

Melanie Subbiah

New York, NY

✉ melanie.s.subbiah@gmail.com

🌐 <https://melaniesubbiah.github.io>

Education

- 2020 – present 📖 **Ph.D. Computer Science, Columbia University** (*in progress*)
Topic: Natural Language Processing
Advisor: Kathleen McKeown
- 2020 – 2022 📖 **M.S. Computer Science, Columbia University**
Completed as part of MS/PhD.
- 2013 – 2017 📖 **B.A. Computer Science, Williams College**, Magna Cum Laude, Phi Beta Kappa
Thesis: *Using Text Abstraction and LSTM Language Models for Domain-Independent Narrative Generation*
Advisor: Andrea Danyluk

Work Experience

- (Feb. – Sep.) 2024 📖 **Early Stage Startups** New York, NY
Advisor/Consultant
Scoping and testing technical requirements for customers of a cost effective green data center solution (BuildAI). Advising on models for educational children's story generation (Nookly).
- (Jun.–Aug.) 2022 📖 **Meta** New York, NY
Machine Learning Research Intern - AI for Augmented Reality Input & Interaction
Deep learning to interpret EMG data for human-computer neural interfaces.
- 2019 – 2020 📖 **OpenAI** San Francisco, CA
Member of Technical Staff - Language
Evaluation suite for GPT-3, co-first-author on the GPT-3 paper (Language Models are Few-Shot Learners).
- 2017 – 2019 📖 **Apple** Cupertino, CA
Machine Learning Engineer - AI Research
Data center HVAC efficiency, reward function design, domain randomization for sim-to-real transfer in computer vision, and effective QA methods for crowd-sourced annotated data.
- (Jun.–Aug.) 2016 📖 **Facebook** Menlo Park, CA
Software Engineer Intern - Site Efficiency
Internal C++ tool to monitor/visualize the efficiency of multi-threaded functions.
- (Jun.–Aug.) 2015 📖 **Fathom Information Design** Boston, MA
Data Science Intern - Activity Characterization
Clustering and visualizing activity patterns in daily movement data from wearables.
- (Jun.–Aug.) 2014 📖 **Dartmouth College** Hanover, NH
Research Assistant - Epidemiology
Statistical analysis for study on in utero arsenic exposure in Margaret Karagas's lab.

Honors and Awards

- 2024 • **Most Influential Scholar Award Honorable Mention, AMiner AI 2000**
Recognized as a top 100 most impactful machine learning scholar between 2014 and 2023.

Honors and Awards (continued)

- 2023
 - **Top 100 AI Achievements from 1943-2021**, *BenchCouncil AI100*
Recognized for my work on the GPT series of models.
 - **Fellowship**, *Amazon CAIT PhD Fellowship*
Funding for two years of the PhD awarded by the Columbia-Amazon AI Center.
- 2021
 - **NSF GRFP**, *Honorable Mention*
Awarded an honorable mention for my submission.
- 2020
 - **Best Paper**, *NeurIPS*
Awarded for "Language Models are Few-Shot Learners".
 - **Fellowship**, *Presidential and SEAS Fellowship*,
Columbia University funding award for incoming PhD students.
- 2017
 - **Commencement Speech**, *Phi Beta Kappa Student Speaker*
Voted to speak at Williams College graduation by the Phi Beta Kappa students.
 - **Computer Science Thesis**, *Highest Honors & Best Thesis Presentation award*
Williams College computer science department awards for the graduating class.
 - **Short Story Writing**, *Honorable Mention*
Williams College writing competition.

Publications

Conference Proceedings/Findings and Journals

- Subbiah, M., Zhang, S., Chilton, L. B., & McKeown, K. (2024). Reading Subtext: Evaluating Large Language Models on Short Story Summarization with Writers, Transactions of the Association for Computational Linguistics (TACL, to be presented at EMNLP).
- Storek, A., Subbiah, M., & McKeown, K. (2023). Unsupervised Selective Rationalization with Noise Injection. In *Proceedings*, Association for Computational Linguistics (ACL).
- Wang, G., Chillrud, L., Harwood, K., Ananthram, A., Subbiah, M., & McKeown, K. (2023). Check-COVID: Fact-Checking COVID-19 News Claims with Scientific Evidence. In *Findings*, Association for Computational Linguistics (ACL).
- Levy, S., Allaway, E., Subbiah, M., Chilton, L., Patton, D., McKeown, K., & Wang, W. (2022). SafeText: A Benchmark for Exploring Physical Safety in Language Models. In *Proceedings*, Empirical Methods in Natural Language Processing (EMNLP).
- Mei, A., Kabir, A., Levy, S., Subbiah, M., Allaway, E., Judge, J., ... Wang, W. (2022). Mitigating Covertly Unsafe Text within Natural Language Systems. In *Findings*, Empirical Methods in Natural Language Processing (EMNLP).
- Brown*, T., Mann*, B., Ryder*, N., Subbiah*, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language Models are Few-Shot Learners. In *Proceedings*, Neural Information Processing Systems (NeurIPS, **Best Paper Award**).
- Nygaard, U., Li, Z., Palys, T., Jackson, B., Subbiah, M., Malipatlolla, M., ... Nadeau, K. (2017). Cord blood T cell subpopulations and associations with maternal cadmium and arsenic exposures, PLoS One.

Refereed Workshops

- Subbiah*, M., Bhattacharjee*, A., Hua, Y., Kumarage, T., Liu, H., & McKeown, K. (2023). *Detecting Harmful Agendas in News Articles*. WASSA Workshop at ACL.
- Subbiah, M., & McKeown, K. (2021). *Understanding Identity Signalling in Persuasive Online Text*. International Workshop on Social Sensing at ICWSM.

Maher, M., [Subbiah, M.](#), & Apostoloff, N. (2018). *Cascaded Dataset QA*. Women in Machine Learning at NeurIPS.

[Subbiah, M.](#), lesser, J., & Apostoloff, N. (2018). *Augmenting Training Data with Simulated Images*. Women in Machine Learning at NeurIPS.

Other Articles and Preprints

[Subbiah, M.](#), Ladhak, F., Mishra, A., Adams, G., Chilton, L. B., & McKeown, K. (2024). *STORYSUMM: Evaluating Faithfulness in Story Summarization*. In submission.

*Co-first authors

Teaching

- | | |
|-------------------------|--|
| Spring 2024 | • Guest Lecture , <i>Global Teaching Labs - Uruguay</i>
Taught by Yi-Tong Tseo, MIT & University of Montevideo |
| Fall 2023 | • Teaching Assistant & Guest Lecture , <i>Natural Language Generation and Summarization</i>
Taught by Kathleen McKeown, Columbia University
TA Quality Overall: 4.7/5.0 |
| Summer 2023 | • Instructor , <i>Discrete Mathematics</i>
Columbia University
Course Quality Overall: 4.2/5.0, Instructor Quality Overall: 4.5/5.0 |
| Spring 2023 | • Reviewer , Quick Start Guide to Large Language Models
Written by Sinan Ozdemir, Pearson Publishing |
| | • Guest Lecture , <i>Computational Journalism</i>
Taught by Mark Hansen, Columbia University |
| Fall 2022 | • Teaching Assistant and Guest Lecture , <i>Natural Language Generation and Summarization</i>
Taught by Kathleen McKeown, Columbia University
TA Quality Overall: 4.9/5.0 |
| Fall, Spring 2021 | • Tutor , <i>Introductory Computer Science & Discrete Mathematics</i>
Columbia University Athletics |
| Fall 2014 – Spring 2017 | • Teaching Assistant , <i>Introductory Computer Science and/or Data Structures</i>
Williams College |

Invited Speaking

Research Talks and Posters

- | | |
|------|---|
| 2024 | • Talk on "How did we get here?: The rise of large language models and the problem of evaluation"
Computer Science Colloquium at Williams College |
| 2023 | • Talk on "Detecting Harmful Agendas in News Articles"
WASSA Workshop at ACL |
| 2021 | • Talk on "Understanding Identity Signalling in Persuasive Online Text"
International Workshop on Social Sensing at ICWSM |

Invited Speaking (continued)

- 2020
 - **Talks on "Language Models are Few-Shot Learners"**
New York University
Stanford University
Columbia University
Philosophy and Machine Learning Conference @ NYU
G-Research
 - **Computer Science colloquium on Language Models**
Williams College
- 2018
 - **Poster on "Augmenting Training Data with Simulated Images"**
Women in Machine Learning Workshop at NeurIPS

Outreach Speaking

- 2024
 - **Research Overview: Language Models**
Stuyvesant High School's research class
- 2023
 - **Half-day interactive seminar on ChatGPT for educators**
Academy for Teachers' Master Class for 25 high school teachers
 - **Interview on "Artificial: Episode 2, Selling Out"**
The Wall Street Journal's "The Journal" podcast (a top 10 daily news podcast in the US)
 - **Talk on "NLP with ChatGPT"**
Pearson Publishing's AI Catalyst Conference
 - **Talk on "ChatGPT: What is it, how does it work, and what's next?"**
Columbia University Council of Deans
Columbia University Faculty
- 2022
 - **Guest for "GPT-3 for Natural Language Processing."**
SuperDataScience podcast (50,000 listens)
- 2021
 - **Interview on "5 Levels of Difficulty: Machine Learning"**
Wired Magazine YouTube video (2 million views)
- 2017
 - **Phi Beta Kappa Student Speaker**
Williams College Commencement (~2,000 attendees)

Service

- 2024
 - **Reviewer**, *ACL/EMNLP/ARR*
- 2023
 - **Panelist**, *Columbia Engineering School Graduate Council info panel on PhD programs*
- 2022
 - **Reviewer**, *EMNLP*
 - **Reviewer**, *In2Writing Workshop @ ACL*
 - **Reviewer**, *Columbia Pre-submission Application Review program for underrepresented candidates*
- 2021
 - **Co-organizer**, *Workshop on Enormous Language Models @ ICLR*
- 2020
 - **Reviewer**, *Columbia Pre-submission Application Review program for underrepresented candidates*
 - **Talk**, *Columbia's "Demystifying the Dissertation" talk series for undergrads*
- 2019
 - **Reviewer**, *Bay Area Machine Learning symposium*

Mentoring

- 2021 – present
 - **Research Mentor**, *Columbia University research students*
 - Akankshya Mishra (masters → preprint)
 - Sean Zhang (CS bridge → TACL paper)
 - Maksym Bondarenko (undergrad)
 - Kate Harwood (masters) → ACL findings paper, NLP freelancing
 - Adam Storek (undergrad) → ACL paper, Columbia CS PhD
 - Bobby Yilun Hua (undergrad) → WASSA workshop paper, Cornell CS PhD
 - Yu-Chen Huang (undergrad) → Amazon SDE
- 2022
 - **Mentor**, *Williams College CS undergraduate buddy program*
- 2021
 - **Research Mentor**, *Lumiere Education's high school student research program*
- 2020
 - **Program Organizer & Research Mentor**, *OpenAI Scholars deep learning transition program*
- 2019
 - **Mentor**, *Institute of International Education TechWomen program*
- 2016 – 2017
 - **Leader**, *Williams College Underrepresented Identities in CS group*