Hussein Mozannar

hmozannar@microsoft.com, https://husseinmozannar.github.io/, 617-821-4607

Current Role	*Senior Researcher,* **Microsoft Research - AI Frontiers**, *Seattle, WA*	2024-

Working with the Human-AI eXperiences (HAX) Team on building AI agents that interact with people.

Education	***Massachusetts Institute of Technology***, *Cambridge, MA*	2019 - 2024

Ph.D, Social and Engineering Systems and Statistics
Adviser: David Sontag,   Committee: Arvind Satyanarayan, Eric Horvitz, and Elena Glassman
Member of MIT-IBM Watson AI Lab, MIT IDSS and CSAIL
M.S, Social and Engineering Systems   2023

American University of Beirut, *Beirut, Lebanon*   2015 - 2019
B.E. in Computer and Communications Engineering, GPA: 4.0/4.0

Work Experience

Research Intern, **Microsoft Research - HAX team**   2022
Worked on understanding how programmers interact with GitHub Copilot. Advised by Gagan Bansal, Adam Fourney and Eric Horvitz.

Machine Learning Research Intern, **ASAPP**   2021
Worked on optimizing AI suggestions for customer service agents. Advised by Ethan Elenberg, Ramya Ramakrishnan, and Kilian Weinberger.

Intern, **Toyota Technological Institute at Chicago**   2018
Studied the long term effects of imposing demographic parity on selection processes. Advised by Nati Srebro and Mesrob Ohannessian

Publications

[12] ***Reading Between the Lines: Modeling User Behavior in AI-Assisted Programming.*** Hussein Mozannar, Gagan Bansal, Adam Fourney, Eric Horvitz. *To appear in Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI 2024)*

[11] **When to Show a Suggestion? Integrating Human Feedback in AI-Assisted Programming.** Hussein Mozannar, Gagan Bansal, Adam Fourney, Eric Horvitz. *To appear in Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2024)*

[10] ***Effective Human-AI Teams via Learned Natural Language Rules and Onboarding.*** Hussein Mozannar, Jimin J Lee, Dennis Wei, Prasanna Sattigeri, Subhro Das, David Sontag . *In Proceedings of The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS 2023),* **Spotlight (top 3.6% of submissions)**.

[9] ***In Defense of Softmax Parametrization for Calibrated and Consistent Learning to Defer.*** Yuzhou Cao, Hussein Mozannar, Lei Feng, Hongxin Wei, Bo An. *In Proceedings of The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS 2023).*

[8] ***Who Should Predict? Exact Algorithms For Learning to Defer to Humans.*** Hussein Mozannar, Hunter Lang, Dennis Wei, Prasanna Sattigeri, Subhro Das, David Sontag . *In Proceedings of The 26th International Conference on Artificial Intelligence and Statistic (AISTATS 2023),* **Oral (top 1.9% of submissions)**

[7] ***Sample Efficient Learning of Predictors that Complement Humans.*** Mohammad-Amin Charusaie*, Hussein Mozannar*, David Sontag and Samira Samadi. *In Proceedings of the Thirty-*

Ninth International Conference on Machine Learning (ICML 2022) . (*: equal contribution)

[6] **Teaching Humans When To Defer to a Classifier.** <u>Hussein Mozannar</u>, David Sontag, and Arvind Satyanarayan. *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2022, 15% acceptance rate) and in and CHI-TRAIT 2022 Workshop.* (**Covered by MIT News, NSF Research News, Tech Explorist**)

[5] **Consistent Estimators for Learning to Defer to an Expert.** <u>Hussein Mozannar</u> and David Sontag. *In Proceedings of the Thirty-Seventh International Conference on Machine Learning (ICML 2020)* (**Covered by MIT News, Forbes, MIT Tech Review**)

[4] **Fair Learning with Private Demographic Data.** <u>Hussein Mozannar</u>, Mesrob I Ohannessian, and Nathan Srebro. *In Proceedings of the Thirty-Seventh International Conference on Machine Learning (ICML 2020) and in NeurIPS 2019 ML with Guarantees Workshop* (**Contributed Talk top 5%**)).

[3] **From fair decision making to social equality.** <u>Hussein Mozannar</u>, Mesrob I Ohannessian, and Nathan Srebro. *In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT 2019).*

[2] **Neural Arabic Question Answering.** <u>Hussein Mozannar</u>, Karl El Hajal, Elie Maamary and Hazem Hajj. *In Proceedings of the Fourth Arabic Natural Language Processing Workshop (ACL 2019)* (**Contributed Talk**)

[1] **Damage Identification in Social Media Posts using Multimodal Deep Learning.** <u>Hussein Mozannar</u>, Yara Rizk, and Mariette Awad. *In 15th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2018)*

Press	

Consistent Estimators for Learning to Defer to an Expert. [5]
- MIT News: AI systems that work w/doctors and know when to step in   2020
- Forbes: Can AI Tell Us When To Use AI And When Not To?   2020
- MIT Tech Review: AI is learning when it should and shouldn't defer to a human   2020

Teaching Humans When To Defer to a Classifier [6]
- MIT News: When should someone trust an AI assistant's predictions?   2022
- NSF Research News: When should someone trust an AI assistant's predictions?   2022
- Tech Explorist: New technique guides humans when to trust an AI   2022

Effective Human-AI Teams via Learned Natural Language Rules and Onboarding [10]
- MIT News: Automated system teaches users when to collaborate with an AI assistant   2023

Workshop Papers

[2] **Simulating Iterative Human-AI Interaction in Programming with LLMs.** <u>Hussein Mozannar</u>, Valerie Chen, Dennis Wei, Prasanna Sattigeri, Manish Nagireddy, Subhro Das, Ameet Talwalkar and David Sontag . *Workshop on Instruction Tuning and Instruction Following NeurIPS23. 2023*

[1] **Machine Teaching with Generative Models for Human Learning.** Michael Doron*, <u>Hussein Mozannar*</u>, David Sontag and Juan Caicedo. *In Third ICML Workshop on Human In the Loop Learning (HILL). 2021*

Working Papers

[4] **The Real HumanEval: Evaluating Large Language Models' Abilities to Support Programmers** <u>Hussein Mozannar*</u>, Valerie Chen*, Mohammed Alsobay, Subhro Das, Sebastian Zhao, Dennis Wei, Manish Nagireddy, Prasanna Sattigeri, Ameet Talwalkar, David Sontag. *In Submission*

[3] ***Impact of Large Language Model Assistance on Patients Reading Clinical Notes: A Mixed-Methods Study.*** *Niklas Mannhardt, Elizabeth Bondi-Kelly, Barbara D. Lam, Chloe P O'Connell, Mercy Nyamewaa Asiedu,* <u>*Hussein Mozannar*</u>*, Monica Agrawal, Tatiana Urman, Catherine E. Ricciard, Alejandro Buendia, David Sontag, Marzyeh Ghassemi . In Submission and arXiv preprint arXiv:2401.09637*

[2] ***Closing the Gap in High-Risk Pregnancy Care Using Machine Learning and Human-AI Collaboration.*** <u>*Hussein Mozannar*</u>**, Yuria Utsumi*, Irene Y Chen, Stephanie S Gervasi, Michele Ewing, Aaron Smith-McLallen, David Sontag. In Submission and arXiv preprint arXiv:2305. 17261*

[1] ***Aligning Language Models at Inference-Time using Per-Token Rewards.*** *Karim El Husseini,* <u>*Hussein Mozannar*</u>*, Mesrob Ohannessian. In Submission*

Awards	KAUST Rising Stars in AI	2024
	NeurIPS Top Reviewer	2023
	ICLR Outstanding Reviewer	2022
	Dean's Award For Best Innovative Software Development Project (AUB)	2019
	Dean's Honor List (AUB)	2015-2019

Service	***Community Engagement***	
	MIT IDSS Student Council: Diversity, Equity, and Inclusion co-chair	2021-
	MIT SES PhD program Graduate Assistance Application Program organizer	2021-
	Co-organizer of LebNet Technology Fellowship program	2023-2024
	Area Chair	
	EMNLP	2024
	Reviewer	
	ACM Conference on Human Factors in Computing Systems (CHI)	2024
	International Conference on Machine Learning (ICML)	2023
	International Conference on Artificial Intelligence and Statistics (AISTATS)	2023
	ACM Web Conference (WWW)	2023
	ACM Conference on Fairness, Accountability, and Transparency (FAccT)	2022
	Neural Information Processing Systems (NeurIPS)	2021, 2022, 2023, 2024
	International Conference on Learning Representations (ICLR)	2022
	Workshops	
	Workshop on Representation Learning for Responsible Human-Centric AI (AAAI 2023 R2HCAI)	
	Workshop on Human-Machine Collaboration and Teaming (ICML 2022 HMCaT)	
	Workshop on Trust and Reliance in AI-Human Teams (CHI TRAIT 2022)	

Teaching	Machine Learning for Healthcare (6.7930/HST.956 MIT), Head TA	2023
	Introduction To Programming (AUB), TA	2016

Talks	JetBrains	2024
	Allen Institute for AI "Training Human-AI Teams"	2024
	Cornell University "Training Human-AI Teams"	2024
	Microsoft Research "Training Human-AI Teams"	2024
	NeurIPS 2023 Tutorial "How to Work With Real Humans in Human-AI Systems" (panel)	2023
	Northwestern University	2023
	University of Chicago Human+AI Lab	2023
	University of Ilinois at Chicago	2023
	IBM Research Zurich Seminar	2023
	Microsoft Research Redmond	2022

	MIT Horizon	2022
	MIT 6.871/HST.956: Machine Learning for Healthcare, Guest Lecture	2022
	MIT AI Ethics group	2022
	University of Washington Lab for Human-AI Interaction	2022
	Boston University AI Research Seminar	2021
	Alan Turing Institute and DeepMind Reading Group	2020
Mentoring	Michael Peng, Undergrad	2024
	Karim El Husseini (UIC intern), Undergrad	2023
	Yuria Utsumi, MEng	2021-2022
	Jimin J Lee, Undergrad	2022-2023