David Manheim

Email / **LinkedIn** / **Google Scholar**

February, 2026

Academic and Research Experience

Head of Research and Policy, Association for Long Term Existence and Resilience, Israel **2022-26**

- Created and led research fellowship program, performed and managed public policy research projects
- Contributed to AI standards development with NIST AISIC, EU Code of Practice, and ISO/IEC JTC1 SC42
- Co-PI on RAND Corporation Delphi elicitation project on Limits of Biorisk from AI (in progress)
- Engaged with international bodies and government on health policy and biological weapons policy.
- Worked with Knesset members to require universal iodization of salt.

Visiting Lecturer, Technion **2021-26**

- Collaborations and research on ethics, public policy, and evaluating benefits and risks of technologies.

Lead Researcher, 1DaySooner **2020-22**

- Research on risk and ethical considerations for Human Challenge Trials (HCTs) for COVID-19 (1, 2, 3, 4).
- Policy engagement with US and UK administrations on use of HCTs for future pandemics.

Independent Grant-Funded Research **2017-22**

- Multiagent Failure Modes for Artificial Intelligence, AI Forecasting, and related research
- Primary Investigator for Modeling Transformative AI Risks project (2020-2022)
- Risk and risk reduction policy analyses for pandemics, biological warfare, and other global catastrophic risks (Contract work for Future of Humanity Institute, Oxford University, 2018-2022)
- Expert elicitation for large-scale and existential pandemic risks (2017-18 Grant)

Post-Doctoral Researcher, University of Haifa, Health and Risk Communication Research Center **2019-20**

- Understanding NGO and government organization social media use related to MMR and measles (link).

Visiting Researcher, Future of Humanity Institute, Oxford University **2018, 19**

- Risk and risk reduction policy analyses for pandemics, biological warfare, and other existential risks.

PhD, Public Policy and Decision Analysis, RAND Corporation / Pardee Rand **2012-17**

Dissertation – Value of Information for Policy Analysis (terrorism risk evaluation and disease biosurveillance)

- Infectious Disease Modeling and Decision Support (1, 2)
- Terrorism Risk and Cryptocurrency (1, 2, Op-Ed)
- Resilience for US Space Command (1, 2, 3)

BS, Mathematics, Lander College (Coursework at Georgia Institute of Technology, Hebrew University)

Fellowships, Residencies, and Grants

ARIA UK - Safeguarded AI, 2024-2025

Accelerator for Systemic Risk Assessment (ASRA), Member, 2025-Present

Accelerator for Systemic Risk Assessment (ASRA) - NextGen Fellow, 2024,

Survival and Flourishing Fund Grants, 2022-2024

Long Term Futures Fund, Multiple grants, 2020-2022

Foresight Fellow in Positive Long Term Futures, Foresight Institute, 2021

Postdoctoral Researcher, Health and Risk Communication Research Center, University of Haifa, 2019-2020

University of Oxford, Future of Humanity Institute. Visiting Researcher, Spring 2018.

Berkeley Existential Risk Initiative, 2018, Grant on Multi-agent machine learning dynamics

Open Philanthropy, 2017, Grant on Existential Risk from Infectious Disease.

Publications

Manheim, D., Homewood, A. Limits of Safe AI Deployment: Differentiating Oversight and Control. 3rd International AI Governance Workshop, in conjunction with AAAI 2026.

Manheim, D. Language Models' Hall of Mirrors Problem: Why AI Alignment Requires Peircean Semiosis (2026) Philosophy & Technology, 39 (9). https://link.springer.com/article/10.1007/s13347-025-00975-5

Manheim, D., Williams, A., Aveggio, C., Berke, A. Understanding the Theoretical Limits of AI-Enabled Pathogen Design: Insights from a Delphi Study. (2025) Rand Health Quarterly 13 (1), 10

Meusel, I., Manheim, D., Delaney, O., Greene, D,. Tobolsky, R., Palya, H., Shapiro, N., Sharma S. A Metagenomic Biosurveillance Network for Emerging Infectious Diseases: A Simulation-Based Model (2025) Health Security, (Forthcoming June 2026)

Manheim, D., Martin, S., Bailey, M., Samin, M., Greutzmacher, R. The Necessity of AI Audit Standards Boards. (2025) AI & Society https://link.springer.com/article/10.1007/s00146-025-02320-y

Elsey, J., Manheim, D. Marsh, A., Schmit, G., Moss, D. Ethical Acceptability of Human Challenge Trials: Consultation with the US public and with research personnel. (2024) PloS one 19 (10), e0307808

G Dal Prá, G., Chan, C., Burkhanov, T., et al and Manheim, D. A Horizon Scan of Global Catastrophic Risks (Under Review, 2024)

Manheim, D. Building less-flawed metrics: Understanding and creating better measurement and incentive systems. Patterns, Volume 4, Issue 10, 2023, https://doi.org/10.1016/j.patter.2023.100842

Winter, Christoph, Nick Hollman, and David Manheim. "Value alignment for advanced artificial judicial intelligence." American Philosophical Quarterly (2023) https://doi.org/10.5406/21521123.60.2.06

Liang, C., Wagstaff, J., Aharony, N., Schmit, V., and Manheim, D. Managing the Transition to Widespread Metagenomic Monitoring: Policy considerations for future biosurveillance. Health Security, Jan 2023 https://doi.org/10.1089/hs.2022.0029

Adams-Phipps J, Toomey D, Więcek W, Schmit V, Wilkinson J, Scholl K, Jamrozik E, Osowicki J, Roestenberg M, Manheim D. A Systematic Review of Human Challenge Trials, Designs, and Safety. Clin Infect Dis. 2022 Oct 11:ciac820. https://doi.org/10.1093/cid/ciac820 PMID: 36219704.

Waites W., Cavaliere M., Danos V., Datta R., Eggo R. M., Hallett T. B., Manheim D., Panovska-Griffiths J., Russell T. W. and Zarnitsyna V. I. (2022) "Compositional modelling of immune response and virus transmission dynamics" Phil. Trans. R. Soc. A.3802021030720210307 http://doi.org/10.1098/rsta.2021.0307

Manheim, D., and Lewis, G. High-Risk Human-Caused Pathogen Exposure Events from 1975-2016 (2022) F1000Research, https://f1000research.com/articles/10-752

Rando HM, Brueffer C, Lordan R, et al. Molecular and Serologic Diagnostic Technologies for SARS-CoV-2. Arxiv. 2022 Apr:arXiv:2204.12598v2. PMID: 35547240; PMCID: PMC9094103.

Clarke, S., Cottier, B., Englander, A., Eth, D., Manheim, D., Dylan Martin, S., & Rice, I. (2022). Modeling Transformative AI Risks (MTAIR) Project--Summary Report. arXiv preprint arXiv:2206.09360. 2022 Jun 19.

Manheim, D. Ancient and Modern Charity; Orthodox Judaism Considering Effective Altruism, Chapter in "Effective Altruism and Religion: Synergies, Tensions, Dialogue," 2022, Dominic Roser (Ed.) Stefan Riedener (Ed.) Markus Huppenbauer (Ed.) https://doi.org/10.5771/9783748925361

Manheim, D. (2021) “Results of a 2020 Survey on Reporting Requirements and Practices for Biocontainment Laboratory Accidents” *Health Security 19 (6)*, https://doi.org/10.1089/hs.2021.0083

Manheim, D., Wiecek, W., Schmit, V., & Morrison, J. (2021). Exploring Risks of Human Challenge Trials for COVID-19. *Risk Analysis* 41 (5), 710-720. https://doi.org/10.1111/risa.13726

Waites, W., Cavaliere, M., Manheim, D., Panovska-Griffiths, J., & Danos, V. (2021). Rule-based epidemic Models. Journal of Theoretical Biology, 530, 110851.

Gruetzemacher, R., Dorner, FE., Bernaola-Alvarez, N., Giattino, C., Manheim, D. (2021). Forecasting AI progress: A research agenda Technological Forecasting and Social Change 170, 120909 https://doi.org/10.1016/j.techfore.2021.120909

Sturniolo, S., Waites, W., Colbourn, T., Manheim, D., & Panovska-Griffiths, J. (2021). Testing, tracing and isolation in compartmental models. *PLoS computational biology* 17(3), e1008633 https://doi.org/10.1371/journal.pcbi.1008633

Reed, T., Waites, W., Manheim, D., de Walque, D., Vallini, C., Gatti, R., Hallett, T.B., (2021). Five Ways that COVID-19 Diagnostics Can Save Lives : Prioritizing Uses of Tests to Maximize Cost-Effectiveness. World Bank Research and Policy Brief; No. 43. https://openknowledge.worldbank.org/handle/10986/35150

Nguyen, L. C., Bakerlee, C. W., et al and Manheim, D. (2020). Evaluating Use Cases for Human Challenge Trials in Accelerating SARS-CoV-2 Vaccine Development. *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciaa935

Colbourn, T., Waites, W., Panovska-Griffiths, J., Manheim, D., Sturniolo, S., Colbourn, G., et al. (2020). *Modelling the health and economic impacts of population-wide testing, contact tracing and isolation (PTTI) strategies for COVID-19 in the UK*.

Manheim, D., & Foster, D. (2020). *Option-based guarantees to accelerate urgent, high-risk vaccines: a new market-shaping approach.* F1000Research, 9(1154). https://europepmc.org/article/PMC/PMC7808051

Manheim, D. (2020). The Fragile World Hypothesis: Complexity, Fragility, and Systemic Existential Risk. *Futures*, *122*, 102570. https://doi.org/10.1016/j.futures.2020.102570

Manheim, D., & Gesser-Edelsburg, A. (2020). The Structure of Tweets about Vaccine Safety Between Health Organizations, Experts and the Public: Analyzing Risk Communication Conversations. *Disaster Medicine and Public Health Preparedness*, 1–25. https://www.cambridge.org/core/journals/disaster-medicine-and-public-health-preparedness/article/structure-of-tweets-about-vaccine-safety-between-health-organizations-experts-and-the-public-analyzing-risk-communication-conversations/EE76CCEEFD82AF76024CEC503CE069B9

Manheim, D., & Denkenberger, D. (2020). Review of potential high-leverage and inexpensive mitigations for reducing risk in epidemics and pandemics. *Journal of Global Health Reports*, *4*. (Submitted Nov 11, 2019, accepted February 24, 2020.) https://www.joghr.org/article/12530-review-of-potential-high-leverage-and-inexpensive-mitigations-for-reducing-risk-in-epidemics-and-pandemics

Manheim, D. (2019). Multiparty Dynamics and Failure Modes for Machine Learning and Artificial Intelligence. *Big Data and Cognitive Computing*, *3*(2), 21. https://doi.org/10.3390/bdcc3020021

Johansson, M. A., et al. And Manheim, D. (2019). An open challenge to advance probabilistic forecasting for dengue epidemics. *Proceedings of the National Academy of Sciences*, *116*(48), 24268–24274. https://doi.org/10.1073/pnas.1909865116

Dion-Schwarz, C., Manheim, D., & Johnston, P. B. (2019). *Terrorist Use of Cryptocurrencies: Technical and Organizational Barriers and Future Threats*. RAND Corporation, Santa Monica, CA. https://doi.org/10.7249/rr3026

Manheim, D. (2018). *Value of Information for Policy Analysis: Understanding and Performing Value of Information Estimation for Complex Systems*. Dissertation. The Pardee RAND Graduate School.

Manheim, D. (2018). Questioning Estimates of Natural Pandemic Risk. *Health Security*, *16*(6), 381–390. https://doi.org/10.1089/hs.2018.0039

Manheim, D., & Garrabrant, S. (2018). *Categorizing Variants of Goodhart's Law*. *ArXiv:1803.04585*

Paul, C., Clarke, C. P., Triezenberg, B. L., Manheim, D., & Wilson, B. (2018). *Improving C2 and Situational Awareness for Operations in and Through the Information Environment https://doi.org/10.7249/rr2489*

Lakens, D., et al. and Manheim, D. (2018). Justify your alpha. *Nature Human Behaviour*, *2*(3). https://doi.org/10.1038/s41562-018-0311-x

McLeod, G., Nacouzi, G., Dreyer, P., Eisman, M., Hura, M., Langeland, K., Manheim, D. Torrington, G. (2016). *Enhancing Space Resilience Through Non-Materiel Means*. *RAND Corporation, Santa Monica, CA.* https://doi.org/10.7249/rr1067

Langeland, K. S., Manheim, D., Mcleod, G., & Nacouzi, G. (2016). *How Civil Institutions Build Resilience*. https://doi.org/10.7249/j.ctt1btc0m7.7

Dreyer, P., Langeland, K., Manheim, D., McLeod, G., & Nacouzi, G. (2016). *RAPAPORT (Resilience Assessment Process and Portfolio Option Reporting Tool): Background and Method*. https://doi.org/10.7249/rr1169

Davis, P., Perry, W., Hollywood, J., & Manheim, D. (2016). *Uncertainty-Sensitive Heterogeneous Information Fusion: Assessing Threat with Soft, Uncertain, and Conflicting Evidence*. https://doi.org/10.7249/rr1200

Manheim, D., Willis, H. H., Shelton, S., Vardavas, R., & Rosoff, H. (2016). Understanding Biosurveillance Value of Information Using an Exploratory Tool. *Final Report to the National Biosurveillance Integration Center, RAND Corporation, Santa Monica, CA. (Non-Public)*

Manheim, D., Chamberlin, M., Osoba, O., Vardavas, R., & Moore, M. (2016) *Improving Decision Support for Infectious Disease Prevention and Control*. RAND Corporation, Santa Monica, CA. https://doi.org/10.7249/rr1576

Manheim, D. (2015). *Causal Models and Exploratory Analysis in Heterogeneous Information Fusion for Detecting Potential Terrorists*. RAND Corporation, Santa Monica, CA. https://doi.org/10.7249/WR1124

Davis, P. K., Manheim, D., Perry, W. L., & Hollywood, J. (2015). *Using causal models in heterogeneous information fusion to detect terrorists.* 2015 Winter Simulation Conference (WSC) 2586–2597. https://doi.org/10.1109/WSC.2015.7408367

Baron, J., O'Mahony, A., Manheim, D., & Dion-Schwarz, C. (2015). *National Security Implications of Virtual Currency: Examining the Potential for Non-state Actor Deployment*. RAND Corporation, Santa Monica, CA. https://doi.org/10.7249/rr1231

Dixon, L., Clancy, N., Bender, B., Kofner, A., Manheim, D., & Zakaras, L. (2013). *Flood Insurance in New York City Following Hurricane Sandy*. RAND Corporation, Santa Monica, CA, for the New York City Mayor's Office of Long-Term Planning and Sustainability. https://www.jstor.org/stable/10.7249/j.ctt5vjwhn

Service and Public Outreach

Grant Evaluator, Foresight Institute **2023-2026**

Summer Fellow Supervisor / Mentor, Pivotal Research Fellowship **2024**

Grant Advisor (Ad-hoc), Long Term Futures Fund **2020-2023**

Course Facilitator, Bluedot Impact **2022-2026**

Summer Fellow Supervisor / Mentor, Swiss Existential Risk Initiative **2022,2023**

Summer Fellow Supervisor / Mentor, Stanford Existential Risk Initiative **2021,2022**

Advisor, PIBBSS.ai **2021-2022**

Board of Scientific Advisors, 1DaySooner **2021-2022**

Journal Opinions and Non-Peer Reviewed Publications

R Gruetzemacher, K Kilian, I Kumar, D Manheim, M Bailey, et al. Envisioning a Thriving Ecosystem for Testing & Evaluating Advanced AI. NIST AI EO Comment, 2024.

Manheim, D. Monrad, J.T. *Letter: Reporting accidents at laboratories must be universal practice*. Financial Times, November 23, 2021.

Manheim, D., & Sandberg, A. (2021) *What is the upper limit of value?* https://philpapers.org/rec/MANWIT-6

Colbourn, T., Waites, W., Manheim, D., Foster, D., Sturniolo, S., Sculpher, M., … et al. (2020). *Modelling the health and economic impacts of different testing and tracing strategies for COVID-19 in the UK*.

Manheim, D. (2020). *Evolving ethics of COVID-19 challenge trials. The Lancet Infectious Diseases.*

Manheim, D. (2018). *Building Less Flawed Metrics: Dodging Goodhart and Campbell's Laws*. Preprint.

Manheim, D. (2017) *Prescientific Organizational Theory*. (V. Rao, Ed.) *Ribbonfarm: Experiments in Refactored Perception*, Vol. 2017

Manheim, D., Johnston, P., Baron, J., & Dion-Schwarz, C. (2017). Are Terrorists Using Cryptocurrencies? *Foreign Affairs*.

Manheim, D. (2016). *Goodhart's Law and Why Measurement is Hard* (V. Rao, Ed.). *Ribbonfarm: Experiments in Refactored Perception*, Vol. 2016.

Manheim, D., Raffensperger,J. Xu, J. *Regulating drone airspace using 'smart markets'*, TechCrunch. April 19, 2016

Manheim, D. *Oversight of Unsafe Systems via Dynamic Safety Envelopes.* ArXiv:1811.09246.

Manheim, D. *A Generative Bayesian Approach for Incorporating Biosurveillance Sources into Epidemiological Models*. https://doi.org/10.1101/328518

Manheim, D. *Self-Limiting Factors in Pandemics and Multi-Disease Syndemics*. BioRxiv, 401018. https://doi.org/10.1101/401018

Invited Talks and Presentations

- The Value of Information Sources in Complex Systems. (Poster) APPAM, 2017. *https://www.researchgate.net/publication/312056226_The_Value_of_Information_Sources_in_Complex_Systems*.

- Risk of Multipandemics, May 14, 2018 Future of Humanity Institute Biosecurity Research Group, Oxford University

- A Proposal to Accelerate Vaccine Production Now, April 14, 2020. Foresight Institute, California.

- Cooperation for AI, Slatestarcodex online Meetup, August 30, 2020

- The Value of Value of Information, Oct 16, 2020. Invited speaker, Economic Evaluation and Health Technology Assessment Speaker Series, Centre for Health Economics, University of York

- Goodhart's Law, Management, and Metrics for OODA Loops, October 28th, 2020 (Non-public)

- Visiting Speaker, Introduction to Value of Information, King's College London Mathematics School, November 30, 2020.

- Effective Altruism (and Judaism), December 20, 2020. Yakar Social Justice Beit Midrash, Jerusalem, Israel.

- Long Term Future and Existential Risks, December 21, 2020. Effective Altruism Israel Fellowship Speaker Series.

- How to Use Money To Do Things (COVID-19 Vaccine Finance Edition), March 17, 2021, Future of Humanity Institute, University of Oxford, E-Seminar Series.

- The Right Side of Maybe (Podcast) How David Manheim Forecasted COVID-19 Vaccines. Clay Graubard, Andrew Eaddy, and Michał Dubrawski, Apr 15, 2021

- It Could Have Been Worse: Understanding the Probability of Worst-Case Pandemic Risks, April 28, 2021 Bi-Weekly General Seminar, Tel Aviv University School of Public Health

-Ups and Downs of (Eu)catastrophes, November 17, 2021. Foresight Institute, California.

- The Interplay of Systemic Sociotechnical and Geopolitical Risks to Climate, RISK-KAN Workshop on Understanding and Modeling Complex Risks in Coupled Human-Environment Systems, February 8, 2022

- Whence Infinite Value? Or: Value is Finite Unless Physics is Wrong in Specific Ways. Technion departmental seminar, March 30, 2022.

- Avoiding the Next (Intentional) Pandemic: Israel and the Biological Weapons Convention. Seminar talk at the Minerva Center for the Rule of Law under Extreme Conditions. March 9, 2022

- Biosecurity policy in the age of AGI risk. EAGxVirtual, Center for Effective Altruism. October 23, 2022

- Israeli AI Regulation, December 8, 2022, Invited Speaker, Center for Security and Emerging Technology, Georgetown University.

- Panelist, "A Torah-based Discussion of Artificial Intelligence" May 16, 2023

- AI Ethics: Cooperating with Stochastic Parrots? April 9, 2024, Chalmers AI Research Centre, University of Chalmers

- Moderator, "AI Standards Setting: Mechanisms and Contributions for Safety" May 16th, 2024

- Panel Talk, "Convergence of Goals: Will Mitigating Extreme Risks Be Otherwise Beneficial?" Cambridge Conference on Catastrophic Risk, 17 September 2024.

- Presenter and Panel Member, "AIxBio Tools and Biological Weapons: Recent Research on Risks and Mitigations" at the Biological Weapons Convention Working Group Meeting, December 10, 2025.

- Invited Speaker "When Should We Worry about AI Being Used for Nefarious Purposes, and What Can We Do About It?" Council for Regulation of Research into Biological Disease Agents, Israel Ministry of Health, at Tel Aviv University. December 30, 2025.

Publication Peer Reviews

Reviewed articles for: PLOS ONE, Lancet Infectious Diseases, Healthcare, Journal of Applied Microbiology, Journal of the Royal Society Interface, Discover Biotechnology, Philosophy and Technology, Futures, and International Journal of Environmental Research and Public Health, and Journal of Biosafety and Biosecurity.

Non-Academic Job History

Senior Policy Advisor, Guarding Against Pandemics **2021-22**

- Policy engagement with, and advice for lobbying federal, state, and local governments in the United States.
- Policy for supporting Biological Weapons Convention, World Bank FIF, and other global governance.

Superforecaster, Good Judgement Project / Good Judgement, Inc. **2011-21**

Catastrophe and Systems Analyst, Everest Reinsurance **2011-12**

- Doubled efficiency by conceiving and executing project to parallelize SQL and C#-based pricing tool.
- Built mini-simulation for flood losses in Asian countries to assess non-modeled risk.

Terrorism Analyst, Risk Management Solutions **2009-11**

- Performed $100,000 of consulting analyses, supporting sales of more than $1 million annually.
- Led training for transition of RMS Probabilistic Terrorism Model from emerging risk team to model development team, spearheaded and trained team of a dozen client facing subject matter experts.

Product Controller, Equity Derivatives, Credit Suisse **2007-9**

- Calculated daily P&L, guidance to senior management for portfolios with notional value over $5b.
- Redesigned and automated inter-entity reconciliation process, reducing process time by 50%.