

Ruth Elisabeth Appel

rappel@cs.stanford.edu | linkedin.com/in/rutheappel | github.com/ruthappel

EXECUTIVE SUMMARY

Award-winning researcher combining deep technical expertise in AI with empirical social science to shape safe AI development.

AREAS OF EXPERTISE

AI, trust and safety, bias, misinformation, persuasion, elections, causal inference, data science, wellbeing, ethics, privacy

SKILLS

Computer Skills	R, Python, C++, SQL, GitHub, Markdown, LaTeX, and Mendeley
Research Methods	NLP (e.g., AI evaluations, fine-tuning), causal inference, field experiments, survey experiments
Data Skills	Large-scale platform observational and experimental data, lab experiment data
Soft Skills	Cross-functional collaboration, research communication, policy engagement
Language Skills	German (native), English (fluent), French (advanced), Spanish (basic)

PROFESSIONAL EXPERIENCE

09/2019 – present Stanford, US	Stanford University Stanford Impact Labs Postdoctoral Fellow (08/2024 – present) <ul style="list-style-type: none">Led research on generative AI content moderation preferences to inform AI alignmentShowed how to approximate political neutrality in AI and evaluate AI for political biasCo-led project on strengthening third-party AI evaluationShowed how social media regulation can inform generative AI regulationLed research on deceptive campaigns and election interference as part of a large-scale academic-industry collaboration with Meta analyzing data from 250 million usersShaped research with Meta on testing misinformation interventions at scaleWorked with cross-functional stakeholders (e.g., research, engineering, legal) SAP Graduate Fellow in Science and Engineering & PACS PhD Fellow (09/2019 – 06/2024) <ul style="list-style-type: none">Led research uncovering why partisans disagree about content moderation in the USLed research on a scalable intervention to counter vaccine misinformationLed research on the ethics and privacy implications of new technologiesCore contributor to research on an intervention to improve youth wellbeing using an appCommunicated results with relevant stakeholders and the public via meetings with high-level decision-makers, conference presentations, popular press articles, and inputs to policy drafts Teaching Assistant (10/2021 – 12/2021, 01/2021 – 03/2021) <ul style="list-style-type: none">Led research methods and R programming sections and developed technical course materials
06/2020 – 09/2020 San Francisco, US	Google LLC Quantitative User Experience Research Intern <ul style="list-style-type: none">Analyzed large-scale Google Cloud user log data to provide UX recommendationsAdded to Google's research infrastructure by importing machine learning software packages
02/2019 – 06/2019 Durham, US	Duke University, Center for Advanced Hindsight Health Division Research Associate <ul style="list-style-type: none">Designed large-scale experiments to improve health behaviors such as vaccine uptake
09/2016 – 12/2016 New York, US	Delegation of the European Union to the United Nations Fifth Committee Section (Administration and Budget) Intern <ul style="list-style-type: none">Prepared and participated in high-level meetings on budget issues and the 2030 Agenda

EDUCATION

09/2019 – 06/2024 Stanford, US	Ph.D. in Political Communication (<i>Nathan Maccoby Dissertation Award</i>) Stanford University; Ph.D. Advisor: Jennifer Pan; GPA: 4.0 Dissertation: Countering Social Media Manipulation: Three Studies of Social Media Manipulation and Efforts to Counter It Dissertation Committee: Jennifer Pan, Jeffrey Hancock, Susan Athey, Dan Jurafsky (Chair) Course focus: Research methods, causal inference, science and interpersonal communication Selected courses: ECON293 ML and Causal Inference, ECON281 Experiments for Impact
09/2023 – 06/2024 Stanford, US	M.S. in Computer Science Stanford University, Department of Computer Science; Advisor: Dan Jurafsky; GPA: 4.0 Course focus: Artificial intelligence, natural language processing, data science Selected courses: CS224N NLP with Deep Learning, CS399 Political Bias in LLMs, CS246 Mining Massive Data Sets, CS145 Data Systems, CS161 Algorithms

09/2017 – 06/2019 **Master in Public Policy** (*summa cum laude*)
Paris, France Sciences Po, School of Public Affairs
Course focus: Emerging technologies and public policy, policy analysis

06/2018 – 08/2018 **Visiting Researcher**
New York, US Columbia University, Columbia Business School; Advisor: Sandra Matz

08/2013 – 10/2016 **B.Sc. Economics**
Mannheim, Germany University of Mannheim, Department of Economics; Minor: Political Science

PEER-REVIEWED PUBLICATIONS

- Appel, R. E.**, Pan, J., & Roberts, M. E. (2023). Partisan conflict over content moderation is more than disagreement about facts. *Science Advances*. <https://doi.org/10.1126/sciadv.adg6799>
★ top 5% of all research scored by Altmetric
- Matz, S. C., **Appel, R. E.**, & Croll, B. (2022). Privacy and ethics in the age of Big Data. In S. C. Matz (Ed.), *The psychology of technology: Social science research in the age of Big Data* (pp. 379–420). American Psychological Association. <https://doi.org/10.1037/0000290-012>
- Appel, R. E.**, & Matz, S. C. (2021). Psychological targeting in the age of Big Data. In *Measuring and Modeling Persons and Situations*. Elsevier. <https://doi.org/10.1016/b978-0-12-819200-9.00015-6>
- Matz, S. C., **Appel, R. E.**, & Kosinski, M. (2020). Privacy in the Age of Psychological Targeting. *Current Opinion in Psychology*, 31. <https://doi.org/10.1016/j.copsyc.2019.08.010>

WORKSHOP PAPERS

- Appel, R. E.** (2024). Generative AI Regulation Can Learn from Social Media Regulation. *NeurIPS Regulatable ML Workshop*, December 2024, Vancouver, Canada. <https://doi.org/10.48550/arXiv.2412.11335>

UNDER PEER REVIEW

- Appel, R. E.**, Kim, Y. M., Pan, J., Xu, Y., Thomas, D., Allcott, H., Barberá, P., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., Guess, A. M., González-Bailón, S., Iyengar, S., Lazer, D., Malhotra, N., Moehler, D., Nimmo, B., Nyhan, B., Rivera, C. V., Settle, J., Thorson, E., Tromble, R., Wilkins, A., Wojcieszak, M., Xiong, B., Jonge, C. K., Franco, A., Mason, W., Stroud, N. J., & Tucker, J. A. How deceptive online networks reached millions in the US 2020 elections. Under review at *Nature Human Behaviour*.
- Appel, R. E.**, Roozenbeek, J., Rayburn-Reeves, R., Corbin, J., Basol, M., Compton, J., & van der Linden, S. Psychological inoculation improves resilience to and reduces willingness to share vaccine misinformation. <https://doi.org/10.31234/osf.io/ek5pu>. Under review at *Scientific Reports*.
- Pei, R., Grayson, S. J., **Appel, R. E.**, Soh, S., Garcia, S., Huang, E., Jackson, M. O., Harari, G. M., & Zaki, J. Bridging the empathy gap: Reducing empathy misperceptions increases social connectedness. Under review at *Nature Human Behaviour*.
- Appel, R. E.** Generative AI Regulation Can Learn from Social Media Regulation. <https://doi.org/10.48550/arXiv.2412.11335>
- Fisher, J., **Appel, R. E.**, Park, C., Potter, Y., Jiang, L., Sorensen, T., Feng, S., Tsvetkov, Y., Roberts, M., Pan, J., Song, D., Choi, Y. Political Neutrality in AI is Impossible—But Here Is How to Approximate It. <https://doi.org/10.48550/arXiv.2503.05728>.
- Longpre, S., Klyman, K., **Appel, R. E.**, Kapoor, S., Bommasani, R., Sahar, M., McGregor, S., Ghosh, A., Blili-Hamelin, B., Butters, N., Nelson, A., Elazari, A., Sellars, A., Ellis, C., Sherrets, D., Song, D., Geiger, H., Cohen, I., McIlvenny, L., Srikumar, M., Jaycox, M., Anderljung, M., Johnson, N., Carlini, N., Mialhe, N., Marda, N., Henderson, P., Portnoff, R., Weiss, R., Westerhoff, V., Jernite, Y., Chowdhury, R., Liang, P., & Narayanan, A. In-House Evaluation Is Not Enough. Towards Robust Third-Party Evaluation and Flaw Disclosure for General-Purpose AI. <https://sites.google.com/view/thirdparty-ai-evaluation/third-party-ai>
★ featured in *WIRED* (2025, March) *Researchers Propose a Better Way to Report Dangerous AI Flaws*

SELECTED WORK IN PROGRESS

- Appel, R. E.**, Pan, J., & Roberts, M. E. How partisanship affects preferences for content moderation in large language models.
- Appel, R.E.**, Athey, S., Karlan, D., Koutout, K., Luca, M., Manjeer, U., Sacher, S., & Wernerfelt, N. Combating misinformation on social media.
- Minson, J., Erlich, A., Gans-Morse, J., Higgins, C., **Appel, R. E.**, Baumann, C., Bai, H., Chen, S., Thomander, S. D., Dorison, C. A., Halperin, E., Jost, J. T., Koenig, F., Lee, I., Lerner, J. S., Makarin, A., Moore, M., Nir, N., Radnitz, S., Risen, J. L.,

Schweitzer, M. E., Shirikov, A., Spohn, M., Syunyaev, G., Vlasceanu, M., Wang, K., & Zeldin, N. Combating information avoidance in wartime Russia: A multidisciplinary field experiment.

POLICY IMPACT AND PUBLIC ENGAGEMENT

Bommasani, R., Singer, S., **Appel, R. E.**, Cen, S., Cooper, A. F., Cryst, E., Gailmard, L. A., Gonzalez, J. E., Ho, D. E., Klaus, I., Lee, M. M., Liang, P., Reuel, A., Song, D., Spence, D., Wan, A., Wang, A., Zhang, D., Zittrain, J., Chayes, J. T., Cuéllar, M.-F., & Fei-Fei, L. (2025). Report of the Joint California Policy Working Group on AI Frontier Models. <https://www.cafrontieraigov.org/>

Longpre, S., **Appel, R. E.** (2025). General-Purpose AI Needs Coordinated Flaw Reporting. <https://crfm.stanford.edu/2025/03/13/thirdparty.html>

Appel, R. E. (2024). Strengthening AI Accountability Through Better Third Party Evaluations. <https://hai.stanford.edu/news/strengthening-ai-accountability-through-better-third-party-evaluations>

Appel, R. E. (2024). EU DSA Election Guidelines Input. <https://ruthappel.github.io/assets/pdf/EU-DSA-Election-Guidelines-Input.pdf>

Appel, R. E. (2023). It's not just about facts: Democrats and Republicans have sharply different attitudes about removing misinformation from social media. <https://theconversation.com/its-not-just-about-facts-democrats-and-republicans-have-sharply-different-attitudes-about-removing-misinformation-from-social-media-216809>

★ read 40k times

Schaake, M., **Appel, R. E.**, Duplichen, D. M., Einstein, L., Elhai, W., Dhafer Muhammad Faishal, M., ... Zhao, J. (2020). *Input on the European Commission White Paper "On Artificial Intelligence – A European approach to excellence and trust."* https://hai.stanford.edu/sites/default/files/2020-07/HAI_WhitePaper_v4B.pdf

Matz, S. C., **Appel, R. E.**, & Kosinski, M. (2019). Rethinking privacy in the age of psychological targeting. <https://blogs.lse.ac.uk/businessreview/2019/11/21/rethinking-privacy-in-the-age-of-psychological-targeting/>

INVITED TALKS

2024 Stanford Data Science Democracy Day, Stanford, US
Social Technologies Lab at Cornell Tech, New York, US
Hertie School, Berlin, Germany
Graphika, virtual
Atlantic Council DFRLab, virtual
Google Jigsaw, virtual

2023 NABE TEC, Santa Clara, US
Stanford Data Science Democracy Day, Stanford, US
Content Moderation and Free Speech on Social Media Workshop, Munich, Germany

2021 Promoting Research in Social Media and Health Symposium, San Francisco, US

CONFERENCE TALKS

Appel, R. E., Kim, Y. M., Pan, J., Xu, Y., ..., Kiewiet de Jonge, C., Franco, A., Mason, W., Stroud, N. J. & Tucker, J. A. (2025, January). *How deceptive online networks reached millions in the US 2020 elections*. Information and Misinformation in Elections 2025 Conference, Caltech, virtual.

Appel, R. E., Kim, Y. M., Pan, J., Xu, Y., ..., Kiewiet de Jonge, C., Franco, A., Mason, W., Stroud, N. J. & Tucker, J. A. (2024, September). *How deceptive online networks reached millions in the US 2020 elections*. Trust and Safety Research Conference, Stanford, US.

Appel, R. E., Pan, J., & Roberts, M. E. (2024, February). *Partisan conflict over content moderation is more than disagreement about facts*. SPSP Annual Convention, San Diego, US.

Appel, R. E., Pan, J., & Roberts, M. E. (2023, September). *Partisan conflict over content moderation is more than disagreement about facts*. Trust and Safety Research Conference, Stanford, US.

Appel, R. E., Pan, J., & Roberts, M. E. (2023, July). *Partisan conflict over content moderation is more than disagreement about facts*. IC2S2 Conference, Copenhagen, Denmark.

Appel, R. E., Basol, M., Rayburn-Reeves, R., Corbin, J., van der Linden, S., & Roozenbeek, J. (2022, May). *Bad Vaxx Online Game Intervention Increases Discernment and Reduces Willingness to Share Vaccine Misinformation*. ICA Conference, Paris, France.

Appel, R. E. (2019, August). *Privacy in the Age of Psychological Targeting*. SPSP Summer Psychology Forum, St. Louis, US.

INVITED POSTERS

Appel, R. E., Pan, J., & Roberts, M. E. (2023, March). *How Partisanship Influences Perceptions of Censorship*. D3 Conference on Inequality in the Digital Age at Harvard Business School, Cambridge, US.

CONFERENCE POSTERS

Appel, R. E., Pan, J., & Roberts, M. E. (2025, March). *How Partisanship Affects Preferences for Content Moderation in Large Language Models*. Designing Trustworthy AI in a Polarized World, The Hoover Institution, Stanford, US.

Appel, R. E., Pan, J., & Roberts, M. E. (2023, February). *How Partisanship Influences Perceptions of Censorship*. SPSP Annual Convention, Atlanta, US.

Appel, R. E., Basol, M., Rayburn-Reeves, R., Corbin, J., van der Linden, S., & Roozenbeek, J. (2022, February). *Bad Vaxx Online Game Intervention Increases Discernment and Reduces Willingness to Share Vaccine Misinformation*. Psychology of Media and Technology Preconference at the SPSP Annual Convention, San Francisco, US.

Appel, R. E. (2020, February). *Privacy in the Age of Psychological Targeting*. SPSP Annual Convention, New Orleans, US.

SELECTED PRIZES AND AWARDS

06/2024	Stanford Communication Department Nathan Maccoby Dissertation Award (\$1,000)
10/2023	LLM Agents Hackathon at AGI House, 3 rd prize for personalized AI shopping agent

SELECTED FELLOWSHIPS AND GRANTS

09/2024 – present	Stanford Impact Labs Postdoctoral Fellowship
09/2024 – 09/2025	Stanford High Impact Technology Fund Award (\$38,000)
04/2024 – 04/2025	Stanford TomKat Center Innovation Transfer Grant (\$30,000)
12/2023 – 12/2024	Center on Philanthropy and Civil Society Grant (\$2,000)
09/2019 – 06/2024	SAP Stanford Graduate Fellow in Science and Engineering (\$52,920 p.a.)
09/2022 – 07/2023	Siegel Family Endowment Research Fellow
09/2022 – 06/2023	Center on Philanthropy and Civil Society PhD Research Fellow (\$79,650)
06/2022 – 09/2022	Stanford Impact Labs Summer Collaborative Research Fellow (\$4,000)
02/2020 – 01/2021	Stanford Center for American Democracy Fellow (\$2,000)
01/2014 – 06/2019	Student Fellow of the German Academic Scholarship Foundation (€27,665)

SELECTED ACTIVITIES AND VOLUNTEER WORK

07/2020 – 06/2024 Stanford, US	Communication for Diversity Group Member <ul style="list-style-type: none">• Advised PhD applicants from underrepresented groups and organized information events
05/2020 – 06/2024 Stanford, US	Stanford German Student Association Vice President <ul style="list-style-type: none">• Co-led Transatlantic Summit on geopolitics and tech (200 participants, \$150,000 budget)• Met with international leaders from politics and business
09/2023 – 06/2024 01/2020 – 05/2021 Stanford, US	Wonks and Techies at Stanford University Group Member <ul style="list-style-type: none">• Co-authored a white paper with suggestions for the EU AI strategy• Participated in discussions and a talk series on ethics, technology and policy
07/2017 – 06/2019 Multiple locations, Germany	German Academic Scholarship Foundation Seminar Organizer, Seminar Participant <ul style="list-style-type: none">• Organized research seminar on democracy and digitization (80 participants, highly positive feedback)• Participated in research seminar on digital ethics, conducting research on the future of work
09/2013 – 06/2015 Mannheim, Germany	UNICEF Student Group Mannheim Executive Board Member, Donor Relations Team Member <ul style="list-style-type: none">• Coordinated large donation campaigns (e.g., student run: 600 runners, €10,000 donations)• Represented the group at national UNICEF conferences

PROFESSIONAL SERVICE

Reviewer	Nature Communications, PNAS Nexus, Humanities and Social Sciences Communications, Journal of Politics, Journal of Elections, Public Opinion & Parties, British Journal of Political Science, ACM Designing Interactive Systems (DIS)
-----------------	--