E7879.

## Conference Proceedings.....

- Lin, Y., Hoover, J., Portillo-Wightman, G., Park, C., Dehghani, M., & Ji, H.** (2018). Acquiring background knowledge to improve moral value prediction. *Proceedings of IEEE/ACM 2018 International Conference on Advances in Social Networks Analysis and Mining, Barcelona, Spain*.
- Hoover, J., Horton, K., & Dehghani, M.** (2017b). The meanings of morality: Investigating the psychometric properties of distributed representations of latent moral concepts. *In the proceedings of 38th Annual Conference of the Cognitive Science Society (CogSci) Philadelphia, PA*.
- Boghrati, R., Hoover, J., Garten, J., & Dehghani, M.** (2016). Syntax accommodation in social media conversations. *In the proceedings of 37th Annual Conference of the Cognitive Science Society*.
- Garten, J., Boghrati, R., Hoover, J., Johnson, M., K, & Dehghani, M.** (2016). Morality between the lines: Detecting moral sentiment in text. *Proceedings of IJCAI 2016 workshop on Computational Modeling of Attitudes*.

## Selected Invited Talks

<b>Train and deploy large language models on Amazon SageMaker</b> AWS re:Invent	<b>Las Vegas, NV</b> January, 2022
<b>Investigating the role of group-based morality in acts of hate</b> Kellogg School of Business, Northwestern University	<b>Evanston, IL</b> January, 2020
<b>Investigating the role of group-based morality in acts of hate</b> Crockett Lab, Department of Psychology, Yale University	<b>New Haven, CT</b> January, 2019
<b>Moralization in social networks and violence during protests</b> Wharton Behavioral Insights from Text Conference	<b>Philadelphia, PA</b> January, 2019

## Facilitated Workshops

<b>A Modern Intro to Theory-Driven Text Analysis</b> Soc. for Personality and Social Psychology Conf.	<b>New Orleans, LA</b> February, 2020
<b>Geospatial modeling for Personality and Social Psychology</b> Soc. for Personality and Social Psychology Conf.	<b>Atlanta, GA</b> February, 2018