ANNUAL
REPORT
2023
+ Q1 2024

Centre for the
Governance of AI

Table of Contents

A Note From Our Director

This past year was the most consequential year for AI governance yet.

To note just a handful of major events from the year:

- The United Kingdom hosted the first global AI Safety Summit and launched the first national AI Safety Institute.

- The United States issued one of the most extensive Executive Orders in its history, directing agencies to begin to study and manage a wide range of risks from AI.

- The European Union prepared to pass the long-awaited EU AI Act—the world's most comprehensive framework for AI regulation to date—after a flurry of last-minute controversy and negotiation about how to regulate general-purpose AI.

This past year was the most consequential year for AI governance yet.

Policymakers around the world have, quite abruptly, awakened to the possibility that AI could have truly transformative effects. Concerns about risks have grown especially quickly over the past year. At the global AI Safety Summit, for example, 28 countries gathered to discuss potential risks ranging from misinformation to AI-enabled cyberattacks to the possibility that society will lose control of the technology it is building. Earlier in the year, the heads of the leading AI companies signed an open letter suggesting that future AI systems could even threaten human survival.

Leading institutions—including both governments and companies—now feel pressure to act.

However, by and large, these institutions are still unsure about what they should do. Many of the risks and benefits of AI remain uncertain, the high-level policy options are still hazy, and concrete policy proposals are in short supply. These institutions are also struggling to hire or train the experts they need to make progress on addressing these questions.

Fortunately, this is the challenge that the Centre for the Governance of AI ("GovAI") was founded to solve. Our mission—which we have been pursuing since 2016—is to positively shape AI's lasting impact on the world, by helping key institutions make better decisions. We pursue our mission in two ways. First, we produce research and expert advice to help inform these institutions. Second, we develop and place AI governance talent to help reduce expertise shortages.

The past year has made the importance of our mission unmistakeable. And—since we made a bet on the importance of AI, especially general-purpose AI, nearly a decade ago—we are now better placed than any other non-profit to execute on this mission. In many ways, the present moment is one for which we have been preparing for years.

This annual report describes what we have done over the past year—our most productive year yet—to seize the present moment.

BEN GARFINKEL
Director

> In many ways, the present moment is one for which we have been preparing for years.

II.
Organisational
Updates

2023 was GovAI's most
productive year yet.

Research

PROGRESS IN 2023

Throughout the year we produced a number of high-quality research outputs, helping to advance public conversations and policymakers' understanding around key issues in AI governance. We published the results of an extensive survey of AI experts from industry, civil society, and academia on **best practices in AGI safety and governance**. The survey revealed an impressive level of consensus for a range of measures that frontier AI companies should take to better manage risks. The results were shared widely among policymakers in the EU and UK.

Further, our report "**Open-Sourcing Highly Capable Foundation Models**" presented an extensive examination of the risks and benefits of open-sourcing powerful AI models. It proved timely, attracting the attention of policymakers as questions around open-sourcing became a focus of policy discussions in the US, UK, and EU. The report forms the basis of a forthcoming OECD report on open-source regulation and is referenced in the EU AI Act trilogue.

Several of our researchers participated in writing "**Frontier AI Regulation: Managing Emerging Risks to Public Safety**". This paper was among the earliest and most thorough treatments of the case for regulating frontier AI systems (while also noting its downsides), a policy question now receiving considerable attention from major governments. It has been cited in expert testimony to the US **House** and **Senate**, and in the **Interim International Scientific Report on the Safety of Advanced AI**.

We also put considerable effort into understanding the foundations of compute governance—identifying where, whether, and how computing hardware can be used as a lever to pursue AI policy goals. We produced papers on the **implications of increased compute efficiency** and on **Know-Your-Customer requirements for compute providers**, measures **now being explored** by the US government.

Over the past year, we have sought to increase the range of our research expertise and position ourselves to stay ahead of shifting policy priorities in AI governance. With this in mind, we hired Sam Manning, our first staff researcher focused on the impacts of labour automation, a topic that we expect to become increasingly important in the coming years.

> Over the past year, we have sought to increase the range of our research expertise and position ourselves to stay ahead of shifting policy priorities in AI governance.

> We will increase our efforts to make our research insights more accessible to key stakeholders, policymakers, and broader audiences.

AMBITIONS FOR 2024

Beyond continuing our efforts to produce and publish high-quality research outputs, our priority in 2024 will be to improve our research processes by hiring a dedicated Research Manager. The Research Manager will be responsible for maintaining a vibrant intellectual environment at GovAI, helping our researchers coordinate their efforts, supporting the operational side of research, and improving our ability to put our research into the hands of the most relevant stakeholders.

We will also increase our efforts to make our research insights more accessible to key stakeholders, policymakers, and broader audiences. Translating research into impact requires making it easier for audiences to quickly grasp our key findings and recommendations. We will put additional effort into creating summaries and compelling presentations of our work, as well as identifying other ways of making our work more useful to a wider range of stakeholders.

Talent Development

PROGRESS IN 2023

Our talent development work currently comprises two major programmes:

- Our **Research Scholar Programme** offers one-year visiting positions at GovAI. It is designed to support the career development of AI governance researchers and practitioners, while offering Scholars opportunities to do high-impact work.

- Our **Seasonal Fellowships** bring a cohort of early-stage researchers or individuals new to the field of AI governance to Oxford to spend three months working on an AI governance research project, learning about the field, and making connections with other researchers and practitioners.

The events of 2023 have reaffirmed the value of this work. First, demand for AI governance talent continues to grow at pace—we are seeing more openings for AI governance roles than ever before. Second, we continue to get more direct evidence of the value of these programmes. Among the first cohort of Research Scholars, one now leads a team at Demos, a major UK think tank; another heads a major workstream at RAND; and a third founded an organisation to facilitate AI governance dialogues between US and Chinese researchers.

Our surveys of Seasonal Fellowship participants show that they have overwhelmingly found the programme valuable, particularly with respect to developing their subject matter expertise and making useful connections within GovAI and the wider AI governance world. Fellows have gone on to take full-time roles in government, including at the UK AI Safety Institute; in think tanks, such as the AI Now Institute and RAND; and in leading AI labs, including Google DeepMind and OpenAI.

Partly because of these positive signals, we worked to expand these programmes over the last year. In 2023 we welcomed 24 researchers to our Seasonal Fellowship Programme, compared with 12 the previous year. And since the start of 2023, we have hosted 13 Research Scholars. The depth and diversity of skills continues to increase with each cohort. We have continued to refine both programmes to adapt to the evolving needs of the AI governance space, and in response to extensive feedback from participants and supervisors.

In addition to our current main programmes, we also piloted an eight-week Policy Programme in early 2023 to enable current and future US policy practitioners to get up to speed on AI, its risks, and the policy tools for mitigating these risks. The programme was structured around guided self-study, workshops, and seminars; it aimed to give participants legible expertise on the issues central to AI policy today and in the coming years. The outcome of that trial was encouraging, with many participants finding both the materials and exercises valuable.

AMBITIONS FOR 2024

Addressing the shortage of AI governance talent will remain one of our top priorities in 2024. There are several avenues we are exploring to increase our impact in this area.

First, we will explore whether it is feasible to significantly expand our Seasonal Fellowships, with the aim of training and placing even larger cohorts—in 2024 we are hosting 37 Fellows across the Summer and Winter Fellowships, roughly 50% more than in 2023. We will likely need to experiment with different structures and additional external supervision capacity to identify the most effective ways to quickly train larger numbers of promising researchers.

At the same time, we will look for ways to improve learning outcomes and the experience of our Fellows. While the post-Fellowship survey results are encouraging, we expect that we can implement many smaller and larger changes to help Fellows find better learning and placement opportunities after the Fellowship and ultimately provide better advice to relevant stakeholders.

Addressing the shortage of AI governance talent will remain one of our top priorities in 2024.

We also intend to experiment with structural changes to our Research Scholar programme, potentially providing more guidance and formal learning opportunities, as well as integrating Scholars more closely into existing workstreams to provide more hands-on experience and better opportunities for directly impactful work. This year, we expect to take on Scholars with a more diverse range of skills, with additional emphasis on technical expertise and experience in setting up research-adjacent projects.

In addition to scaling and improving our existing programmes, we plan to pilot at least one new talent programme. Most likely this will be an evolution of last year's Policy Programme. Other possibilities we are considering include a programme to facilitate mid- to late-career professionals transitioning into AI governance from adjacent fields.

To enable both the structural improvement of our existing programmes and the exploration of other possible programmes, we are considering opening a senior role to lead our overall talent efforts. This role—tentatively titled Head of Talent Development— would serve a critically important function in our organisation and could shape the direction of this workstream for years to come. We will therefore only hire for this position if we can identify an exceptional candidate whom we can trust to lead our talent development efforts towards the largest possible impact.

Policy Engagement

PROGRESS IN 2023

2023 saw tremendous growth in policymakers' interest in AI across the world. In many ways, this was the moment we had spent years preparing for. Doing our best to seize the opportunity, we substantially expanded our engagement with decision makers in government.

We substantially expanded our engagement with decision makers in government.

The UK government established itself as a global leader in AI governance, establishing an AI Safety Institute and convening the first global AI Safety Summit. GovAI was exceptionally well-positioned to contribute to these developments. We offered extensive advice to UK policymakers, producing a number of memos and supporting the Department of Science, Innovation and Technology with implementation capacity through secondments.

This work enabled us to engage not only with domestic AI policymaking, but also with the international governance efforts being spearheaded by the UK—we wrote a **blog post** on what the first AI Safety Summit should aim to accomplish, and offered remarks at the summit itself.

We have also continued to engage closely with policymakers in the US. In January 2024, GovAI joined the National Institute of Science and Technology's **AI Safety Institute Consortium**. We also provided advice on several key issues, including comments on semiconductor export controls for the Department of Commerce and inputs on **US national priorities for AI** and **AI accountability**. These efforts were complemented with in-person engagement—Ben Garfinkel and Markus Anderljung met with policymakers in Washington, DC, and Anton Korinek **testified** before the US Senate AI Insight Forum on the labour market impacts of advanced AI.

AMBITIONS FOR 2024

In 2024, as more institutions recognise the importance of AI policy and seek out expertise, we expect demand for our work to continue to grow. We will grow our team and increase our output to meet this demand and maximise our impact during this pivotal period.

In addition to building on the strong foundations we have created in the UK, we will seek to expand our engagement with policymakers in the US. The US is likely to be one of the most influential actors in the global AI governance landscape. We therefore hope to increase our presence there by hiring one or two policy researchers based in Washington, DC. If this initial exploration goes well, we may set up a permanent team in Washington, DC, over the coming years.

> In addition to building on the strong foundations we have created in the UK, we will seek to expand our engagement with policymakers in the US.

Organisational Capacity

PROGRESS IN 2023

Since the beginning of 2023, we have also continued to build out our non-research team, improving the level of service and support for our various workstreams. We hired a Talent Management Associate, Anna Ringvold, to specialise in running and improving our hiring processes. As a result, the efficiency of our hiring rounds has improved substantially, reducing the strain on other parts of the organisation.

We also hired a Head of Operations, Ryan Fugate, to improve the effectiveness of the team and free up time for our Chief of Staff, who had been covering this portfolio since the summer of 2023. In early 2024 we set out to increase our capacity further, hiring a Research Management Associate, Valerie Belu, to assume responsibility for planning and managing our Seasonal Fellowships. Valerie took over this crucial workstream from our outgoing Research Manager, Emma Bluemke. In addition, we hired William Ehlers as Operations Associate. We expect that these additional hires will substantially increase our implementation capacity.

AMBITIONS FOR 2024

As the urgency of our mission grows, we need to increase our organisational capacity further. In 2024, we will navigate a significant organisational transition as we spin out from our fiscal sponsor, Effective Ventures, and establish GovAI as an independent entity in both the US and the UK. The spin-out process will require significant staff time and careful coordination to ensure that it goes smoothly. Once complete, we expect that independence will enable us to more easily build a global presence, and to expand and deepen our relationships with key partners.

The spin-out also coincides with a push to build more robust operational capabilities under the leadership of our new Head of Operations. After a period of rapid team growth, 2024 will be an opportunity to consolidate our progress by strengthening our internal infrastructure. By investing in our organisational foundations, we can continue to grow our team and impact in a sustainable way.

By investing in our organisational foundations, we can continue to grow our team and impact in a sustainable way.

As part of our maturation as an organisation, we are also committed to diversifying our funding sources. Historically, we have relied on a small number of primary funders who provide us with multi-year support, with Open Philanthropy being our largest funder to date. While we are deeply grateful to these donors, we believe that cultivating a wider range of supporters is important for our organisational resilience. In 2024, we expect to hire a full-time fundraiser to support those efforts. Attracting new funding from both foundations and individual donors who share our mission is challenging and important work, and we expect that a dedicated hire will substantially increase our ability to do so.

As our team continues to grow in 2024, we are also likely to outgrow our current office space. We will begin the search for a new permanent office that can house our larger team and facilitate a productive work environment.

Notable Departures

We have bid farewell to core staff members Anne le Roux, Lennart Heim, and Emma Bluemke.

Since our last update, we have bid farewell to core staff members Anne le Roux, Lennart Heim, and Emma Bluemke.

Anne was part of the founding team when GovAI spun out of Oxford University in 2021 and played an invaluable role in setting up the new version of the organisation. She was GovAI's Head of Programme Operations before becoming our Head of Partnerships. Anne worked tirelessly to set up our early structures and played a pivotal role in evaluating a range of ideas for what an independent GovAI should focus on. She moved on in early 2023 to join Anthropic as Chief of Staff to the President.

Lennart is in many ways an advertisement for our talent development efforts. He initially joined GovAI on our Winter Fellowship Programme, with limited experience, then became a Research Scholar, before receiving a staff researcher position. In the space of just two years, he has become a leading figure in the field of AI governance. His advice is regularly sought by policymakers, and his extensive work on compute governance has shaped AI policy in both the US and UK. Lennart leaves to lead a technical governance workstream at RAND and will remain affiliated with GovAI as an Adjunct Fellow.

Emma led GovAI's research management portfolio for about two years, developing many processes and structures that have enabled our researchers to do their most high-impact work. Emma was instrumental in the development of GovAI's Seasonal Fellowship programmes. She developed the Fellowship from a four-person cohort with very little structure into a large, well-organised programme that offers value to Fellows not only through immediate supervision but also regular Q&A sessions, career planning advice, and opportunities for hands-on policy work. The work that Emma has done at GovAI will continue to benefit the organisation for many years to come. We are excited to see the important work that she will do in her next role building key new programmes at Anthropic.

Finances

We are grateful to all our funders whose generosity enables us to achieve our mission.

We are grateful to all our funders whose generosity enables us to achieve our mission. This past year we received large donations from Open Philanthropy, the Waking Up Foundation, the Fickel Family Foundation, and an anonymous donor, in addition to over $100,000 from small donors. Altogether, in 2023 we drew upon $4.2 million in donations from both recent donations and earlier commitments. Our expenditures for the year were $3.9 million, which went primarily towards staff and contractors ($2.5 million), overheads ($1.1 million), and travel and venue costs ($200,000).

At the time of writing, we are poised to secure a substantial new commitment from Open Philanthropy, which will cover the majority of our expenses over the next two years, including room for growth.

You can donate to us at **www.givingwhatwecan.org/charities/govai**. If you are interested in supporting our work and would like to develop a more detailed understanding of our funding needs and specific opportunities, please get in touch at **contact@governance.ai**.

In summary

The decisions made about AI governance in 2024 are likely to have effects that last for decades and shape the lives of billions.

2024 will be a year of exciting progress and evolution for GovAI across all of our priority areas. We will continue to meet the challenges and opportunities posed by the progress in AI. To achieve this goal, we will grow and develop our team, accelerate our research and policy work, and support the maturation of our organisational infrastructure. The decisions made about AI governance in 2024 are likely to have effects that last for decades and shape the lives of billions. We are committed to doing everything in our power to ensure the safe development of AI and steer it towards a positive future.

III.
Outputs

Since our last update, GovAI researchers have authored or contributed to over 60 research outputs.

Highlighted Research

Towards Best Practices in AGI Safety and Governance
Jonas Schuett, **Noemi Dreksler**, **Markus Anderljung**, **David McCaffary**, **Lennart Heim**, **Emma Bluemke**, and **Ben Garfinkel**

AGI labs should take steps to mitigate the serious risks posed by developing and deploying advanced AI systems, but best practices for safety are still emerging. To support the development of best practices, we surveyed leading experts from labs, academia, and civil society, presenting them with 50 statements about what AGI labs should do. We found remarkably high levels of agreement for a range of measures.

Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives
Elizabeth Seger, **Noemi Dreksler**, **Richard Moulange**, Emily Dardaman, **Jonas Schuett**, **K. Wei**, Christoph Winter, Mackenzie Arnold, Seán Ó hÉigeartaigh, **Anton Korinek**, **Markus Anderljung**, **Ben Bucknall**, **Alan Chan**, **Eoghan Stafford**, **Leonie Koessler**, Aviv Ovadya, **Ben Garfinkel**, **Emma Bluemke**, Michael Aird, Patrick Levermore, **Julian Hazell**, and Abhishek Gupta

We examine the risks and benefits of open-sourcing highly capable foundation models, a question facing AI labs that has sparked significant debate. While open-sourcing has historically provided substantial net benefits, we argue that for some highly capable foundation models likely to be developed in the near future, open-sourcing may pose sufficiently extreme risks to outweigh the benefits. We make recommendations for developers, standard-setting bodies, and governments for establishing safe and responsible model sharing practices and preserving open-source benefits where safe.

Frontier AI Regulation: Managing Emerging Risks to Public Safety
Markus Anderljung, Joslyn Barnhart, **Anton Korinek**, Jade Leung, Cullen O'Keefe, Jess Whittlestone, Shahar Avin, Miles Brundage, Justin Bullock, Duncan Cass-Beggs, Ben Chang, Tantum Collins, Tim Fist, Gillian Hadfield, Alan Hayes, Lewis Ho, Sara Hooker, Eric Horvitz, Noam Kolt, **Jonas Schuett**, Yonadav Shavit, Divya Siddarth, **Robert Trager**, and Kevin Wolf

Frontier AI models pose a distinct regulatory challenge: dangerous capabilities can arise unexpectedly; it is difficult to robustly prevent a deployed model from being misused; and it is difficult to stop a model's capabilities from proliferating broadly. Industry self-regulation is an important first step. However, wider societal discussions and government intervention will be needed to create standards and to ensure compliance with them. We consider several options to this end, including granting enforcement powers to supervisory authorities and licensure regimes for frontier AI models. Finally, we propose an initial set of safety standards.

Publications, Reports, and Working Papers

Scenarios for the Transition to AGI
Anton Korinek and Donghyun Suh

Governing Through the Cloud: The Intermediary Role of Compute Providers in AI Regulation
Lennart Heim et al.

How to Design an AI Ethics Board
Jonas Schuett, Ann-Katrin Reuel, and Alexis Carlier

Computing Power and the Governance of Artificial Intelligence
Girish Sastry et al.

Visibility into AI Agents
Alan Chan et al.

Hardware-Enabled Governance Mechanisms: Developing Technical Solutions to Exempt Items Otherwise Classified Under Export Control Classification Numbers 3A090 and 4A090
Gabriel Kulp et al.

Generative AI for Economic Research: Use Cases and Implications for Economists
Anton Korinek

Scenario Planning for an A(G)I Future
Anton Korinek

Three Lines of Defense Against Risks from AI
Jonas Schuett

- **Increased Compute Efficiency and the Diffusion of AI Capabilities**
  Konstantin Pilz, Lennart Heim, and Nicholas Brown

- **Towards Publicly Accountable Frontier LLMs: Building an External Scrutiny Ecosystem under the ASPIRE Framework**
  Markus Anderljung et al.

- **Trends in Machine Learning Hardware**
  Marius Hobbhahn, Lennart Heim, and Gökçe Aydos

- **Compute at Scale: A Broad Investigation into the Data Center Industry**
  Konstantin Pilz and Lennart Heim

- **Structured Access for Third-Party Research on Frontier AI Models: Investigating Researchers' Model Access Requirements**
  Benjamin Bucknall and Robert Trager

- **Oversight for Frontier AI Through a Know-Your-Customer Scheme for Compute Providers**
  Janet Egan and Lennart Heim

- **AI's Economic Peril**
  Stephanie A. Bell and Anton Korinek

- **Coordinated Pausing: An Evaluation-Based Coordination Scheme for Frontier AI Developers**
  Jide Alaga and Jonas Schuett

- **Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives**
  Elizabeth Seger et al.

- **Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT**
  Jai Vipra and Anton Korinek

- **International Governance of Civilian AI**
  Robert Trager et al.

- **Confidence-Building Measures for Artificial Intelligence: Workshop Proceedings**
  Sarah Shoker et al.

- **Risk Assessment at AGI Companies: A Review of Popular Risk Assessment Techniques from Other Safety-Critical Industries**
  Leonie Koessler and Jonas Schuett

- **Frontier AI Regulation: Managing Emerging Risks to Public Safety**
  Markus Anderljung et al.

- **International Institutions for Advanced AI**
  Lewis Ho et al.

- **Will Humanity Choose Its Future?**
  Guive Assadi

- **Auditing Large Language Models: A Three-Layered Approach**
  Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi

- **AGI Labs Need an Internal Audit Function**
  Jonas Schuett

- **Model Evaluation for Extreme Risks**
  Toby Shevlane et al.

- **Towards Best Practices in AGI Safety and Governance**
  Jonas Schuett et al.

- **Spear Phishing with Large Language Models**
  Julian Hazell

Policy Advice

■ **Response to the RFI Related to NIST's Assignments Under the Executive Order Concerning AI**
Jonas Schuett, Leonie Koessler, and Markus Anderljung

■ **Accessing Controlled AI Chips via Infrastructure-as-a-Service (IaaS): Implications for Export Controls**
Lennart Heim and Janet Egan

■ **Preparing the Workforce for an Uncertain AI Future: Written Statement to the US Senate AI Insight Forum on Workforce**
Anton Korinek

■ **National Priorities for Artificial Intelligence: Response to the OSTP Request for Information**
Jonas Schuett, Markus Anderljung, Lennart Heim, and Elizabeth Seger

■ **Evaluate Advanced AI Models for Potential Autocratic Misuse: Response to the OSTP Request for Information**
Eoghan Stafford

■ **Response to the NTIA AI Accountability Policy**
Everett Thornton, Jonas Schuett, Markus Anderljung, and Lennart Heim

■ **Response to the UK's Future of Compute Review**
Jess Whittlestone et al.

Blog Posts

Goals for the Second AI Safety Summit
Ben Garfinkel et al.

Preventing AI Misuse: Current Techniques
Ben Clifford

Summer Fellowship 2023 Wrap Up—What Did Our Fellows Work On?
GovAI

The Case for Including the Global South in AI Governance Discussions
Sumaya Nur Adan

Putting New AI Lab Commitments in Context
Shaun Ee and Joe O'Brien

What Should the Global Summit on AI Safety Try to Accomplish?
Ben Garfinkel and Lennart Heim

Proposing a Foundation Model Information-Sharing Regime for the UK
Nikhil Mulani and Jess Whittlestone

Beijing Policy Interest in General Artificial Intelligence is Growing
Fynn Heide

Winter Fellowship 2023 Wrap Up—What Did Our Fellows Work On?
GovAI

Preliminary Survey Results: US and European Publics Overwhelmingly and Increasingly Agree that AI Needs to be Managed Carefully
Noemi Dreksler et al.

Opinion Articles

- **To Govern AI, We Must Govern Compute**
  Lennart Heim, Markus Anderljung, and Haydn Belfield

- **Frontier AI Regulation: Safeguards Amid Rapid Progress**
  Markus Anderljung and Anton Korinek

- **How to Prevent an AI Catastrophe**
  Markus Anderljung and Paul Scharre

- **Chinese Firms Are Evading Chip Controls**
  Tim Fist, Lennart Heim, and Jordan Schneider

- **Exploring the Impact of Language Models on Cognitive Automation with David Autor, ChatGPT, and Claude**
  Anton Korinek

IV.
People

Our team has grown to 25 people across research and operations.

Core Researchers

Our core research team includes both staff researchers and academic researchers not officially employed by GovAI who are heavily integrated with the team.

BEN GARFINKEL
Director

Ben leads GovAI and is responsible for setting the direction of the organisation, making key decisions, and overseeing its research. His own research has focused on the security implications of AI, the causes of war, and the methodological challenge of forecasting risks from technology. He earned a BS in Intensive Physics and in Mathematics and Philosophy from Yale University before studying for a DPhil in International Relations at the University of Oxford.

MARKUS ANDERLJUNG
Head of Policy; Senior Research Fellow

Markus leads GovAI's policy team, aiming to produce rigorous recommendations for governments and AI companies. His research focuses on frontier AI regulation, responsible cutting-edge development, national security implications of AI, and compute governance. He is an Adjunct Fellow at the Center for a New American Security, and a member of the OECD AI Policy Observatory's Expert Group on AI Futures. In 2022 he was seconded to the UK Cabinet Office as a Senior Policy Specialist. He's previously been GovAI's Deputy Director, and has worked as a Senior Consultant at EY Sweden.

ROBERT TRAGER
International Governance Lead

Robert is Director of the Oxford Martin AI Governance Initiative at the University of Oxford and a Senior Research Fellow at Oxford's Blavatnik School of Government. He is a recognised expert in the international governance of emerging technologies, diplomatic practice, institutional design, and technology regulation. He regularly advises government and industry leaders on these topics. Much of his recent research has focused on international governance regimes for AI. Previously, he was Professor of Political Science at the University of California, Los Angeles.

ANTON KORINEK
Economics of AI Lead

Anton is a Professor of Economics at the University of Virginia. He works on the economics of transformative AI and offers a MOOC on the topic. Previously he conducted research on financial crises, and worked in the IT and financial sectors.

## JONAS SCHUETT
Senior Research Fellow

Jonas leads our workstream on risk management. Before joining GovAI, he advised the UK Government on AI regulation, interned at Google DeepMind's Public Policy team, and helped found the Institute for Law and AI (LawAI), where he is still a non-executive board member. He holds a law degree from Heidelberg University and is currently completing his PhD in law at Goethe University Frankfurt.

NOEMI DREKSLER
Senior Research Fellow

Noemi leads GovAI's survey work, which has included surveying AI researchers' views on the ethics, governance, and future of AI; leading economists' views on AI and economic growth; the views of local US policymakers before and after ChatGPT was released; public attitudes towards AI in Europe and the US; and experts' views on AGI safety and governance best practices. She holds a DPhil in Experimental Psychology from Oxford.
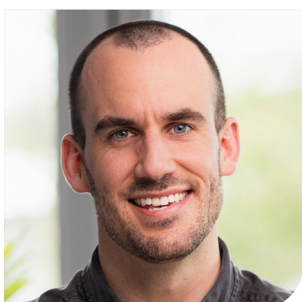
SAM MANNING
Senior Research Fellow

Sam's work focuses on measuring the economic impacts of frontier AI systems and designing policy options to help ensure that advanced AI can foster broadly shared economic prosperity. He previously conducted research at OpenAI and worked on a randomised controlled trial of a guaranteed income programme in the US. Sam has a MSc in International and Development Economics from the University of San Francisco.

LENNART HEIM
Adjunct Fellow

Lennart is a researcher at RAND. His research focuses on the role of compute for advanced AI systems and how compute can be leveraged as an instrument for AI governance, with an emphasis on policy development and security implications. Lennart's publications cover the impacts and governance of advanced AI systems and empirical trends in machine learning, such as compute, data, and AI hardware. He has provided policy and technical advice to several governmental agencies and leading tech companies, including the US and UK governments. Lennart was previously a Senior Research Fellow at GovAI, and remains affiliated as an Adjunct Fellow.

Research Scholars

BEN CLIFFORD
Research Scholar

Ben's work is focused on practical projects to strengthen AI safety and governance. He was project manager for the International Scientific Report on Safety of Advanced AI—a report chaired by Yoshua Bengio—involving thirty countries and presented at the Seoul AI Safety Summit. At GovAI, Ben has authored an explainer on current AI safety techniques and worked with the UK government on the AI Safety Summit and AI Safety Institute. Previously, Ben founded a startup, ran sales teams, and worked in product management. He holds a BA in Philosophy from the University of Warwick.

ALAN CHAN
Research Scholar

Alan's research focuses on agent governance, particularly on infrastructure to manage the interactions of AI agents with the world. He is also interested in model evaluations and bridging AI safety and AI ethics. He has previously done research at the Center on Long-Term Risk and the University of Cambridge. He is a final-year PhD student in machine learning at Mila (Quebec AI Institute).

LEONIE KOESSLER
Research Scholar

Leonie's research focuses on frontier AI regulation and risk management, in particular the EU AI Act and risk thresholds. Before joining GovAI, she worked for the German government. Leonie is also a Research Affiliate at the Institute for Law and AI (LawAI) and holds a Master of Laws (LL.M.) from King's College London and the German First Law Examination (1. Staatsexamen) from her studies at Humboldt-University Berlin.

JIDE ALAGA
Research Scholar

Jide's research focuses on responsible development and deployment policies for frontier AI systems. He was previously at the Center on Long-Term Risk where he conducted research on risk management techniques for preventing AI misuse. Jide earned his BA in Political Science and Philosophy from the University of Toronto.

BEN BUCKNALL
Research Scholar

Ben's research focuses on technical topics with downstream implications for AI policy and governance. His prior work has investigated system access requirements for AI research and evaluation, as well as responsible deployment best practices. Ben holds an MMath in Mathematics from Durham University and an MSc in Computational Science from Uppsala University. He was previously an intern at the University of Cambridge.

CLAIRE DENNIS
Research Scholar

Claire's research focuses on international AI governance regimes and global benefit-sharing. Before joining GovAI, Claire served as a US diplomat and consultant in the United Nations Executive Office of the Secretary-General. She holds a Master in Public Affairs from Princeton University and a BA in International Affairs from George Washington University.

JOHN HALSTEAD
Research Scholar

John's research focuses on AI misuse risks and the relevance of cost-benefit analysis to AI policy design. He has a DPhil in political philosophy from the University of Oxford and previously worked for a TV company, at the Forethought Foundation, and at Founders Pledge.

PETER WILLS
Research Scholar

Peter's research focuses on attributing responsibility in tort law for AI systems. He is a DPhil Candidate in the Faculty of Law of the University of Oxford. Previously, he was a clerk at the Court of Appeal for British Columbia and the Supreme Court of Canada. He holds law degrees from Harvard Law School and the University of Cambridge, in addition to an MEng and a BA from the University of Toronto.

OLIVER RITCHIE
Research Scholar

Oliver's research is focused on UK policy, particularly how the government can make effective AI policy that supports economic growth and better public services while also protecting people from harm or unfair treatment. He has previously worked as a social researcher, helped academics maximise real-world impact, and—at HM Treasury—advised government ministers on topics including tax reform, international negotiations, corruption, decarbonisation, and the COVID-19 response.

Non-Research Staff

GEORG ARNDT
Chief of Staff

Georg supports GovAI's Director in day-to-day decision-making for the organisation, retains a high-level overview of GovAI programmes, and manages GovAI's non-research staff. He has previously worked as an economic consultant for NERA, as Project Manager for the Future of Humanity Institute, University of Oxford, and as Chief Executive of the Future of Humanity Foundation.

RYAN FUGATE
Head of Operations

Ryan leads GovAI's Operations team and is responsible for the efficient day-to-day functioning of the organisation, as well as ensuring that GovAI's operating model continuously and effectively supports its mission. He has previously worked in both the private and non-profit sectors and has experience leading teams focused on strategy, operations, product management, and corporate development.

VALERIE BELU
Research Management Associate

Valerie is GovAI's Research Management Associate. She organises and runs the Seasonal Fellowships. Valerie holds a DPhil in Politics, an MPhil in Comparative Government, and a BA in Philosophy, Politics, and Economics, all from the University of Oxford. Before joining GovAI, she was a Fellow at the LSE's European Institute and a Stipendiary Lecturer at St. Hilda's College, Oxford.

GINA MOSS
Executive Assistant

Gina is executive assistant to GovAI's Director. She manages the Director's time and communication channels, while also supporting him in the prioritisation and organisation of his work. Gina has over 10 years experience working with CEOs, directors, and high-profile individuals in a range of industries from law and finance to film and television.

PAUL HARDING
Operations Manager

Paul is responsible for executing GovAI's priorities and programmes, running the day-to-day operations of the organisation, and supporting our outreach and public relations work. He previously ran an outreach project in Cambridge and worked in both the private and non-profit sectors in China.

AQUILA HASSAN
Operations Associate

Aquila helps ensure the smooth running of GovAI's organisational processes, recruitment, and research programmes. She has a background in engineering and previously worked as a bridge engineer.

ANNA RINGVOLD
Talent Management Associate

Anna is responsible for our outreach, recruitment, and hiring processes. Previously, Anna worked as an IT project manager in the fintech industry.

WILLIAM EHLERS
Operations Associate

William helps ensure that various GovAI processes and projects run smoothly. He has previously interned as a commercial project manager and worked in operations for several US-based nonprofits.

Advisory Board

Our Advisory Board consists of representatives from academia, industry, philanthropy, and policy. With GovAI spinning out of our fiscal sponsor and establishing itself as an independent organisation, we expect to revise our Board in 2024. The new Board will then transition into a more formal role as GovAI's official Board of Directors after the spin-out is completed.

ALLAN DAFOE
President

Allan is the founder and former Director of GovAI. He now serves as Director of Frontier Safety and Governance at Google DeepMind.

HELEN TONER
Advisory Board Member

Helen is Director of Strategy at the Center for Security and Emerging Technology.

TOBY ORD
Advisory Board Member

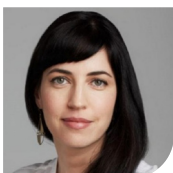Toby is a Senior Researcher at the Oxford Martin AI Governance Initiative.

AJEYA COTRA
Advisory Board Member

Ajeya is a Senior Program Officer at Open Philanthropy.

TASHA MCCAULEY
Advisory Board Member

Tasha is an Adjunct Senior Management Scientist at RAND and co-founded Fellow Robots.