ELINOR POOLE-DAYAN

✉ elinorpd@mit.edu | 🔗 elinor-poole-dayan | 🎓 Google Scholar | 🔗 elinorp-d.github.io

EDUCATION

Massachusetts Institute of Technology (MIT)   Sept. 2023 – May 2025
Master's of Science   *Cambridge, MA, USA*

- Advised by Deb Roy in the MIT Center for Constructive Communication. GPA 5.0/5.0
- Thesis: *From Dialogue to Decision: An LLM-Powered Framework for Analyzing Collective Idea Evolution and Voting Dynamics in Deliberative Assemblies* (Grade: A+)

McGill University   Sept. 2019 – May 2023
Bachelor's in Honours Computer Science and Mathematics   *Montréal, QC, Canada*

- First Class Honours with Distinction. GPA 3.9/4.0

RESEARCH & PROFESSIONAL EXPERIENCE

Predoctoral Researcher   July 2025 – Present
Massachusetts Institute of Technology (MIT)   *Cambridge, MA, USA*

- Research on pluralistic alignment in LLMs [C.1, W.1] and evaluating the fairness of AI's societal impacts; supervised by Michiel Bakker.

Graduate Research Assistant   Sept. 2023 – June 2025
Center for Constructive Communication, MIT Media Lab   *Cambridge, MA, USA*

- Conducted research related to fairness and bias in LLMs [C.2, C.4] and computational social science [W.3, C.3].
- Played a key role in the design and execution of the MIT Student Deliberative Assembly on Sustainability [R.1].
- Supported engineering and design efforts for tech-enhanced deliberative assembly pilots in Ireland and Oregon.
- Led lab efforts to organize and write documentation and developed custom API wrappers for Fora data handling.

Data Science Intern   May 2022 – Aug. 2022
Unity Technologies   *Montréal, QC, Canada*

- Optimized deep learning algorithms to throttle bid requests in Unity's Ad Exchange using Tensorflow.
- Decreased model training time by 25% and reduced model size and number of parameters by 50%.
- Created a text data preprocessing pipeline on Google Cloud Platform Dataflow using Apache beam.

NLP Research Intern   Jan. 2022 – May 2022
McGill University   *Montréal, QC, Canada*

- Identified the geo-indicativeness of text using BERT applied to geosocial datasets to build a safety tool for social media. Supervised by Grant McKenzie.

NLP Research Intern   June 2021 – Aug. 2021
Shamoon College of Engineering   *Be'er Sheva, Israel*

- Classified author gender of books to perform a case study on female authors who wrote under male pseudonyms.
- Preprocessed data using CoreNLP and scikit-learn. Designed and implemented baseline experiments using SVMs.
- Supervised by Irina Rabaev and Marina Litvak.

NLP Research Intern   Jan. 2021 – May 2021
Mila - Quebec AI Institute   *Montréal, QC, Canada*

- Investigated the effect of gender debiasing on fine-tuned language models such as BERT using PyTorch.
- Explored debiasing methods and reformulated bias metrics for racial and religious biases.
- Published paper in ACL with Nick Meade [C.6] and supervised by Siva Reddy.

Academic Conference Publications

C.1 **Benchmarking Overton Pluralism in LLMs**
Elinor Poole-Dayan, Jiayi Wu, Taylor Sorensen, Jiaxin Pei, and Michiel A. Bakker
ICLR 2026.

C.2 **LLM Targeted Underperformance Disproportionately Impacts Vulnerable Users**
Elinor Poole-Dayan, Deb Roy, and Jad Kabbara
AAAI 2026.

C.3 **Computational Analysis of Conversation Dynamics through Participant Responsivity**
Margaret Hughes, Brandon Roy**, Elinor Poole-Dayan**, Deb Roy, and Jad Kabbara
EMNLP 2025 (Oral Presentation).

C.4 **On the Relationship between Truth and Political Bias in Language Models**
Suyash Fulay, William Brannon, Shrestha Mohanty, Cassandra Overney**, Elinor Poole-Dayan**, Deb Roy, and Jad Kabbara
EMNLP 2024.

C.5 **Are Diffusion Models Vision-And-Language Reasoners?**
Benno Krojer**, Elinor Poole-Dayan**, Vikram Voleti, Chris Pal, and Siva Reddy
NeurIPS 2023.

C.6 **An Empirical Survey of the Effectiveness of Debiasing Techniques for Pre-trained Language Models**
Nicholas Meade**, Elinor Poole-Dayan**, and Siva Reddy
ACL 2022.

Workshop Papers & Extended Abstracts

W.1 **Benchmarking Overton Pluralism in LLMs**
Elinor Poole-Dayan, Jiayi Wu, Jiaxin Pei, and Michiel A. Bakker
NeurIPS 2025 LLM Evaluation Workshop (Poster Presentation).

W.2 **Tracing Idea Evolution in Democratic Deliberation with LLMs**
Elinor Poole-Dayan and Deb Roy
COLM 2025 NLP4Democracy Workshop (Oral Presentation).

W.3 **LLMs for Public Narrative Understanding**
Elinor Poole-Dayan, Daniel T. Kessler, Margaret Hughes, Emily S. Lin, Marshall Ganz, and Deb Roy
NAACL 2025 Workshop on Narrative Understanding (Poster Presentation).

W.4 **LLM Targeted Underperformance Disproportionately Impacts Vulnerable Users**
Elinor Poole-Dayan and Deb Roy
NeurIPS 2024 Safe Generative AI Workshop (Poster Presentation).

W.5 **Sandbagging in LLMs: Systematic Underperformance for Uneducated and ESL Users**
Elinor Poole-Dayan and Deb Roy
IC2S2 2024 (Poster Presentation).

Pre-prints & Under Review

R.1 **An AI-Powered Framework for Analyzing Collective Idea Evolution in Deliberative Assemblies**
Elinor Poole-Dayan, Deb Roy, and Jad Kabbara
Aug. 2025.

R.2 **Applying Large Language Models to Characterize Public Narratives**
Elinor Poole-Dayan*, Daniel T Kessler*, Hannah Chiou, Margaret Hughes, Emily S. Lin,
Marshall Ganz, and Deb Roy
May 2025.

INVITED TALKS AND RESEARCH PRESENTATIONS

Feb 2026 – **Mila Quebec AI Institute** FATE & Society Reading Group. Invited talk: *Benchmarking Overton Pluralistic Alignment [C.1]*

Oct 2025 – **COLM NLP4Democracy Workshop**. Oral presentation: *Analyzing Collective Idea Evolution in Democratic Deliberation with LLMs* [W.2]

TEACHING EXPERIENCE

Kaufman Teaching Certificate Feb. 2025 – May 2025
MIT Teaching + Learning Lab *Cambridge, MA, USA*

- Developed teaching skills through eight practice-based workshops, evaluated through 2 microteaching sessions, received individual feedback from peers and teaching experts, and implemented evidence-based teaching techniques grounded in the scholarship of teaching and learning.
- Developed a syllabus for a course titled *Ethics, Fairness, and Bias in Generative Language Models*.

Teaching Assistant: Intro to Media Arts & Sciences Sept. 2024 – Dec. 2024
MIT Media Lab *Cambridge, MA, USA*
Teaching Assistant: Honours Algorithms & Data Structures Jan. 2022 – May 2022
McGill University *Montréal, QC, Canada*

AWARDS AND HONOURS

2025 – **AAAI 2026 Student Scholarship & Volunteer Award** ($2,100)
2023 – **Mila Excellence Scholarship - EDI in Research** ($5,000)
2023 – **Canadian Graduate Scholarship - Master's (NSERC)** ($17,500)
2020 – **Lev Bukhman Internship Award** ($1,070)
2019–2021 – **Dean's Honour List** (top 10% of class by GPA)
2019–2021 – **J W McConnell Scholarship** ($9,000)
2019 – **NY Cardiac Center Scholarship** ($20,000)
2019 – **National AP Scholar**
2019 – **National Merit Scholarship Award** ($2,500)
2018 – **Positive Coaching Alliance Triple-Impact Competitor Scholarship** ($1,000)

SELECTED PRESS

AI chatbots provide less-accurate information to vulnerable users. *MIT News*, Feb 19, 2026.

Some language reward models exhibit political bias. *MIT News*, Dec 10, 2024.

MENTORSHIP

Undergraduate Students
Jiayi Wu, Brown University (Summer + Fall 2025): Automated Benchmarking of Overton Pluralism with LLM-as-a-Judge [C.1, W.1]
Hannah Chiou, Wellesley College (Spring 2025): Computational Analysis of Public Narratives with LLMs [R.2]

High School Students
Abbey Hehman, John Jay High School (Spring 2026): Understanding the Impact of User Context on Biased Sycophantic Model Behavior

ACADEMIC SERVICE

Reviewer for ACL Rolling Review: October 2025 (EACL); May 2025 (EMNLP); December 2024 (ACL); October 2024 (NAACL; Emergency Reviewer)
Reviewer for AAAI: Social Impact Track 2026
Reviewer for ICLR: 2026

TECHNICAL SKILLS

Programming Languages: Python, Java, Javascript, C, Unix/Linux, OCaml, SQL
Machine Learning & Data Science: TensorFlow, PyTorch, Keras, scikit-learn, pandas, NumPy, matplotlib, seaborn, plotly
Cloud Computing: Google Cloud Platform, Amazon Web Services, Docker, slurm cluster