

Claire R. Leibowicz

Homepage: www.claireleibowicz.com

Updated May 2024

Education

PhD **The University of Oxford**, Oxford, UK, beginning part time in 2021
Information, Communication, and the Social Sciences, Oxford Internet Institute, Balliol College
Shirley Scholar

MSc **The University of Oxford**, Oxford, UK, 2017
Social Science of the Internet, Oxford Internet Institute, Balliol College
Clarendon Scholar

A.B. **Harvard University**, Cambridge, MA, 2016
Psychology (Cognitive Science Track), Computer Science Minor
Phi Beta Kappa, Magna Cum Laude with Highest Honors in Field

Primary Research Fields

Technology and society, responsible AI, AI governance, synthetic and manipulated media, digital culture, museum studies

Publications

Refereed Papers and Projects

- [6] Stray, J., Halevy, A., Assar, P., Hadfield-Mennell, D., Boutilier, C., Ashar, A., Bakalar, C., Beattie, L., Ekstrand, M., **Leibowicz, C.R.**, Moon Sehat, C., Johansen, S., Kerlin, L., Vickrey, D., Singh, S., Vrijenhoek, S., Zhang, A., Andrus, M., Helberger, N...Vasan, N. (2023). Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *ACM Transactions on Recommender Systems*, just accepted.
- [5] Saltz, E., Barari, S., **Leibowicz, C.R.**, Wardle, C. (2021). Misinformation interventions are common, divisive, and poorly understood. *Harvard Kennedy School Misinformation Review*, 2(5).
- [4] **Leibowicz, C.R.**, Saltz, E., Coleman, L. 2021. Creating AI art responsibly: A field guide for artists and designers. *Diseña, Special Issue on Visual Methods for Online Images: Collection, Circulation, and Machine Co-creation*, 5.
- [3] **Leibowicz, C.R.**, McGregor, S., Ovadya, A. 2021. The deepfake detection dilemma: A multistakeholder exploration of adversarial dynamics in synthetic media. *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*, 736-744.
- [2] **Leibowicz, C.R.**, Saltz, E., Coleman, L. 2021. How can artists use AI responsibly? *Proceedings of Art Machines 2: International Symposium on Machine Learning and Art*, 194-195.
- [1] Saltz, E., **Leibowicz, C.R.**, Wardle, C. 2021. Encounters with visual misinformation and labels across platforms: An interview and diary study to inform ecosystem approaches to misinformation interventions. *CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-6.

Refereed Papers Under Review and in Preparation

- [2] **Leibowicz, C.R.**, Real rules for fake media: Understanding multistakeholder synthetic media governance. In preparation.
- [1] **Leibowicz, C.R.**, Eccles, K. The art of looking in the 21st century. In preparation.

Non-Refereed Reports and White Papers

- [7] **Leibowicz, C.R.**, Cardona, C., 2024. From Principles to Practices: Lessons Learned from Applying PAI's Synthetic Media Framework to 11 Use Cases *Partnership on AI: Research Publication*.
- [6] **Leibowicz, C.R.**, 2023. Partnership on AI's Responsible Practices for Synthetic Media: A Framework for Collective Action. *Partnership on AI: Research Publication*.
- [5] **Leibowicz, C.R.**, Saltz, E., Coleman, L., 2020. A field guide to making AI art responsibly. *Partnership on AI: Research Publication*.
- [4] **Leibowicz, C.R.**, Stray, J., Saltz, E., 2020. Manipulated media detection requires more than tools: Community insights on what's needed. *Partnership on AI: Research Publication*.
- [3] **Leibowicz, C.R.**, 2020. The deepfake detection challenge: Insights and recommendations on AI and media integrity. *Partnership on AI: Research Publication*.
- [2] **Leibowicz, C.R.**, et al., 2019. Human-AI collaboration framework and case studies. *Partnership on AI: Research Publication*.
- [1] **Leibowicz, C.R.** 2015. The Case for the Classroom, in Student privacy: The next frontier - Emerging and future privacy issues in K-12 learning environments. *Berkman Klein Center Research Publication No. 2015-12, 10*.

Op-Eds and Selected Writing

- [17] **Leibowicz, C.R.**, Why watermarking AI-generated content won't guarantee trust online *MIT Tech Review*, August 9, 2023.
- [16] Stephan, A., **Leibowicz, C.R.**, How will AI change local news in the next 5 years? *PAI Blog*, January 28, 2022.
- [15] **Leibowicz, C.R.**, Stephan, A., Gregory, S. How to share the tools to spot deepfakes (without breaking them). *PAI Blog*, January 13, 2022.
- [14] **Leibowicz, C.R.**, Moon Sehat, C., Stephan, A., Stray, J. If we want platforms to think beyond engagement, we have to know what we want instead. *PAI Blog*, November 9, 2021.
- [13] Saltz, E., **Leibowicz, C.R.** Fact-Checks, Info Hubs, and Shadow Bans: A Landscape Review of Misinformation Interventions *PAI Blog*, June 14, 2021.
- [12] Shane, T., Saltz, E., **Leibowicz, C.R.** There are lots of ways to label AI content. But what are the risks? *First Draft News*, May 13, 2021.
- [11] Shane, T., Saltz, E., **Leibowicz, C.R.** From deepfakes to TikTok filters: How do you label AI content? *Nieman Lab*, May 12, 2021.
- [10] **Leibowicz, C.R.**, Preparing for a world of Holocaust deepfakes. *Tablet Magazine*, May 4, 2021.

- [9] Saltz, E., **Leibowicz, C.R.** Labeling Misinformation Isn't Enough. Here's What Platforms Need To Do Next. *PAI Blog*, March 11, 2021.
- [8] Saltz, E., **Leibowicz, C.R.**, Wardle, C. Warning Labels Won't Be Enough to Stop Vaccine Misinformation. *AI&*, February 18, 2021.
- [7] Saltz, E., Noel, P., **Leibowicz, C.R.**, Wardle, C., Gregory, S. 5 Urgent Considerations for the Automated Categorization of Manipulated Media. *AI&*, June 29, 2020.
- [6] Saltz, E., Shane, T., Kwan, V., **Leibowicz, C.R.**, Wardle, C. It Matters How Platforms Label Manipulated Media. Here are 12 Principles Designers Should Follow. *AI&, The Startup*, June 9, 2020.
- [5] **Leibowicz, C.R.**, Artificial intelligence is not the cure for the COVID-19 infodemic. *The Hill*, May 9, 2020.
- [4] **Leibowicz, C.R.** On AI and Media Integrity: Insights from the Deepfake Detection Challenge. *PAI Blog*, December 11, 2019.
- [3] **Leibowicz, C.R.**, Wardle, C. PAI and First Draft Launch Research Fellowship on Media Manipulation. *First Draft News*, December 2, 2019.
- [2] **Leibowicz, C.R.**, Gregory, S., Ellis, L. Protecting Public Discourse from AI-Generated Mis/Disinformation. *PAI Blog*, June 17, 2019.
- [1] **Leibowicz, C.R.**, Adler, S., Eckersley, P.D., When Is It Appropriate to Publish High-Stakes AI-Research? *PAI Blog*, April 2, 2019.

Presentations

Research Presentations at Conferences

- [8] **Leibowicz, C.R.** 2023. Real rules for fake media: Synthetic media governance in the era of deepfakes. *RightsCon*, San José, Costa Rica.
- [7] **Leibowicz, C.R.**, Gregory, S. 2023. Tackling synthetic media with content authenticity: Insights from civil society, academia, and industry. *International Communication Association Annual Conference*, Toronto, CA.
- [6] **Leibowicz, C.R.**, Saltz, E. 2021. Misinformation interventions are common, divisive, and misunderstood. How should they be designed? *Trust and Truth Online Conference*, London, UK (virtual format).
- [5] **Leibowicz, C.R.**, 2020. Why taking action on manipulated media requires diverse stakeholders. *Trust and Truth Online Conference*, London, UK (virtual format).
- [4] **Leibowicz, C.R.**, 2019. The art of looking in the 21st century. *American Psychological Association: Technology, Mind, and Society Conference*, Washington, DC.
- [3] **Leibowicz, C.R.**, Cikara, M., 2018. Using transcranial direct current stimulation to examine the dorsolateral prefrontal cortex's role in intergroup empathy. *Social and Affective Neuroscience Society Conference*, New York, NY.
- [2] **Leibowicz, C.R.**, 2018. The art of looking in the 21st century. *Theorizing the Web*, New York, NY.
- [1] **Leibowicz, C.R.**, Cikara, M. 2016. Using transcranial direct current stimulation to examine the dorsolateral prefrontal cortex's role in intergroup empathy. *Conte Center, Harvard Brain Science Initiative, and Center for Brain Science Neuroscience and Mental Health Poster Session*, Cambridge, MA.

Invited Lectures, Talks, and Panels

- [40] May 2024, **Speaker:** Exposing Deepfakes and Disinformation, *German Bundestag, Committee on Digital Affairs*, Berlin, Germany (virtual participation).
- [39] May 2024, **Speaker:** On Trust in Digital Media, *Adobe Content Authenticity Initiative*, New York, NY.
- [38] April 2024 **Expert Testimony:** U.S. Courts Advisory Committee on Evidence Rules, *U.S. Department of Justice*, Washington, D.C.
- [37] March 2024. **Guest Lecturer:** Media in U.S. Politics, *Columbia University*, New York, NY.
- [36] December 2023. **Panelist:** Chasing AI: Threatening or Enhancing the News, *NewsTECHForum*, New York, NY.
- [35] October 2023. **Panelist:** Multistakeholder responses to deepfakes and synthetic media, *Oxford Generative AI Summit*, Oxford, UK.
- [34] September 2023. **Panelist:** Are You for Real? Identifying and Detecting AI-Generated Content, *Copyright + Technology Conference*, New York, NY.
- [33] September 2023. **Panelist:** Harnessing Generative Artificial Intelligence for Learning, Teaching, and Working, *Harvard University 30th Presidential Inauguration Academic Symposium*, Cambridge, MA.
- [32] September 2023. **Panelist:** AI and Film: Bridging the Gap Between Innovation and Responsibility, *Toronto International Film Festival*, Toronto, CA.
- [31] September 2023. **Panelist:** Exploring AI Detection and Potential, *Democracy Reporting International*, Berlin, Germany (virtual participant).
- [30] June 2023. **Panelist:** Understanding child safety risks with generative AI, *Tech Coalition and Thorn*, (virtual format).
- [29] June 2023. **Panelist:** Media Literacy in the AI Age, *VivaTechnology*, Paris, France.
- [28] May 2023. **Panelist:** AI, Misinformation and Elections in Conflict Affected Contexts, *Alliance for Peacebuilding*, Washington, D.C.
- [27] April 2023. **Moderator:** Authentic storytelling for the era of generative AI, *Adobe*, (virtual format).
- [26] March 2023. **Panelist:** ChatGPT and DALL-E: What Generative AI means for Journalism *Associated Press*, New York, NY (virtual format).
- [25] March 2023. **Panelist:** Real rules for fake media: Synthetic media governance in the era of deepfakes and Dall*E *South by Southwest*, Austin, TX.
- [24] October 2022. **Panelist:** The deeper the fake, the more dangerous: The disinformation potential of text-to-image generation *Democracy Reporting International*, Berlin, Germany (virtual format).
- [23] June 2022. **Keynote:** Taking Action on Manipulated Media with Diverse Stakeholders *CVPR Workshop on Media Forensics*, New Orleans, LA (virtual format).
- [22] March 2022. **Speaker:** Responsible Deepfake Detection, *BBC AI/ML Seminar Series*, London, England (virtual format).
- [21] January 2022. **Distinguished Engineer Seminar Series Speaker:** Creating AI Art Responsibly, *SONY AI*, Tokyo, Japan (virtual format).
- [20] October 2021. **Panelist:** AI and Local News, *NYC Media Lab*, New York, NY (virtual format).

- [19] September 2021. **Moderator:** Deepfakes and Holocaust Testimony: What is the Future of Truth and Denial?, *USC Shoah Foundation and Partnership on AI*, Los Angeles, CA (virtual format).
- [18] August 2021. With Gregory, S. **Speaker:** Multistakeholder Solutions for Deepfakes and Visual Disinformation, *DARPA Semantic Forensics (SemaFor) Program Computational Disinformation Workshop*, New York University, New York, NY (virtual format).
- [17] May 2021. **Speaker:** On the Individual and Societal Dimensions of Synthetic Media, *Stanford Institute for Human-Centered Artificial Intelligence (HAI) GAN Workshop*, Stanford University, Stanford, CA (virtual format).
- [16] March 2021. **Seminar Speaker:** Multistakeholder Solutions for Deepfakes and Visual Disinformation, *Center for Informed Democracy and Social-cybersecurity (IDeaS)*, Carnegie Mellon University, Pittsburgh, PA (virtual format).
- [15] March 2021. **Panelist:** Show Me the Truth: Countering Disinformation and Deepfakes with Digital Content Attribution, *NYU Center for Cybersecurity New York University*, New York, New York (virtual format).
- [14] March 2021. With Saltz, E., Coleman, L. 2021. **Session:** How to Use AI for Your Art Responsibly, *Mozilla Festival "MozFest," Mozilla Foundation* (virtual format).
- [13] December 2020. With Gregory, S. **Keynote:** Threat Modeling and Access, *DARPA Semantic Forensics (SemaFor) Program Computational Disinformation Workshop*, New York University, New York, NY (virtual format).
- [12] November 2020. **Panelist:** Deepfakes. *IDeaS Conference, Center for Informed Democracy and Social-cybersecurity (IDeaS)*, Carnegie Mellon University, Pittsburgh, PA (virtual format).
- [11] November 2020. **Invited Lecturer:** AI and Information Integrity, *CS 59866: Ethical, Legal, and Policy Issues in Technology and Cybersecurity*, City College of New York: City University of New York, New York, NY (virtual format).
- [10] October 2020. **Panelist:** Manipulated Media in the US Elections - Key Takeaways, *Deepfakes and Elections: Should Europe be Worried*, hosted by *Democracy Reporting International*, Berlin, Germany (virtual format).
- [9] July 2020. **Moderator:** Deepfakes in the 2020 U.S. Elections - A Conversation Among Technologists and Industry, *Addressing Deepfakes in the U.S. 2020 Elections*, hosted by *UW Center for an Informed Public and Microsoft*, Seattle, WA (virtual format).
- [8] July 2020. **Panelist:** How to Use AI for Your Art Responsibly, *Gray Area*, San Francisco, CA (virtual format).
- [7] March 2020. **Participant:** Disinformation and Election Integrity Roundtable, *Transatlantic Commission on Election Integrity (TCEI)*, *Alliance of Democracies Foundation*, San Francisco, CA.
- [6] January 2020. **Participant:** Content Authenticity Initiative Inaugural Summit, *Adobe, Twitter, The New York Times*, San Jose, CA.
- [5] December 2019. **Panelist:** The Deepfake Detection Challenge Launch Event at NeurIPS, *Facebook*, Vancouver, CA.
- [4] November 2019. **Session Facilitator:** Synthetic and Manipulated Media Meeting 2, *Carnegie Endowment for International Peace*, Sunnyvale, CA.

- [3] June 2019. **Session Facilitator:** Synthetic and Manipulated Media Meeting 1, *Carnegie Endowment for International Peace*, San Francisco, CA.
- [2] November 2018. **Panelist:** AI Ethics Case Studies, *Berkman Klein Center, MIT, and Princeton Workshop on AI and Ethics*, Cambridge, MA.
- [1] April 2018. **Panelist:** Curbside Coaching: Silicon Valley and Toronto Tech Meet Civil Society, *RightsCon*, Toronto, CA.

Previous Research Experience

Research Assistant, Harvard Intergroup Neuroscience Laboratory
Department of Psychology, Harvard University, Cambridge, MA
Dr. Mina Cikara, 2014–17

Research Assistant, Youth and Media Lab
Berkman Klein Center for Internet and Society, Harvard Law School, Cambridge, MA
2014–16

Dean's Office Summer Research Fellow
Department of Management, Columbia Business School, New York, NY
Dr. Adam Galinsky, 2014

Research Assistant
Negotiation, Organizations, and Markets Unit, Harvard Business School, Cambridge, MA
Dr. Amy Cuddy, 2013–14

Simons Summer Research Fellow, Laboratory for Computational Neurodiagnostics
Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY
Dr. Lilianne Mujica-Parodi, 2011

Fellowships and Awards

Dieter Schwarz Fellowship, Dieter Schwarz Foundation, 2022–2024

Bellagio Residency Fellowship, The Rockefeller Foundation, 2022

Shirley Scholarship, Oxford Internet Institute, 2021–2027

Journalism Fellow, Tablet Magazine, 2021

Poster Award, Social and Affective Neuroscience Society Annual Conference, 2018

Clarendon Scholarship, The University of Oxford, 2016

Phi Beta Kappa, Alpha Iota Chapter, Harvard University, 2016

Gordon W. Allport Thesis Prize, Department of Psychology, Harvard University, 2016

John Harvard Scholar, Harvard University, 2016

Abramson Fellowship, Harvard College Research Program, Harvard University, 2016

Mind, Brain, and Behavior Summer Research Fellowship, Harvard University 2015

Dean's Office Summer Research Fellowship, Columbia Business School, 2014

Intel (now Regeneron) Science Talent Search Semi-Finalist, Society for Science and the Public, 2012
Simons Summer Research Fellowship, Stony Brook University, 2011

Professional Activities

UNESCO, AI and Holocaust Denial Report, Consultant/Reviewer, April 2024

Media

ABC News, *AI influencers explode on social media. Some are controlled by teens* , March 22, 2024.

Scripps News, *A Royal Conspiracy* , March 12, 2024.

CBC News, *What's wrong with this picture? Plenty, say experts* , March 12, 2024.

TechCrunch, *Women in AI: Claire Leibowicz, AI and media integrity expert at PAI* , March 9, 2024.

DW News, *How to spot AI videos by Sora* , March 9, 2024.

DW News, *Fact Check: Are these influencers real – or virtual?* , January 24, 2024.

WIRED, *The Biden Deepfake Robocall Is Only the Beginning* , January 23, 2024.

IEEE Spectrum, *Content Credentials Will Fight Deepfakes in the 2024 Elections* , December 27, 2023.

Remezcla, *Deepfake or Art? The Ethics Behind Bad Bunny's AI Viral Song* , November 29, 2023.

The New York Times, *'A.I. Obama' and Fake Newscasters: How A.I. Audio Is Swarming TikTok* , October 13, 2023.

The Wall Street Journal, *Real or AI? The Tech Giants Racing to Stop the Spread of Fake Images* , September 15, 2023.

Bloomberg, *AI 'Watermarking' Tools Emerging to Tag Machine-Made Content* , August 7, 2023.

Axios, *Meta and Microsoft join AI standards group on "synthetic media"* , June 14, 2023.

Yahoo! Finance, *12 Ways Web3 Media Could Embrace AI* , June 7, 2023.

Bloomberg, *Deepfakes Are Running Rampant as Tools to Detect Them Lag Behind* , April 20, 2023.

CNN, *Puffer coat Pope. Musk on a date with GM CEO. Fake AI 'news' images are fooling social media users* , April 2, 2023.

IEEE Spectrum, *Can Big AI Make Responsible AI? Major tech companies grapple with guidelines for artificial intelligence* , March 19, 2023.

Associated Press, *From marketing to design, brands adopt AI tools despite risk*, March 8, 2023.

Tech Policy Press, *Mitigating the Ethical and Legal Risks of Synthetic Media and Generative AI*, March 5, 2023.

KQED Forum, *How to Wrap Our Heads Around These New Shockingly Fluent Chatbots*, March 2, 2023.

MIT Technology Review, *How to create, release, and share generative AI responsibly*, February 27, 2023.

ABC News WEAR-TV, *How Advancements in AI Technology Impact Education*, February 16, 2023.

Uzbek and Rica, *Remembering the Holocaust in the Age of Virtual Witnesses*, May 24, 2022.

HumAIIn Podcast, *How Platforms Leverage the Extended AI Community to Address Misinformation*, June 4, 2021.

Kpopcast, *Artificial Love: K-pop, AI, and Aespa's Concept*, February 10, 2021.

Casey Newton's Platformer, *Why labels on misinformation aren't working*, December 14, 2020.

The MacArthur Foundation, *Technology in the Public Interest Experts List*, October 7, 2020.

The Radical AI Podcast, *More Than Fake News: Fighting Media Manipulation*, October 7, 2020.

TechTarget, *The deepfake 2020 election threat is real, but containable*, September 25, 2020.

Consumer Reports, *Facebook Gave Tips to Spot Fake News. A Study Says They Work Surprisingly Well*, June 22, 2020.

IEEE Spectrum, *Facebook AI Launches its Deepfake Detection Challenge*, December 11, 2019.

Axios, *Social Media and the Truth*, August 20, 2019.

The Nonprofit Innovators Podcast, *Fixing the Big Problems in AI*, August 6, 2019.