ALEXANDER SPANGHER

Curriculum Vitae

Phone: (631) 487-7777
Email: alexspan@stanford.edu
Website: **https://alexander-spangher.com**

RESEARCH OBJECTIVES

I work in the intersection of NLP and Computational Journalism. I develop novel AI-driven methods to study and enhance the entire process of human communicative acts: (1) before (i.e. Creative Planning), (2) during (i.e. Structured Generative AI) and after (i.e. Misinformation and Recourse).

EDUCATION

B.S.	**Columbia University** Bachelor of Science in Neuroscience Bachelor of Science in Computer Science	*2010-2014*
M.S.	**Columbia University**[1] Master of Science in Data Science Master of Science in Journalism	*2014-2018*
Ph.D.	**University of Southern California**[2] Graduate Studies, Computer Science	*2019-2025*
Postdoctoral Scholar	**Stanford University** Stanford Artificial Intelligence Lab (SAIL) Fellowship Human-Centered Artificial Intelligence (HAI) Fellowship	*2025-*

EMPLOYMENT

The New York Times, Data Scientist	*2014-2018*
Microsoft Research, Research Intern	*2018*
Carnegie Mellon University, Research Assistant	*2018-2019*
Stanford University, Summer Research Assistant	*2019*
Bloomberg LP, Research Intern	*2020-2024*

PUBLICATIONS

Peer-Reviewed Publications [3]

1. 🏅🧍 **Alexander Spangher**, Nanyun Peng, Sebastian Gehrmann, Mark Dredze. Do LLMs Plan Like Human Writers? Comparing Journalistic Coverage of Press Releases with LLMs. *In the 2024 Conference on Empirical Methods in Natural*

[1] Masters degrees pursued part-time while working full-time at the *New York Times*.
[2] Transferred from Carnegie Mellon University (start date: 2018).
[3] 🏅 indicates paper received an award. 🧍 indicates paper was selected for an oral presentation.

Language Processing. 2024.
Acceptance rate 20.8%

2. **Alexander Spangher**, James Youn, Matt DeButts, Nanyun Peng, Jonathan May. 2024. Explaining Mixtures of Sources in News Articles. *Findings of the 2024 Conference on Empirical Methods in Natural Language Processing*. 2024.
Acceptance rate 20.8%

3. **Alexander Spangher**, Guillaume Sanchez, Honglu Fan, Elad Levi, and Stella Biderman. Stay on Topic with Classifier-Free Guidance. *In Forty-first International Conference on Machine Learning*. 2024.
Acceptance rate 27.5%

4. **Alexander Spangher**, Emilio Ferrara, Ben Welsh, Nanyun Peng, Serdar Tumgoren, and Jonathan May. Tracking the Newsworthiness of Public Documents. In the 2024 Association of Computational Linguistics. 2024.
Acceptance rate 23.5%

5. **Alexander Spangher**, Zihan Xue, Te-Lin Wu, Mark Hansen, Jonathan May. LegalDiscourse: Interpreting When Laws Apply and To Whom. *In the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* 2024.
Acceptance rate 23%

6. **Alexander Spangher**, Nanyun Peng, Emilio Ferrara, and Jonathan May. Identifying Informational Sources in News Articles. *In the 2023 Conference on Empirical Methods in Natural Language Processing*. 2023
Acceptance rate 23.3%

7. **Alexander Spangher**, James Youn, Jonathan May and Nanyun Peng. First Steps Towards a Source Recommendation Engine: Investigating How Sources Are Used in News Articles. *Computation + Journalism*. 2023.

8. **Alexander Spangher**, Xiang Ren, Jonathan May and Nanyun Peng. *NewsEdits*: A News Article Revision Dataset and a Novel Document-level Reasoning Challenge. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2022.
Acceptance rate 26%

9. **Alexander Spangher**, Jonathan May, Sz-rung Shiang, Lingjia Deng. Multi-task Semi-Supervised Learning for Class-Imbalanced Discourse Classification. *The 2021 Conference on Empirical Methods in Natural Language Processing*. 2021.
Acceptance rate 25.6%

10. **Alexander Spangher** and Jonathan May. *StateCensusLaws.org*: A Web Application for Consuming and Annotating Legal Discourse Learning. *Computation + Journalism*. 2022.

11. **Aleander Spangher** and Jonathan May. News Discourse Patterns: A Roadmap for Computational Journalism. *Computation + Journalism*. March 2021.

12. **Alexander Spangher**, Nanyun Peng, Jonathan May, Emilio Ferrara. Enabling Low-Resource Transfer Learning across COVID-19 Corpora by Combining Event-Extraction and Co-Training. *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020.* 2020.

13. **Alexander Spangher**, Nanyun Peng, Jonathan May, Emilio Ferrara. Don't quote me on that: Finding Mixtures of Sources in News Articles. *Computation + Journalism*. 2020

14. **Alexander Spangher**, Nanyun Peng, Jonathan May, Emilio Ferrara. Modeling Newsworthiness for Lead-Generation Across Corpora. *Computation + Journalism.* 2020. *Southern California Natural Language Processing Conference.* 2019.
15. **Alexander Spangher**, Gireeja Ranade, Besamira Nushi, Adam Fourney, Eric Horvitz. Characterizing Search Engine Traffic to Internet Research Agency Web Properties. *The Web Conference 2020.* Taipei, Taiwan. 2020.
16. **Alexander Spangher**, Jia Zhang, Rahul Ramachandran, Manil Maskey, Patrick Gatlin, J.J. Miller, Sundar Christopher. Methodology for Building Scalable Knowledge Graphs using Pre-existing NASA Ontologies. *IEEE Transactions on Cognitive Communications and Networking.* 2019.
Acceptance rate 19.4%
17. **Alexander Spangher**, Berk Ustun. Actionable Recourse in Linear Classification. *Proceedings of the 5th Workshop on Fairness, Accountability and Transparency in Machine Learning, ICML.* 2018. Also presented in: *Workshop on Ethical, Social and Governance Issues in AI, 2018, NIPS.* 2018.

Publications (cont.)

18. Yufei Tian, Tenghao Huang, Miri Liu, Derek Jiang, **Alexander Spangher**, Muhao Chen, Jonathan May, Nanyun Peng. Are Large Language Models Capable of Generating Human-Level Narratives? 2024. *In the 2024 Conference on Empirical Methods in Natural Language Processing..*
Acceptance rate 20.8%
19. Lucas Spangher, Matteo Bonotto, William Arnold, Dhruva Chayapathy, Tommaso Gallingani, **Alexander Spangher**, Francesco Cannarile, Daniele Bigoni, Eliana De Marchi, Cristina Rea. DisruptionBench: A Robust Benchmarking Framework for Machine Learning-Driven Disruption Prediction. *Journal of Fusion Energy.* 2024. Acceptance rate 20%
20. Lucas Spangher, William Arnold, **Alexander Spangher**, Andrew Maris, Cristina Rea. Autoregressive Transformers for Disruption Prediction in Nuclear Fusion Plasmas. *NeurIPS 2023 Workshop: Machine Learning and the Physical Sciences.* 2023.
Acceptance rate 75.6%
21. Te-Lin Wu, **Alexander Spangher**, Pegah Alipoormolabashi, Marjorie Freedman, Ralph Weischedel, Nanyun Peng. Understanding multimodal procedural knowledge by sequencing multimodal instructional manuals. 2023. *In the 61st Annual Meeting of the Association for Computational Linguistics.*
Acceptance rate 24.1%
22. Berk Ustun, **Alexander Spangher**, Yang Liu. Actionable Recourse in Linear Classification. (Expanded Version). *In the Conference on Fairness, Accountability and Transparency (FAT*), 2019, ACM.* 2019.
Acceptance rate 27%
23. Ryan L Boyd, **Alexander Spangher**, Adam Fourney, Besmira Nushi, Gireeja Ranade, James Pennebaker, Eric Horvitz. Characterizing the Internet Research Agency's Social Media Operations During the 2016 US Presidential Election using Linguistic Analyses. *Whitepaper.* 2019.

In Submission

24. **Alexander Spangher**, Kung-Hsiang (Steeve) Huang, Hyundong Justin Cho and Jonathan May. NewsEdits 2.0: Learning the Intentions Behind Updating News. *Proceedings of the 2025 Conference of the North American Chapter of*

the Association for Computational Linguistics: Human Language Technologies. 2025.

25. **Alexander Spangher**, Tenghao Huang, Yiqin Huang, Liheng Lai, Lucas Spangher, Sewon Min, Mark Dredze. A Novel Multi-Document Retrieval Benchmark Grounded on Journalist Source-Selection in Newswriting. *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2025.
26. **Alexander Spangher**, Michael Lu, Hyundong Justin Cho, Weiyan Shi, Jonathan May. NewsInterview: a Dataset and a Playground to Evaluate LLMs' Ground Gap via Informational Interviews. *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2025.
27. Ben Welsh, Naitian Zhou, Arda Kaz, Michael Vu, **Alexander Spangher**. NewsHomepages: Homepage Layouts Capture Information Prioritization Decisions. *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2025.
28. Ryan Lee, **Alexander Spangher**, Xuezhe Ma. PatentEdits: Framing Patent Novelty as Textual Entailment. *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2025.

Newspaper Articles and Graphics (Selected)

1. **Alexander Spangher**. Building the Next New York Times Recommendation Engine. *The New York Times*. https://nyti.ms/2zpGG5g
2. **Alexander Spangher**. How Does This Article Make You Feel? Using data science to predict the emotional resonance of New York Times articles for better ad placement. *The New York Times*. https://nyti.ms/2PyHkcn.
3. **Alexander Spangher**. What the Paris attacks tell us about how foreign news gets made. *Columbia Journalism Review*. https://bit.ly/2DIV2TH
4. **Alexander Spangher**. 19 Countries, 43 States, 327 Cities: Mapping The Times's Election Coverage. *The New York Times*. https://nyti.ms/2ScnEbV
5. **Alexander Spangher**. Eye on the Prize: 100 years worth of Pulitzer Prize Winners by Race, Gender and Location. *Columbia Journalism Review*. https://bit.ly/2r2YEIT
6. **Alexander Spangher**. 3 Smart Data Journalism Techniques that can help you find stories faster. *Medium*. https://bit.ly/2DIUydH.
7. For more articles and graphics, see: alexander-spangher.com/data-vis.html

FUNDING and HONORS

Fellowship Funding (Total: $1,027,000)

Stanford AI Lab (SAIL) Fellowship $85,000	*2025-2026*
Stanford Human-Centered AI (HAI) Fellowship $85,000	*2026-2027*
Bloomberg Data Science PhD Fellowship. $550,000[4]	*2020-2024*
Annenberg Graduate Fellowship Symposia. $6,000.	*2019-2024*

[4]Included in this sum is $410,000 in fellowship funding ($60,000 in tuition, $35,000 in stipend, $5,000 in travel funding each year for over 4 years) and $110,000 in internship funding and $30,000 in gift funding to my advisor. Internships were a required part of the fellowship.

Annenberg Graduate Fellowship $80,000. *2019*

Carnegie Mellon Graduate Student Fellowship. $80,000. *2018*

Columbia School of Journalism Scholarship. $78,000. *2016-2017*

***New York Times* Tuition Scholarship.** $32,000. *2014-2018*

John Jay Scholar Summer Funding. $20,000 *2011-2012*

Intel STS Semi-finalist. $10,000. *2009*

National Merit Scholarship. $1,000. *2009*

Awarded Grants (Total $915,544):

1. Long-Context Natural Language Policy Understanding *2024*
   PI: Jonathan May, professor, Information Sciences Institute, University of Southern California
   MITRE Research Grant. **$230,000**.
   Role: I was the primary grant-writer. My work on *LegalDiscourse* formed the basis of the grant.
2. Multi-News: Debiasing International Angles in Newswriting *2024*
   PI: Monica Lam, professor, Stanford University
   Magic Grant, Brown Institute for Media Innovation. **$70,000**
   Role: I was the primary grant writer. My research on *news discourse structure* was cited in the grant.
3. Deploying Classifier Free Guidance for LLMs-as-Scientists Planning *2024*
   PI: James Aung, PhD, OpenAI
   OpenAI Research Credits Grant **$10,000**
   Role: I was the primary grant writer. I also contributed to the grant by designing rubrics and schemas for evaluation. My research on *Classifier Free Guidance* was cited.
4. News Discourse to Improve Student Journalist Writing *2024*
   USC Annenberg Dean's Faculty Research Fund. **$10,000**
   Role: I contributed substantially to writing the grant. My research on *news discourse structure* formed the basis of the grant proposal.
5. Source Recommendations as a Service *2023*
   PI: **Alexander Spangher**, PhD
   Google Research Credits Grant. **$5,000**
   Role: I substantially wrote and edited this grant, and it was based on my work.
6. Computational Approaches to Explaining Computational Journalism *2019*
   PI: **Alexander Spangher**, PhD
   Google Research Credits Grant. **$5,000**
   Role: I substantially wrote and edited this grant, and it was based on my work.

Paper Awards

1. EMNLP Outstanding Paper Award, 2024:
   Alexander Spangher, Nanyun Peng, Sebastian Gehrmann, Mark Dredze. Do LLMs Plan Like Human Writers? Comparing Journalistic Coverage of Press Releases with LLMs. *In the 2024 Conference on Empirical Methods in Natural Language Processing*. 2024.
   Statistics: 1 of 30 papers selected for awards in this conference, out of 1,271 accepted papers and 6,105 submissions.
   Documentation:
   - https://x.com/emnlpmeeting/status/1857173122598010918

2. EMNLP Outstanding Paper Award, 2024:
Yufei Tian, Tenghao Huang, Miri Liu, Derek Jiang, **Alexander Spangher**, Muhao Chen, Jonathan May, Nanyun Peng. Are Large Language Models Capable of Generating Human-Level Narratives? 2024. *In the 2024 Conference on Empirical Methods in Natural Language Processing.* https://arxiv.org/pdf/2407.13248.
Statistics: 1 of 30 papers selected for awards in this conference, out of 1,271 accepted papers and 6,105 submissions.
Documentation:
   - https://x.com/emnlpmeeting/status/1857173122598010918
3. NAACL Outstanding Paper Award 2022:
Alexander Spangher, Xiang Ren, Jonathan May and Nanyun Peng. *NewsEdits*: A News Article Revision Dataset and a Novel Document-level Reasoning Challenge. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* 2022. https://aclanthology.org/2022.naacl-main.10.pdf.
Statistics: 1 of 8 papers selected for awards in this conference, out of 442 accepted papers and 2103 submissions.
Documentation:
   - https://2022.naacl.org/blog/best-papers/
   - https://x.com/naaclmeeting/status/1542337649007923201
   - https://www.bloomberg.com/company/stories/bloomberg-publishes-3-ai-papers-naacl-2022/
4. ICML Spotlight Award, 2024:
Alexander Spangher, Guillaume Sanchez, Honglu Fan, Elad Levi, Stella Biderman. Stay on Topic with Classifier-Free Guidance. *ICML*. 2024.
Statistics: Top 3.5% of papers are awarded with a Spotlight Award.
Documentation: https://icml.cc/virtual/2024/events/2024SpotlightPosters.
5. C+J Best Paper Award, 2023:
Alexander Spangher, James Youn, Jonathan May and Nanyun Peng. First Steps Towards a Source Recommendation Engine: Investigating How Sources Are Used in News Articles. *Computation + Journalism.* 2023.
Statistics: 1 of 4 papers selected for awards at this conference, out of 40 accepted papers.
Documentation: None available online.
Conference website: https://www.datajconf.com/

EXPERIENCE

Research Experience

Bloomberg LP, New York City, NY *Research Intern* *2019-present*
Advisor: Mark Dredze, Sebastian Gehrmann, Amanda Stent, Sz-Rung Shiang, Yao Ming, Lingja Deng.
Discourse Analysis in Journalism.
- Development of multitask machine learning models for discourse analysis.
- Analysis of semi-supervised data augmentation techniques for sequence classification.

University of Southern California, Los Angeles, CA *Ph.D. Student 2019-present*
Advisor: Jonathan May, Emilio Ferrara, Nanyun Peng.
Discourse Analysis in Journalism and Law.

- Development of machine learning models for discourse-tagging sequences.
- Development of discourse schemas for characterizing news articles edits, source-inclusion and legal text.
- Collection of large-scale corpora with legal and journalistic value for computational linguistic analysis.

Stanford University, Palo Alto, CA *Research Assistant* *2019*
Advisor: James Hamilton.
Computational Journalism Field Survey.

- Practitioner interviews (51 interviews) to gain knowledge on the current innovations, methodologies and challenges relevant to journalists.
- Word-embeddings classification of U.S. patent databases and NSF grants.
- Visualization/tooling in Python/Flask, D3.js and Google App-Engine.

Carnegie Mellon Univ., Mountain View, CA *Research Assistant* *2018-2019*
Advisor: Jia Zhang.
Knowledge Graph Construction for NASA Earth Sciences.

- Text-modeling using hierarchical topic modeling to improve and model existing NASA concept-ontologies.
- Text-matching and word-modeling using custom lexical parsing rules to extract datasets, variables and methods from papers.
- Visualization/tooling in D3.js with an emphasis on interpretability and ease of capturing user feedback.

Microsoft Research, Redmond, WA *Research Intern* *Summer 2018*
Advisors: Gireeja Ranade, Adam Fourney, Besamira Nushi, Eric Horvitz.
Large-scale analysis of user-behavior changes in response to misinformation

- Data analysis merging data from Facebook, Twitter and Microsoft. Causal modeling using counterfactual analysis.
- Text-modeling using TF-IDF to track search query changes over time.
- Intensive fact-checking, informal Congressional briefing, contact with staff of Congressman Adam Schiff (Representative, D-CA 28th District).

Employment

New York Times, NYC, NY *Data Scientist* *2014-2018*
Advisors: Chris Wiggins, Jose Muanis Castro, Thompson Marzagao.
Collaborative Topic Models for Article Recommendations:

- Created an improved article recommendation-engine by building a topic model to incorporate information from article-text and user clicks. Scale to millions of users and provide recommendations in real-time.
- Modeling: Custom-designed Bayesian model that extends Latent Dirichlet Allocation, coded in C++.
- Collaboration: Dr. David Blei, Jake Hoffman, and Prem Gopalan, Columbia University. See https://arxiv.org/pdf/1311.1704.pdf.
- Deployment: MySQL, WSGI API, Luigi data pipelines.
- Extensions: multi-armed bandit and contextual bandit algorithms.

Project Feels:

- Modeled different emotions in *New York Times* article body text. The purpose was to predict tragic, happy and polarizing articles for downstream decision-making.
- Modeling: 7 different deep learning architectures were tested alongside ensemble methods and other linear methods.
- Data collection: Crowd-sourcing on Amazon Mechanical Turk, using active learning to select successive batches of articles to label.
- Deployment: Google Cloud Services (GKE, Datastore, BigQuery).

Newsroom Tools and Other:

- Used Latent Dirichlet Allocation and TF-IDF to build a related-articles feature for journalists doing research, directly into their publishing platform.
- Used simple character-level modeling and K-Means to cluster text-messages journalists received from Q&A sessions with readers to facilitate responding.
- Used custom Bayesian model to perform newsletter recommendations for users.
- Used Random Forest and simple decision trees to create powerful and interpretable models of user retention likelihood.

Open Source Code

Classifier Free Guidance for **Huggingface** and **EleutherAI**

- Implements contrastive logit manipulations to steer language modeling outputs (see publication https://arxiv.org/pdf/2306.17806, listed above).
- *Contributors:* Guillaume Sanchez, Honglu Fan, **Alexander Spangher**. https://github.com/huggingface/transformers/issues/24536.
- Huggingface is the standard open-source framework for large language models (LLMs), model improvements, and model sharing. EleutherAI is an open-science research lab dedicated to training and improving LLMs.

Actionable Recourse Implementation for **IBM Fairness 360 Project**.

- Implements Mixed-Integer Program (MIP), using CPLEX and a Pyomo based optimizer for providing actionable recourse auditing (see publication https://bit.ly/48GLLrV, listed above.)
- *Contributors:* Berk Ustun, **Alexander Spangher**. https://github.com/ustunb/actionable-recourse
- The IBM Fairness 360 project is an open-sourced project integrating different fairness and transparency algorithms.

Broca: A Battery of Natural Language Processing Methods for **Mozilla Knight OpenSource News Fellows**

- A pipeline system of organize a sequence of text-transformations. Automatic intermediate caching for time-saving and debugging.
- *Contributors:* Francis Tseng, **Alexander Spangher**. https://github.com/frnsys/broca.
- The Mozilla Knight OpenSource News Fellows program brings together coders and data journalists to create tools for jouranlism. *Blog:* Francis Tseng. Introducing Broca. https://bit.ly/2DVdrwH.

PRESENTATIONS and CONFERENCES

Invited Talks (Selected)

1. **Alexander Spangher**, Ke Isherwood-Huang and Amber Lynn Scott. "What the Diff': The Narrative of a Coronavirus Crisis in the U.S. Navy. *Symposium talk. USC Annenberg Research and Creative Project Symposium.* April 2021. https://www.instagram.com/p/CNsUlUflYV0/.
2. **Aleander Spangher**, Jillian Kwong. The Effects of a 2020 Census Undercount: A Computational Law Approach. *Symposium talk. USC Annenberg Research and Creative Project Symposium.* April 2020. https://www.instagram.com/p/B_VijsdFZwu/.
3. **Alexander Spangher**, Nanyun Peng, Jonathan May, Emilio Ferrara. Modeling Newsworthiness for Lead-Generation Across Corpora. *Symposium Talk. Southern California Natural Language Processing Conference.* 2019. https://arxiv.org/abs/2104.09653.
4. **Alexander Spangher**. Project Feels and Actionable Recourse. *Keynote. Open Data Science Conference. San Francisco, California.* 100+ attendees. October 2018.
5. Adam Grant, **Alexander Spangher**. *Interview. New York Times* Young Professionals Interview Series, New York City, NY. 100+ attendees. February 20th, 2018.
6. Nicholas Kristof, **Alexander Spangher**, Hannah Cassius. *Interview. New York Times* Young Professionals Interview Series, New York City, NY. 100+ attendees. March 18th, 2017.
7. **Alexander Spangher**. Project Feels: Deep Text Models for Predicting the Emotional Resonance of *New York Times* Articles. Open Data Science Conference. Boston, Massachusetts. 150+ attendees. April 30th-May 3rd, 2018.
8. **Alexander Spangher**, Adam Kelleher. Recommender Systems in Digital Media. *Keynote. DataEngConf.* New York City. 175+ attendees. March 15th, 2018.
9. **Alexander Spangher**. Building the Next *New York Times* Recommendation Engine. *Keynote. DataEngConf.* New York City, NY. 100+ attendees. December 13th-15th, 2015.

Guest Lectures (Selected)

1. University of Southern California, Viterbi School of Engineering. Class: Advanced Natural Language Processing. Professor: Jonathan May. Nov. 20, 2024. *Talk Title: "Discourse in Natural Language Processing".*
2. University of Southern California, Annenberg School of Communication and Journalism. Class: Data Science for Communication and Social Networks. Professor: Luca Luceri. Sept. 26th, 2024. *Talk Title: "Overview of Probabilistic Topic Models".*
3. University of Southern California, Annenberg School of Communication and Journalism. Class: Seminar in Social Analysis: Journalism and the Shifting Information Landscape. Professor: Laura Davis. April 16th, 2024. *Talk Title: "Computational Approaches to Journalism".*
4. Stanford University, Department of Communication. Class: Exploring Computational Journalism. Professor: Serdar Tumgoren. Feb. 13th, 2024. *Talk Title: "Computational Approaches to Covering Local News".*

5. Stanford University, Department of Communication. Class: Investigative Watchdog Reporting. Professor: Cheryl Phillips. May 3rd, 2023. *Talk Title: "NLP Methods for Local Journalism".*
6. Cornell Tech, School of Information Science. Class: Data-Driven Product Development. Professor: Mor Namaan. May 20th, 2018. *Talk Title: "Project Feels at the New York Times: Tools for Advertisers in Journalism"*
7. Columbia University, Graduate School of Journalism. Class: Data Journalism. Professor: Jonathan Stray. Feb. 20th, 2018. *Talk Title: "Project Feels: Neural Networks for Interpreting News"*
8. Columbia University, Graduate School of Journalism. Class: Investigative Journalism. Professor: Steven Coll. Dec. 4th, 2017. *Talk Title: "NLP Techniques for Investigative Journalism"*
9. Columbia University, Graduate School of Journalism. Class: Data Journalism. Professor: Jonathan Stray. Feb. 14th, 2017. *Talk Title: "Computational Methods at the New York Times"*
10. Columbia University, Graduate School of Journalism. Class: Data Journalism. Professor: Jonathan Stray. Feb. 2nd, 2015. *Talk Title: "Recommendation Systems in Journalism"*

Teaching Assistantships

1. Jonathan May, Advanced Natural Language Processing, University of Southern California, Fall 2024.
2. Mohammad Reza Rajati, Machine Learning for Data Science, University of Southern California. Summer 2024.
3. Mark Core, Applied Natural Language Processing, University of Southern California. Spring 2020.
4. Francis Champagne, The Developing Brain. Department of Psychology, Columbia University. Spring 2012.

Reviewer

1. Association of Computational Linguistics, Rolling Review. 2024.
2. Association of Computational Linguistics, Rolling Review. 2023.
3. Association of Computational Linguistics, Rolling Review. 2022.
4. Association of Computational Linguistics. 2021.
5. The Web Conference. 2020.
6. Association of Computational Linguistics. 2020.
7. Automated Knowledge Base Construction 2019 Conference. Amherst, Massachusetts, May 20th-21st 2019.

OTHER EXPERIENCES

Diversity

- *Co-chair and Founder:* NYT Young Professionals.
- *Contributor:* Unpublished Black History https://nyti.ms/2KvaGDM.

Professional Music Experiences

- *Double Bassist* for critically-acclaimed off-Broadway play, *The Dybbuk*. Review: https://bit.ly/2FDSrNB.
- *Double Bassist*. New York Youth Symphony, Carnegie Hall. *2006-2010*.
- *Pianist* All-State Piano Recital and Orchestra. *2008-2010*

MEDIA MENTIONS

(Selected)

1. *USC ISI*. Avery Anderson. 2022. Rescuing Local Journalism, One AI Tool At A Time. https://bit.ly/3UiWTFy
2. *USC Viterbi Magazine*. Caitlin Dawson. 2021.The Good Fight: New York Times Data Scientist Turned USC Computer Science Student. https://bit.ly/3Yn6MmA
3. *Bloomberg*. 2020. Introducing the Third Class of Bloomberg Data Science Ph.D. Fellows (2020-2021). https://bit.ly/3AoQL7E
4. *BuzzFeed*. Peter Aldous. How Russia's Trolls Engaged American Voters. https://bit.ly/2TFrhsB
5. *WIRED*. Louise Matsakis. What Does a Fair Algorithm Actually Look Like? https://bit.ly/2C8uyKN.
6. *The Wall Street Journal*. Benjamin Mullin. New York Times Adapts Data Science Tools for Advertisers. *https://on.wsj.com/2sA4yof*.
7. *National Public Radio*. Le Show, February 18th, 2018. https://bit.ly/2TGEpxF. (A discussion on Project Feels.)
8. *Business Insider Japan*. Fumiaki Ishiguro. AI in Advertising at the *New York Times* (Translated). https://bit.ly/2Q57g0D.
9. *Language Log, University of Pennsylvania*. Mark Liberman. Recommended for You. https://bit.ly/2U7qktu.
10. *KnightLab*. Shakeeb Asrar. A quick look at recommendation engines and how the New York Times makes recommendations. https://bit.ly/2Sa6T15. https://bit.ly/2AIhkBQ.
11. *Women Who Code*. Ema Kaminskaya. ODSC Event Reflections:

    > "Alex's ability to captivate and connect with the audience was a sight to behold. The whole talk felt like an informal conversation between the presenter and 150+ people in the audience. That's definitely a skill and a bit of a talent to manage such a big crowd in a very conversational way, encouraging questions and sparking curiosity."
    >
    > –March 2018. Published on *Women Who Code*. https://bit.ly/2AIhkBQ

REFERENCES

Emilio Ferrara
Email: emiliofe@usc.edu
Relationship: Primary Advisor
Position: Professor of Computer Science and Communication, Associate Department Chair
Phone Number: (310) 448-8661
Institution: University of Southern California

Jonathan May
Email: jonmay@isi.edu
Relationship: Co-Advisor
Position: Research Associate Professor of Computer Science
Phone Number: (310) 448-9157
Institution: University of Southern California, Information Sciences Institute

Nanyun Peng
Email: violetpeng@cs.ucla.edu
Relationship: Co-Advisor
Position: Associate Professor
Phone Number: (310) 206-7284
Institution: University of California Los Angeles

Mark Dredze
Email: mdredze@cs.jhu.edu
Relationship: Bloomberg Mentor
Position: John C Malone Professor
Phone Number: (410) 516-6786
Institution: Johns Hopkins University

James Hamilton
Email: jayth@stanford.edu
Relationship: Stanford Visiting Student Sponsor and Mentor
Position: Hearst Professor of Communication; Chair, Department of Communication
Phone Number: (650) 723-5448
Institution: Stanford University