DR. MARGARET MITCHELL

Computer Scientist & Researcher

margarmitchell@gmail.com | Seattle, WA, U.S.A. | m-mitchell.com
https://scholar.google.com/citations?user=5na92fcAAAAJ | margaret-mitchell-9b13429/

About Me

Focused on foundational machine learning research and AI ethics operationalization, bringing together machine learning, ethics, social science, cognitive science, language, law and policy. My work is interdisciplinary, and tends to include math, computer vision, and natural language processing. I also like to work on applied assistive and augmentative technology. I wear many hats when I work, adapting my role to the needs of each project.

Research Areas

Artificial/Machine Intelligence, Ethical/Responsible AI, Machine Learning (ML), Personal identity in ML, Assistive/Clinical Technology, Natural Language Generation, Natural Language Processing, Multimodal models (Computer Vision and Natural Language), Referring Expression Generation/Object Reference, Information Extraction

Some Notable Things

- Recently recognized in TIME: 100 Most Influential People of 2023, TIME100/AI, WIRED, and FORTUNE: 13 Top AI Innovators 2023 — 2023
- Received Association for Computational Linguistics "Test of Time" Award for work I led on generating descriptions of visual scenes — 2022
- Research Lead for Google Cloud Model Cards product, which won 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 prestigious Helen Keller Achievement Award from the American Foundation for the Blind — 2018
- Co-founder of Widening Natural Language Processing — 2016
- First woman (perhaps?) to win a deep learning competition, image captioning at CVPR — 2015

Work

Chief Ethics Scientist and Researcher, Hugging Face — 2021-now
Manager: Clément Delangue

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 — 2018-2021
Manager: Samy Bengio

Grew team to 12+ people with my 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 2016-2017

Manager: Blaise Agüera y Arcas

Defined and implemented responsible practices for bias evaluation. Developed "ML Fairness" organization and founded an "Ethical AI" team.

Researcher, Microsoft Research, Cognition Group 2014-2016

Manager: Pushmeet Kohli

Began to deeply focus on how to advance AI to maximally benefit people. Advanced the state of the art in image description, storytelling, and visual 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 went on to win the Helen Keller Achievement Award from the American Foundation for the Blind

Researcher, Microsoft Research, NLP Group 2013-2014

Manager: Bill Dolan

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 2012-2013

Supervisor: Benjamin Van Durme

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 2011

Supervisors: Tamara Berg, Alex Berg

Engineered data-driven language generation system, "Midge", to read in the output of state-of-the-art computer vision systems and generate syntactically/semantically well-formed descriptions of images. Collaborated on automatic classification of visually descriptive text, and characterizing description. The work I led has since been recognized with the Test of Time award.

Visiting Scholar, Center for Spoken Language Understanding 2009-2012

Supervisors: Brian Roark, Richard Sproat

Developed system to generate personal language for teenagers with cerebral palsy. Completed work on system that helps predict Mild Cognitive Impairment.

Research Assistant/Associate, Center for Spoken Language Understanding 2005-2007

Supervisors: Brian Roark, John-Paul Hosom, Jan van Santen

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.

Education

University of Aberdeen, PhD: Computing Science — 2009-2012
Thesis: Generating Reference to Visible Objects
Advisors: Kees van Deemter, Ehud Reiter
Examined factors affecting reference in visual settings, including the production of color and size words, visual perception, and stored object representations. Created an algorithm that generates human-like reference to visible real world objects.

University of Washington, MA/MS: Computational Linguistics — 2007-2008
Thesis: Towards the Generation of Natural Reference
Advisors: Scott Farrar, Emily M. Bender
Introduced algorithm to generate referring expressions comparable to human-produced expressions. Developed class-based system for the prenominal ordering of modifiers.

Reed College, BA: Linguistics, Allied Field: Psychology — 2001-2005
Senior Thesis: On the Generation of Referring Expressions
Advisors: John Haviland, Matt Pearson
Critiqued the problem of generating human-like reference from a knowledge base, examining linguistic, psychological, and computational factors.

Recent Public Tools

Diffusion Bias Explorer — 2023
Tool to analyze visual social biases in generative image models.

Data Measurement Tool — 2022
Interactive tool for quantifying different aspects of datasets, to assist in dataset comparison and curation.

Know Your Data — 2020-2021
Interactive tool to explore datasets with the goal of improving data quality and identifying bias issues. Conceived the idea of & led the math behind the tool's "Relations" module, which show how closely associated different characteristics are.

Diversity and Inclusion Explorable — 2020-2021
Building from my work on formal methods for measuring diversity and inclusion, this interactive tool focused on explaining diversity metrics methods and visually demonstrating tradeoffs between different diversity goals.

Model Card Toolkit — 2019-2020
Tensorflow library that streamlines and automates generation of Model Cards.

Data Cards Playbook — 2019-2021
Playbook to help interdisciplinary teams adopt a people-centered approach to transparency in dataset documentation for responsible AI systems.

Fairness Indicators — 2017-2019
A suite of tools to enable simple computation and visualization of fairness metrics for binary and multi-class classification.

Google Cloud Model Cards — 2017-2019
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.

Favorite Publications

For full list of publications, please see my Google Scholar profile.
h-index: 43; i10-index: 66

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. *Arxiv preprint*. 2023

Jernite, Y. and Nguyen, H. and Biderman, S. and Rogers, A. and Masoud, M. and Danchev, V. and Tan, S. and Luccioni, A. S. and Subramani, N. and Johnson, I. and Dupont, G. and Dodge, J. and Lo, K. and Talat, Z. and Radev, D. and Gokaslan, A. and Nikpoor, S. and Henderson, P. and Bommasani, R. and Mitchell, M. (2022). Data Governance in the Age of Large-Scale Data-Driven Language Technology. *FAccT 2022*. 2022

Aka, O. and Burke, K. and Bäuerle, A. and Greer, Christina, and Mitchell, M. (2021). Measuring Model Biases in the Absence of Ground Truth. *AIES 2021*. 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? 🦜 *FAccT 2021*.
`NB: Not included in h-index. "Shmitchell, S." references multiple authors.`

Hutchinson, B. and Smart, A. and Hanna, A. and Denton, E. and Greer, C. and Kjartansson, O. and Barnes, P. and Mitchell, M. (2021). Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure. *FAccT 2021*.

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. *AIES 2020*. 2020

Raji, I. D. and Gebru, T. and Mitchell, M. and Buolamwini, J. and Lee, J. and Denton, E. (2020). Saving face: Investigating the ethical concerns of facial recognition auditing. *FAccT 2020*.

Raji, I. D. and Smart, A. and White, R. N. and Mitchell, M. and Gebru, T. and Hutchinson, B. and Smith-Loud, J. and Theron, D. and Barnes, P. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. *FAccT 2020*.

Prabhakaran, V. and Hutchinson, B. and Mitchell, M. (2019). Perturbation sensitivity analysis to detect unintended model biases. *EMNLP 2019*. 2019

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. *FAT* 2019*.

Hutchinson, B. and Mitchell, M. (2019). 50 years of test (un) fairness: Lessons for machine learning. *FAT* 2019*.

Zhang, B. H. and Lemoine, B. and Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. *AIES 2018*. 2018

Benton, A. and Mitchell, M. and Hovy, D. (2017). Multi-task learning for mental health using social media text. *EACL 2017*. 2017

Agüera y Arcas, B. and Mitchell, M. and Todorov, A. (2017). Physiognomy's New Clothes. *Medium*.

Ryu, H. J. and Mitchell, M. and Adam, H. (2017). Improving smiling detection with race and gender diversity. *FAT/ML 2017*.

Huang, T.-H., and Ferraro, F., and Mostafazadeh, N. and Misra, I. and Agrawal, A. and Devlin, J. and Girshick, R. and He, X. and Kohli, P. and Batra, D. and Zitnick, C. L. and Parikh, Devi and Vanderwende, L. and Galley, M. and Mitchell, M. (2016). Visual Storytelling. *NAACL 2016*. 2016

Misra, I. and Zitnick, C. L. and Mitchell, M. and Girshick, R. (2016). Seeing through the Human Reporting Bias: Visual Classifiers from Noisy Human-Centric Labels. *CVPR 2016*.

Mason, R. and Gaska, B. and Van Durme, B. and Choudhury, P. and Hart, T. and Dolan, B. and Toutanova, K. and Mitchell, M. (2016). Microsummarization of Online Reviews: An Experimental Study. *AAAI 2016*.

Antol, S. and Agrawal, A. and Lu, J. and Mitchell, M. and Batra, D. and Zitnick, C. L. and Parikh, D. (2015). VQA: Visual Question Answering. *ICCV 2015*. 2015

Devlin, J. and Cheng, H. and Fang, H. and Gupta, S. and Deng, L. and He, X. and Zweig, G. and Mitchell, M. (2015). Language Models for Image Captioning: The Quirks and What Works. *ACL 2015*.

Mitchell, M. and Hollingshead, K. and Coppersmith G. (2015). Quantifying the Language of Schizophrenia in Social Media. *the 2nd CLPsych Workshop, NAACL 2015*.

Fang, H. and Gupta, S. and Iandola, F. and Srivastava, Rupesh K. and Deng, L. and Dollar, P. and Gao, J. and He, X. and Mitchell, M. and Platt, J. C. and Zitnick, L. and Zweig, G. (2015). From Captions to Visual Concepts and Back. *CVPR 2015*.

Gormley, M., and Mitchell, M., and Van Durme, B., and Dredze, M. (2014). Low Resource Semantic Role Labeling. *ACL 2014*. 2014

Beller, C., and Knowles, R., and Harman, C., and Bergsma, S., and Mitchell, M., and Van Durme, B. (2014). I'm a Belieber: Social Roles via Self-identification and Conceptual Attributes. *ACL 2014*.

Mitchell, M., and Aguilar, J., and Wilson, T., and Van Durme, B. (2013). Open Domain Targeted Sentiment. *EMNLP 2013*. 2013

Mitchell, M., and van Deemter, K., and Reiter, E. (2013). Attributes in Visual Reference. *PRE-CogSci 2013*.

Mitchell, M. and Reiter, E., and van Deemter, K. (2013). Typicality and Object Reference. *CogSci 2013*.

Mitchell, M. and van Deemter, K., and Reiter, E. (2013). Generating Expressions that Refer to Visible Objects. *NAACL 2013*.

Mitchell, M., Dodge, J., Goyal, A., Yamaguchi, K., Stratos, K., Han, X., Mensch, A., Berg, A., and Berg, T. L., Daumé III, H. (2012). Midge: Generating Image Descriptions From Computer Vision Detections. *EACL 2012*. 2012

Mitchell, M., Dunlop, A., and Roark, B. (2011). Semi-Supervised Modeling for Prenominal Modifier Ordering. *ACL 2011*. 2011

Mitchell, M., van Deemter, K., and Reiter, E. (2011). On the Use of Size Modifiers When Referring to Visible Objects. *CogSci 2011*.

Roark, B., Mitchell, M., Hosom, J., Hollingshead, K., and Kaye, J. (2011). Spoken Language Derived Measures for Detecting Mild Cognitive Impairment. *IEEE Transactions on Audio, Speech, and Language Processing.*

Mitchell, M., van Deemter, K., and Reiter, E. (2010). Natural Reference to Objects in a Visual Domain. *INLG 2010.* 2010

Favorite Public Projects

Founder & Lead, Model Cards 2017-2021

Worked cross-Alphabet to operationalize "Model Cards", transparent documentation that accompanies machine learning models. Launches include Google Cloud, Perception, and Jigsaw, as well as launches from OpenAI, Nvidia, Allen AI, Salesforce, Hugging Face.

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 2015-2016

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 smartglasses. Connected groups throughout Microsoft to move forward as one: MSR Labs, Cognitive Services, Windows, Narrator, Bing, Garage, Outreach, NExT.

Won the prestigious Helen Keller Achievement Award from the American Foundation for the Blind in 2018.

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.

Has been the basis of multiple storytelling workshops and competitions.

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.

Some of the first work on fairness in health settings.

Top Dishes Researcher, Cortana and Bing Local Recommendations. 2014-2015

Developed end-to-end system to mine social media for positive/negative sentiment expressed towards most common items.

Incorporated into Microsoft products.

Academic Service (Subset)

FAccT Program Chair 2020-2021 and FAccT/FAT* Steering Committee 2018-2021

Co-Founder, Widening NLP (WiNLP) group and annual workshop

Co-Founder, Annual 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) in ACL

Program & Publication Chair for ACL 2014 and North American ACL (NAACL) 2015
First-of-its-kind gathering of medical practitioners and computational linguists. Became an annual event, still ongoing.

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

Knowledge Base Population - Sentiment Slot Filling Track Owner, National Institute of Standards and Technology (NIST) Text Analysis Conference (TAC) 2014
For the first time, defined and organized task for the NIST TAC to predict sentiment between knowledge base entities.

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)
- The International Natural Language Generation Conference (INLG)

Systems & Languages

Program in 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, LaTeX, Praat, Wavesurfer, version control systems, Tensorflow

Personal Details

Citizenship: U.S.
Residence: Seattle, Washington
Date of birth: 18th November, 1983
Hobbies: Gardening, exercising, political analysis. Occasionally all at once.
Nerdy obsession with collecting vinyl records: Yes