



DR. MARGARET MITCHELL

Computer Scientist & Researcher

📍 Seattle, WA, USA @ margarmitchell@gmail.com 🌐 m-mitchell.com

🔗 scholar.google.com/citations?user=5na92fcAAAAJ 📄 margaret-mitchell-9b13429

👤 Current Affiliations: Hugging Face, Harvard BKC, Ada Lovelace, IASEAI



About

I'm an interdisciplinary computer scientist and researcher, with work additionally incorporating programming, math, ethics, social science, cognitive science, psychology, linguistics, and policy.

My ideas have been implemented by multiple technology companies, academic institutions, and government agencies, and I've widely published in top journals and conferences on the topics of ML training, ML data, visual description, assistive and clinical technology, and operationalizing AI ethics.

I'm most known for pioneering “[Model Cards](#)”, now an industry standard; inventing methods to address bias in AI; co-developing “[Seeing AI](#)” to assist low-vision individuals; and co-authoring the “[Stochastic Parrots](#)” paper.

I'm also a regular subject matter expert for news coverage on the interaction of AI and society, and a frequent invited speaker across academia and industry.

Throughout my work, I tend to take a *people-first* approach to AI development: Centering what people need before diving in to how technology can be applied to people.

Research Areas

Artificial Intelligence, Machine Learning, Ethical/Responsible AI, Assistive/Clinical Technology, Bias/Fairness, Natural Language Processing (Generation, Understanding), Multimodal models, Referring Expression Generation/Object Reference, Information Extraction

Select Awards & Achievements

- Co-creator of [Visual Question Answering \(VQA\)](#), awarded the [Mark Everingham Prize](#) for stimulating a new strand of vision and language research. 2025
- Recognized for public-facing AI influence and leadership in [TIME: Most Influential People](#), [TIME100/AI](#), [WIRED](#), [Fortune](#), [Business Insider](#), and other media. 2023
- Received Association for Computational Linguistics “[Test of Time](#)” Award for work I led on [generating descriptions of visual scenes](#). 2022
- Led creation of [Model Cards](#), now a tech industry staple. Received former U.S. Secretary of Defense Ash Carter’s Tech Spotlight award. 2020
- Included in [100 Brilliant Women in AI Ethics](#). 2019
- Research Lead for Microsoft’s [Seeing AI](#) product, which won the [Helen Keller Achievement Award](#) from the American Foundation for the Blind. 2018
- Founder of Google Ethical AI team. 2017
- Co-founder of [Widening Natural Language Processing](#). 2016
- First woman (perhaps?) to win a deep learning competition, [image captioning at CVPR](#). 2015
- Co-founder of [The Workshop on Computational Linguistics and Clinical Psychology](#). 2014

Employment

CHIEF ETHICS SCIENTIST and RESEARCHER, HUGGING FACE <i>Manager: Clément Delangue</i>	2021- now
Implement ethically informed tools for AI releases, such as watermarking and consent. Define and oversee organizational processes aligned with human values. Create new methods for analyzing machine learning data and models in terms of bias, fairness, inclusion, representation, and safety. Provide input on AI policy, including in the U.S. and E.U.	
STAFF RESEARCH SCIENTIST, GOOGLE BRAIN, ETHICAL AI <i>Manager: Samy Bengio</i>	2018- 2021
Grew team to 12+ people with co-lead Dr. Timnit Gebru. Worked on defining and operationalizing AI practices aligned to human values under the broad umbrella of “ethics” across Google. Work included developing methods for accountability, transparency, design processes, dataset construction and usage, and building tools to measure bias issues. Worked closely with Google Cloud on auditing practices and ethical launch protocols. Published theoretical and experimental aspects of our work externally.	
SENIOR RESEARCH SCIENTIST, GOOGLE CEREBRA <i>Manager: Blaise Agüera y Arcas</i>	2016- 2017
Defined and implemented responsible practices for bias evaluation. Co-developed “ML Fairness” organization and founded an “Ethical AI” team.	
RESEARCHER, MICROSOFT RESEARCH, COGNITION GROUP <i>Manager: Pushmeet Kohli</i>	2014- 2016
Focused on developing AI for maximum human benefit, advancing state of the art in image description, storytelling, and descriptions for the visually impaired. Won 1st and 3rd place in the first deep learning image captioning competition at CVPR. Developed Seeing AI, leveraging my work in image captioning and Microsoft’s cloud platform, which won the Helen Keller Achievement Award from the American Foundation for the Blind.	
RESEARCHER, MICROSOFT RESEARCH, NLP GROUP <i>Manager: Bill Dolan</i>	2013- 2014
Created “microsummarization” for Bing & Cortana. Advanced methods for conversation generation and image description, resulting in several patents. Led clinical NLP research.	
POSTDOCTORAL RESEARCHER, JOHNS HOPKINS UNIVERSITY <i>Supervisor: Benjamin Van Durme</i>	2012- 2013
Implemented graphical models for semantic role labeling, named entity recognition, and sentiment detection. With Dr. Matt Gormley, developed Pacaya graphical modeling toolkit.	
JOHNS HOPKINS WORKSHOP GRADUATE, CENTER FOR LANGUAGE AND SPEECH PROCESSING <i>Supervisors: Tamara Berg, Alex Berg</i>	2011
Collaborated on identifying and characterizing visually descriptive text. Led sub-project on a data-driven language generation system, “Midge”, to describe images, which has since been recognized with the Association for Computational Linguistics Test of Time award.	

Education

VISITING SCHOLAR, CENTER FOR SPOKEN LANGUAGE UNDERSTANDING <i>Supervisors: Brian Roark, Richard Sproat</i> Developed system to generate personal language for teenagers with cerebral palsy. Completed work on system that helps predict Mild Cognitive Impairment.	2009-2012
RESEARCH ASSISTANT/ASSOCIATE, CENTER FOR SPOKEN LANGUAGE UNDERSTANDING <i>Supervisors: Brian Roark, John-Paul Hosom, Jan van Santen</i> Led phonetic and syntactic transcription work for projects on Alzheimer's, Aphasia, Apraxia, Autism, Dysarthria, Parkinson's. Developed system to help automatically diagnose Mild Cognitive Impairment.	2005-2007
UNIVERSITY OF ABERDEEN, PHD: COMPUTING SCIENCE Thesis: Generating Reference to Visible Objects <i>Advisors: Kees van Deemter, Ehud Reiter</i> Examined factors affecting reference in visual settings, including the production of color and size words, visual perception, and stored object representations. Created systems that generate human-like reference to visible real world objects.	2009-2012
UNIVERSITY OF WASHINGTON, MA/MS: COMPUTATIONAL LINGUISTICS Thesis: Towards the Generation of Natural Reference <i>Advisors: Scott Farrar, Emily M. Bender</i> Introduced algorithm to generate referring expressions comparable to human-produced expressions. Developed class-based system for the prenominal ordering of modifiers.	2007-2008
REED COLLEGE, BA: LINGUISTICS, ALLIED FIELD: PSYCHOLOGY Senior Thesis: On the Generation of Referring Expressions <i>Advisors: John Haviland, Matt Pearson</i> Critiqued the problem of generating human-like reference from a knowledge base, examining linguistic, psychological, and computational factors.	2001-2005

Select Public Tools

WATERMARKING CODE Code for visible disclosure of AI generation in videos, images, and chat text. Deployed within the Hugging Face Gradio framework, used by developed to share apps.	2025
GATE FOR VOICE CLONING CONSENT Code for adding a "gate" to voice cloning that requires consent.	2025
DIFFUSION BIAS EXPLORER Interface and code for analyzing visual social biases in generative image models.	2023
DATA MEASUREMENT TOOL Interface and code for quantifying different aspects of datasets, to assist in dataset comparison and curation.	2022
KNOW YOUR DATA App for exploring dataset quality and identify biases. Conceived & led the math behind the tool's "Relations" module, which shows how closely associated different characteristics are.	2020

Favorite Publications

MODEL CARD TOOLKIT	2019-2020
Tensorflow library that streamlines and automates generation of Model Cards.	
DATA CARDS PLAYBOOK	2019-2021
Guide to help interdisciplinary teams adopt a people-centered approach to transparency in dataset documentation for responsible AI systems.	
GOOGLE CLOUD MODEL CARDS	2017-2020
Transparent documentation on how well machine learned models work. Focused in particular on the performance of face detection models, with evaluation results disaggregated by face size, facial orientation, perceived gender presentation, age, and skin tone.	
MACHINE LEARNING "CRASH COURSE" ON FAIRNESS.	2019
Teaching module looking at different types of human biases that can manifest in training data, and providing strategies to identify them and evaluate their effects.	
<i>For full list of publications, please see my Google Scholar profile.</i>	
Blili-Hamelin, B. and Graziul, C. and Hancox-Li, L. and Hazan, H. and El-Mhamdi, E.-L. and Ghosh, A. and Heller, K. and Metcalf, J. and Murai, F. and Salvaggio, E. and Smart, A. and Snider, T. and Tighanimine, M. and Ringer, T. and Mitchell, M. and Dori-Hacohen, S. (2025). Stop treating 'AGI' as the north-star goal of AI research. <i>ICML</i> .	2025
Mitchell, M. and Ghosh, A. and Luccioni, A. S. and Pistilli, G. (2025). Fully autonomous agents should not be developed. <i>Arxiv</i> .	2025
Mitchell, M. and Luccioni, A. S. and Lambert, N. and Gerchick, M. and McMillan-Major, A. and Ozoani, E. and Rajani, N. and Thrush, T. and Jernite, Y. and Kiela, D. (2023). Measuring Data. <i>Arxiv</i> .	2023
Aka, O. and Burke, K. and Bäuerle, A. and Greer, C. and Mitchell, M. (2021). Measuring Model Biases in the Absence of Ground Truth. <i>AIES</i> .	2021
Gebru, T. and Bender, E. and McMillan-Major, A. and Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? <i>FAccT</i> .	2021
NB: Not included in h-index. "Shmitchell, S." references multiple authors.	
Mitchell, M. and Baker, D. and Moorosi, N. and Denton, E. and Hutchinson, B. and Hanna, A. and Gebru, T. and Morgenstern, J. (2020). Diversity and inclusion metrics in subset selection. <i>AIES</i> .	2020
Mitchell, M. and Wu, S. and Zaldivar, A. and Barnes, P. and Vasserman, L. and Hutchinson, B. and Spitzer, E. and Raji, I. D. and Gebru, T. (2019). Model Cards for Model Reporting. <i>FAT*</i> .	2019
Hutchinson, B. and Mitchell, M. (2019). 50 years of test (un)fairness: Lessons for machine learning. <i>FAT*</i> .	2019
Benton, A., and Mitchell, M., and Hovy, D. (2017). Multi-task learning for mental health using social media text. <i>EACL 2017</i> .	2017
Agüera y Arcas, B. and Mitchell, M. and Todorov, A. (2017). Physiognomy's New Clothes. <i>Medium</i> .	2017

Select Project Leadership

Huang, T.-H., Ferraro, F., Mostafazadeh, N., Misra, I., Agrawal, A., Devlin, J., Girshick, R., He, X., Kohli, P., Batra, D., Zitnick, C. L., Parikh, D., Vanderwende, L., Galley, M., Mitchell, M. (2016). Visual Storytelling . <i>NAACL</i> .	2016
Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., Parikh, D. (2015). VQA: Visual Question Answering . <i>ICCV</i> .	2015
Devlin, J., Cheng, H., Fang, H., Gupta, S., Deng, L., He, X., Zweig, G., Mitchell, M. (2015). Language Models for Image Captioning: The Quirks and What Works . <i>ACL</i> .	2015
Mitchell, M. and Hollingshead, K. and Coppersmith, G. (2015). Quantifying the Language of Schizophrenia in Social Media . <i>CLPsych Workshop, NAACL</i> .	2015
Fang, H., Gupta, S., Iandola, F., Srivastava, R. K., Deng, L., Dollar, P., Gao, J., He, X., Mitchell, M., Platt, J. C., Zitnick, L., Zweig, G. (2015). From Captions to Visual Concepts and Back . <i>CVPR</i> .	2015
Beller, C., Knowles, R., Harman, C., Bergsma, S., Mitchell, M., Van Durme, B. (2014). I'm a Believer: Social Roles via Self-identification and Conceptual Attributes . <i>ACL</i> .	2014
Mitchell, M., Reiter, E., and van Deemter, K. (2013). Typicality and Object Reference . <i>CogSci</i> .	2013
Mitchell, M., Dodge, J., Goyal, A., Yamaguchi, K., Stratos, K., Han, X., Mensch, A., Berg, A., Berg, T. L., Daumé III, H. (2012). Midge: Generating Image Descriptions From Computer Vision Detections . <i>EACL</i> .	2012
Mitchell, M., Dunlop, A., and Roark, B. (2011). Semi-Supervised Modeling for Prenominal Modifier Ordering . <i>ACL</i> .	2011
Mitchell, M., van Deemter, K., and Reiter, E. (2011). On the Use of Size Modifiers When Referring to Visible Objects . <i>CogSci</i> .	2011
Roark, B., Mitchell, M., Hosom, J., Hollingshead, K., and Kaye, J. (2011). Spoken Language Derived Measures for Detecting Mild Cognitive Impairment . <i>IEEE Transactions on Audio, Speech, and Language Processing</i> .	2011
Mitchell, M., van Deemter, K., and Reiter, E. (2010). Natural Reference to Objects in a Visual Domain . <i>INLG</i> .	2010
PROJECT LEAD, Ethics of AI Agents Analyzing ethical considerations in the development, use, procurement, and deployment of AI agents; developing recommendations for future research and governance. Research has been showcased in multiple podcasts and news articles, and featured as an MIT Tech Review OpEd . Relevant preprint available on Arxiv .	2024-current
PROJECT LEAD, SHADES dataset With researchers all over the world, created multilingual, multi-cultural dataset for examining multiple bias types. Main project paper published at NAACL 2025 ; Available on Hugging Face .	2023-2024
Co-CHAIR, BigScience Data Governance Working Group With co-chair Yacine Jernite , led international group of engineers and researchers in defining a fully open structure for the governance of data used in Large Language Models. Main project paper published at FAccT 2022 .	2021-2022

FOUNDER & LEAD, Model Cards	2017-2021
Worked cross-Alphabet to operationalize “ Model Cards ”, transparent documentation that accompanies machine learning models. Launches include multiple Google orgs, followed by launches from other companies: OpenAI (I assisted with the model card for GPT2), Nvidia, Allen AI, Salesforce, Hugging Face, and more.	
GOOGLE REPRESENTATIVE, Partnership on AI	2018-2020
Worked within the Fairness, Transparency, and Accountability group to establish norms on these issues across AI-focused organizations.	
CO-FOUNDER & RESEARCH LEAD, ABOUT ML in the Partnership on AI	2019-2021
Building from my work on Model Cards, created collaboration with Microsoft and IBM on transparency documentation for machine learning models.	
TECHNICAL & RESEARCH LEAD, Seeing AI	2014-2016
Leveraged my image captioning work to produce visual descriptions for the visually impaired. Streamed technology through cloud services to land on user’s mobile device or smart-glasses. Connected groups throughout Microsoft to move forward as one: MSR Labs, Cognitive Services, Windows, Narrator, Bing, Garage, Outreach, NExT. App available here .	
<i>Won the prestigious Helen Keller Achievement Award from the American Foundation for the Blind in 2018.</i>	
MICROSOFT STRATEGIC INTERN PROJECT LEAD, Visual Storytelling	2015-2016
Led group of 8 researchers, 2 visiting professors, and 5 students on summer project to drive forward research on selecting key frames in a photo album and generating a coherent story for them. Project resulted in full working system, with all 4 full-time interns first-authoring a paper at a top-tier conference. Main project paper published at NAACL 2016 .	
<i>Has been the basis of multiple storytelling workshops and competitions.</i>	
SENIOR RESEARCHER, JSALT Mental Health Summer Workshop	2016-2017
Exploring how we can utilize shared patient data alongside shared patient social media feeds to monitor PTSD and depression. Relevant paper published at EACL 2017 .	
<i>Some of the first work on fairness in health settings.</i>	
TOP DISHES RESEARCHER, Cortana and Bing Local Recommendations.	2014-2015
Developed concept of “microsummarization” and led team implementing an end-to-end system that mined social media for positive/negative sentiment expressed towards common items. Relevant paper published at AAAI 2016 .	
<i>Incorporated into Microsoft products.</i>	

Academic Service (Subset)

MENTOR, Women in Machine Learning (WiML) (multiple years)

AREA CHAIR, FAccT (multiple years), PROGRAM CHAIR 2020-2021 and FAccT/FAT* STEERING COMMITTEE 2018-2021

CO-FOUNDER, Widening NLP (WiNLP) group and annual workshop

CO-FOUNDER, Ethics in NLP workshop in the Association for Computational Linguistics (ACL)

CO-FOUNDER, PROGRAM & PUBLICATION CHAIR, Annual Storytelling workshop at ACL

CO-FOUNDER, Computational Linguistics and Clinical Psychology workshop (CLPsych) at ACL

PROGRAM & PUBLICATION CHAIR for ACL 2014 and North American ACL (NAACL) 2015

AREA CHAIR, Generation, ACL 2016

PUBLICATION CHAIR & AREA CHAIR, Generation, NAACL 2016

Publication chair contributions still used in official style guidelines.

AREA CHAIR, GENERATION & SUMMARIZATION, Empirical Methods in Natural Language Processing (EMNLP) 2015

GENERAL, PROGRAM & PUBLICATION CHAIR, the International Natural Language Generation Conference (INLG) 2014

TRACK OWNER, KNOWLEDGE BASE POPULATION - SENTIMENT SLOT FILLING, National Institute of Standards and Technology (NIST) Text Analysis Conference (TAC) 2014

REGULAR REVIEWER FOR:

- Various workshops on fairness, vision-to-language, clinical NLP work
- Neural Information Processing Systems (NeurIPS)
- Empirical Methods in Natural Language Processing (EMNLP)
- The Annual Meeting of the Association for Computational Linguistics (ACL)
- The North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)
- ACM Conference on Fairness, Accountability, and Transparency (FAccT, formerly FAT*)
- The International Natural Language Generation Conference (INLG)

Systems & Languages

Python, Bash, C#, Java. Experience with Tcl, C++, HTML, JavaScript, PHP, sed, awk, etc.

Comfortable working with Windows, MacOS 9 and X, Linux (Ubuntu, Fedora), all MS Office software, most Google Enterprise software, all OpenOffice software, $\text{\LaTeX}$, Praat, Wavesurfer, version control systems, Tensorflow

Personal Details

Citizenship: U.S.

Residence: Seattle, Washington

Date of birth: 18th November, 1983

Hobbies: Gardening, exercising, birds

Nerdy obsession with collecting vinyl records: Yes