

Emilio Barkett

EDUCATION	Columbia University M.A. in Media Studies & Sociology Research Advisors: Diane Vaughan, James Chu GPA 4.01/4.00 Thesis: <i>Status Hierarchies in Language Models</i> Brigham Young University–Hawaii B.A. in Communication; Minor in Political Science Research Advisor: Mason Allred GPA 3.81/4.00	2024 – 2025 2022 – 2024
AREAS OF INTEREST	Language Model Behavior; AI Risk Management; Human-AI Interaction; Organizational Risk	
RESEARCH EXPERIENCE	Supervised Program for Alignment Research <i>Research Mentor</i> Columbia AI Alignment Club <i>Research Manager / Fellowship Facilitator</i> Columbia Business School <i>Summer Research Intern</i> Research Advisor: Alan Zhang Columbia University <i>Research Assistant</i> Research Advisor: James Chu	Berkeley, CA (<i>Remote</i>) Dec 2025 – Present New York, NY Jan 2025 – Present New York, NY June 2025 – Aug 2025 New York, NY Jan 2025 – July 2025
BOOK	[1] Outstanding Stories from Kiribati. Emilio Barkett <i>Barnes and Noble Press</i> , 2022. [link]	
PUBLICATIONS	[7] Whose Alignment? Process Alignment Across Organizational Contexts. Niklas Weller, Emilio Barkett <i>ICML Workshop</i> , 2026. [arXiv] (SPAR). [6] Representation Without Control: Testing the Realization Effect in Language Models. Ciaran Walsh, Emilio Barkett <i>Preprint</i> , 2026. [arXiv] (SPAR). [5] The Compulsory Imaginary: AGI and Corporate Authority. Emilio Barkett <i>Preprint</i> , 2026. [arXiv] [4] Status Hierarchies in Language Models. Emilio Barkett <i>Master’s Thesis</i> , 2026. [arXiv]. [3] Getting out of the Big-Muddy: Escalation of Commitment in LLMs. Emilio Barkett , Olivia Long, Paul Kröger <i>Preprint</i> , 2026. [arXiv]. [2] Don’t Change My View: Ideological Bias Auditing in Large Language Models. Paul Kröger, Emilio Barkett <i>Preprint</i> , 2026. [arXiv]. [1] Reasoning Isn’t Enough: Examining Truth-Bias and Sycophancy in LLMs. Emilio Barkett , Olivia Long, Madhavendra Thakur <i>ICML Workshop</i> , 2025. [arXiv].	

WORKS IN
PROGRESS

[4] Normalization of Deviance in AI Development.
Emilio Barkett
Working Paper.

[3] Evaluating Moral Reasoning in Embodied Language Model Agents.
Roshni Lulla, Saloni Modi, **Emilio Barkett**
Working Paper. (SPAR).

[2] Underrepresented Languages in LLMs.
Raghav Pant, Dan Sachs, **Emilio Barkett**
Working Paper. (SPAR).

[1] Illusion of Explanatory Depth in LLMs.
Hayzie Chu, Andrew Xie, Shruti Das, **Emilio Barkett**
Working Paper. (SPAR).

AWARDS &
FELLOWSHIPS

Supervised Program for Alignment Research, \$5,250.	Dec 2025 – Present
Claude Campus Ambassador, Anthropic, \$1,750.	Aug 2025 – Dec 2025
AI Policy Fellowship, Columbia AI Alignment Club.	Jan 2025 – June 2025
Service Scholarship, Brigham Young University–Hawaii, \$1,500.	2022
Dean’s List Scholarship, Brigham Young University–Hawaii, \$500.	2022

PRESENTATIONS

- Conferences**
- ICML 2026 Workshop, Seoul, South Korea, July 2026. *Whose Alignment? Process Alignment Across Organizational Contexts*
 - Beyond Our Screens: Reimagining Critical Media Literacy Symposium, University of Notre Dame, August 2025. *The Compulsory Imaginary: AGI and Corporate Authority*
 - ICML 2025 Workshop, Vancouver, Canada, July 2025. *Reasoning Isn’t Enough: Examining Truth-Bias and Sycophancy in LLMs*
 - Cultural Studies Association 2025 Conference, California Institute of the Arts, May 2025. *Sociotechnical Imaginaries in AGI Development*
 - Data Science Day, Columbia University, April 2025. *The Compulsory Imaginary: AGI and Corporate Authority*
 - AI Summit, Columbia University, March 2025. *The Compulsory Imaginary: AGI and Corporate Authority*
 - Artifacts of Control, Lincoln Center for the Performing Arts, Dec 2024. *Surveillance on Campus*
 - Undergraduate Research Conference, Brigham Young University–Hawaii, Dec 2024. *Media and War: A Genealogy of the Alternate Narrative*

ACADEMIC
SERVICE

- Program Committee Member**
- Workshops: AI for Social Impact (AAAI 2026)
- Reviewer**
- Conferences: NeurIPS (2025, 2026), IASEAI (2026)
 - Workshops: Pluralistic Alignment (ICML 2026), Models of Human Feedback for AI Alignment (ICML 2025), Socially Responsible Language Modeling Research (COLM 2025)

TECHNICAL
SKILLS

- Languages & Tools:**
Python, R, HTML, Git, $\text{\LaTeX}$, Figma, Qualtrics, Amazon MTurk, Prolific Academic, Open Science Framework, arXiv
- Human Languages:**
English (native), Gilbertese (advanced), American Sign Language (basic)