

# Centre for the Governance of AI Annual Report 2022

April 2023

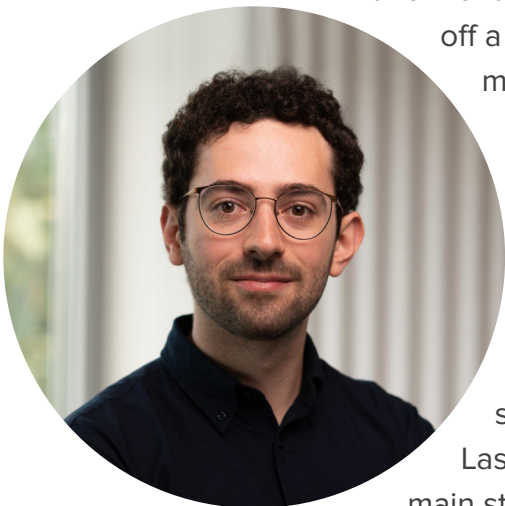
# Contents

<b>A Note From Our Director</b>	<b>3</b>
<b>People</b>	
Advisory Board	10
Core Researchers	12
Research Scholars	16
Non-Research Staff	18
<b>Programmes</b>	
Summer and Winter Fellowships	21
Research Scholar Programme	21
GovAI Policy Programme	21
Policy Team	22
Seminar Series	22
Surveys	22
<b>Research Output</b>	
Highlighted Research	24
Publications, Reports, and Working Papers	25
Opinion Articles	27
Public Advisory Contributions	27
Blog Posts	28
Miscellaneous	28

# A Note From Our Director

2022 was a deeply consequential year for the field of AI governance.

Jurisdictions around the world, most notably the European Union, have come close to finalising the first stringent AI regulations. Technological competition between the US and China has entered a new era, as the US government has adopted the dramatic policy goal of cutting off China's access to advanced chips. The release of the chatbot ChatGPT—which reached 100 million active users within just two months—drew broad public attention and kicked off a race between tech giants worried about missing out on a new market.



Increasingly consequential decisions about AI are being made by governments, companies, and a range of NGOs and international bodies. As progress in AI marches forward, many long-hypothesised risks are also starting to seem less speculative and more concrete.

Last week, when I began writing this note, the main story on the front page of *The New York Times* recounted a journalist's "deeply unsettling" experience with a human-like and emotionally manipulative chatbot being rushed out by a major tech company. The idea that powerful and opaque AI systems could soon change the world—in ways that could be positive, but could also be very damaging—is becoming mainstream.

Unsurprisingly, then, I have never felt more confident in the urgency of our mission. This annual update is intended to give interested readers a clearer view of what the Centre for the Governance of AI ("GovAI") is aiming to do, the progress we are making, and what we hope to accomplish in the coming year. We believe this is an unusually critical period for our work, and we remain immensely grateful to everyone who makes it possible.



## Our Mission

GovAI's mission is to positively shape AI's lasting impact on the world, by helping key institutions make better decisions.

We further this mission in two main ways. First, **we produce helpful research**. Second, **we develop and place AI governance talent**.

Unfortunately, the gap we are working to fill as an organisation has only become more glaring over time. Increasingly consequential decisions are being made: about how to regulate AI, about how to release products that present both benefits and harms, about how to compete responsibly. At the same time, nearly every institution we interact with is very conscious of the challenges it faces in making wise decisions. These institutions often lack the capacity they need to dig into questions that matter. They perceive severe shortages of expertise and talent in critical areas. And these constraints simply are not disappearing as quickly as new decisions and risks are arising. We feel a sense of urgency to do what we can to help.

## How Our Approach to Our Mission Has Evolved

Our approach to our mission has evolved over time, as the world has changed and as we have gained more evidence about our pathways to impact. Here, I will draw attention to two particularly notable changes.

First—although we were previously closer to a traditional academic research centre—we have been gradually deemphasising academic publications in favour of research that is more carefully tailored towards decision-makers. A growing portion of our output has come to take the form of reports, blog posts, unpublished memos, and conversations with relevant stakeholders. We now recognise that these kinds of outputs are often simply more useful for positively influencing decisions. Although our researchers continue to publish academic papers, when it makes sense to do so, this is no longer the default form our outputs take. Our researchers have also become more sensitive to “demand signals” when deciding which projects to embark on. At the same time, we have remained committed to maintaining our financial independence from the institutions we hope to inform.



Second, after noting the successful careers of many of our alumni, we have become increasingly focused on developing and placing AI governance talent.

We have done this mainly by expanding and improving our career development programmes. We believe that there is an increasingly large unmet need for programmes that can help promising people go from being *interested in AI governance* to being *ready to contribute*. As a blunt illustration of this unmet need, the upcoming round of our Summer Fellowship—which brings promising early-career researchers to our office for three months—has received 275 applications for 13 slots. At the same time, though, many important institutions continue to report that they are struggling to find AI governance experts to hire or draw on as advisors. We believe that helping to close this talent gap is currently the most important thing we are doing as an organisation, although we see our talent development efforts as deeply intertwined with our research efforts.

## Our Progress in 2022

### Relaunch

In 2021, GovAI's Founding Director (**Allan Dafoe**) stepped down from his leadership role. I then took over as

Acting Director and led a project to relaunch the organisation outside of the University of Oxford, with plans to expand its activities.

The first part of 2022 was therefore largely devoted to completing this transition. We hired several people to fill critical behind-the-scenes roles, including a Chief of Staff (**Georg Arndt**), a Research Manager (**Emma Bluemke**), and a three-person operations team (**Paul Harding**, **Aquila Hassan**, and **Hunter Muir**). We onboarded these new team members, refined our internal structures and processes, and clarified our direction as an organisation.

### Building Up Our Research Team

Our most significant research priority for the year was to create a policy team, designed to have a particularly strong focus on producing decision-targeted research. Under the leadership of **Markus Anderljung**, the team has established workstreams on AGI lab corporate governance (led by **Jonas Schuett**) and compute governance (led by **Lennart Heim**). Notably, at the start of the year, Markus and Jonas also spent three months in the UK Cabinet Office closely advising the UK's approach to AI regulation.

We also brought **Robert Trager**, a prominent Professor of Political

Science, to Oxford. He has played a central role in developing a workstream on international governance, supervising junior researchers, and deepening our connections with academic experts. We accelerated our survey workstream (led by **Noemi Dreksler**) by bringing on an additional survey researcher (**David McCaffary**). Finally, we launched an experimental workstream on AI governance field-building strategy (led by **Sam Clarke**).

## Building Up Our Career Development Programmes

In 2022, we relaunched our Summer/Winter Fellow Programme after a two-year hiatus. This programme brings cohorts of early-career researchers to Oxford for three months, to do supervised AI governance research, connect with other researchers, and attend relevant events.

To increase the programme's impact, we tripled its cohort size, roughly doubled the volume of supervision we provide, introduced additional events and networking opportunities, and improved the guidance we provide on project selection and career planning. We have also begun to more extensively survey participants and analyse the programme's sources of impact. (See

here for a summary of the initial cohort and the projects they completed. We are currently hosting our second cohort.)

We also soft-launched a new Research Scholar Programme, which supports the careers of promising researchers by offering them one-year visiting positions at GovAI. We accepted our first cohort, which initially consisted of six researchers with a range of backgrounds and career aspirations. We intend to experiment with introducing additional support and structure to increase the programme's impact.

Finally, we began preparations for an experimental Policy Fellowship Programme that is more specifically focused on supporting people who hope to pursue AI policy careers in government.

## Aspirations for 2023

Simply put, our main goal for 2023 is to make our work even more useful.

Last year was an unusually transitional year for us as an organisation. We onboarded a large number of people, launched and relaunched multiple programmes, and redefined ourselves to a significant degree. Although we hope to hire more in 2023—with the aim of expanding our most

promising work streams—our central focus will be on identifying and implementing improvements to our research pipeline and talent programmes.

We want to choose the most useful research projects we can, execute them as successfully as we can, and communicate their results as reliably as we can to the audiences that matter most. We want our talent programmes to do as much as they can to support the development and advancement of the people who pass through them. And we want the under-the-hood aspects of GovAI – our internal processes, role divisions, culture, and financial position – to be as robust and well-tuned as we can make them.

The risks that AI poses for the world are becoming increasingly clear and pressing. We are committed to doing more and more to address them, with each passing year, through continual work to make GovAI the most effective and impact-focused organisation it can be.

## Finances

### Donations, Spending, and Runway

While we received most of our funding to date already in 2021, we want to take this opportunity to thank all our funders, who enable us to achieve our mission.

We are grateful to **Open Philanthropy**; **the Centre for Effective Altruism’s Long-Term Future Fund**; an **independent donor**, through **Effective Giving**; and **Jaan Tallinn**, through **Founders Pledge**, for supporting our work.

For 2022 we additionally would like to thank the **Casey and Family Foundation** for donating \$80,000 and the **Waking Up Foundation** for donating another \$80,000. We are also grateful to everyone who made an **individual contribution through Giving What We Can!** In addition, we were recently able to convert a large individual donation of crypto tokens to dollars. To date, we have received a total of about \$5 million via donations.

We further secured an additional commitment from Open Philanthropy to provide substantial funding over the next two years. At the time of writing, the exact details of the grant are still under consideration.

On the expenditure side, we have incurred costs of \$1.9 million in 2022. Most of those costs were for salaries and contractors (\$1.5 million). Our remaining expenditures were mostly split across travel and events (\$147,000) and overhead (\$233,000). Using our Q4’22 expenditure as an anchor, and not yet accounting for Open Philanthropy’s funding commitment, we currently have a runway of about one year.

## Room for Funding

We have room for additional funding. While we expect to be able to maintain and slightly expand our core staff and programmes over the next two years, we believe we could put significantly more funding to good use. Additional funding could allow us to offer more people spots in our Research Scholars Programme, produce more research by hiring additional Research Fellows (should sufficiently strong candidates be found) and creating a significant budget for research assistants, substantially improve our career development programmes and research pipeline by hiring an additional staff member to oversee them, and ensure that we have room to grow by moving into a larger office. There are also a range of more specific expenditures that we would be keen to make if we had additional funding. In general, funding constraints and associated trade-offs currently play a very active role in our decision-making.

You can donate to us at <https://www.givingwhatwecan.org/charities/govai>. If you may be interested in supporting our work and are interested in developing a more in-depth and detailed understanding of our funding needs, please do not hesitate to contact us at [contact@governance.ai](mailto:contact@governance.ai).

## A Note of Gratitude

I would like to close out this update—as I did last year—with a few thank-you notes.

First, I would like to thank all of our funders for their generous support of our work. We are immensely grateful for all the donations we receive—both large and small—and we aim to never forget our responsibility to use these donations to achieve our mission as best we can. Nothing we do would be possible without our donors.

Second, I would like to thank everyone who joined the GovAI team in the past year. Thank you for believing in our mission, working hard, and helping to make us a better and better organisation with each passing day.

Finally, I would simply like to thank everyone who has decided to focus their career on managing the risks and opportunities posed by AI. I continue to be deeply impressed by the selflessness, industriousness, and kindness of so many people in so many different parts of the AI governance community. I am proud to be part of this community, and I am optimistic that GovAI will find ways to support it for many years to come.



## People

The past year saw GovAI grow substantially. Since the start of the year, we have added five staff in management and operations, five core researchers, and four visiting researchers.



# Advisory Board

Our Advisory Board is comprised of a wide range of experts and provides guidance and oversight for GovAI's activities.



**Allan Dafoe**  
**President**

Allan chairs GovAI's Advisory Board and regularly advises the organisation on matters of strategy. He founded GovAI and now leads DeepMind's Long-term Strategy and Governance team.



**Helen Toner**  
**Advisory Board Member**

Helen is Director of Strategy and Foundational Research Grants at the Center for Security and Emerging Technology (CSET).



**Toby Ord**  
**Advisory Board Member**

Toby is Senior Research Fellow in Philosophy at the University of Oxford.





**Ajeya Cotra**  
**Advisory Board Member**

Ajeya is a Senior Research Analyst at Open Philanthropy.



**Tasha McCauley**  
**Advisory Board Member**

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



# Core Researchers

We currently have a small core research team, including both staff researchers and academic researchers who are heavily integrated into the team despite not being employed by GovAI.



## **Ben Garfinkel** **Acting Director**

Ben leads GovAI and is a Research Fellow at the University of Oxford. He 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.



## **Robert Trager** **International Governance Lead**

Robert Trager is GovAI's International Governance Lead and a Professor of Political Science at the University of California, Los Angeles. He is currently taking a sabbatical with GovAI. He studies the strategic implications of emerging technologies with a particular focus on the transition to advanced AI. He also investigates arms control and nonproliferation; the construction of international orders; and a variety of other topics. Much of his recent research has focused on possible international governance regimes for AI.



## **Anton Korinek**

### **Economics of AI Lead**

Anton is a Professor of Economics at the University of Virginia and a Fellow at the Brookings Institution. His research focuses on the economics of transformative AI. He recently created the first MOOC (massive online open course) on the subject.

---



## **Markus Anderljung**

### **Head of Policy; Research Fellow**

Markus's work aims to identify and improve upon AI governance policy recommendations. He spent 3 months last year seconded to the UK Cabinet Office, advising on the UK's approach to AI regulation. Since his return, he has focussed on setting up GovAI's Policy Team. Markus previously served as Deputy Director of GovAI, when it was based at the University of Oxford.

---



## **Noemi Dreksler**

### **Research Fellow**

Noemi leads GovAI's survey work. Last year, she surveyed economists on their views on long-run economic growth and advanced AI, as well as local US policymakers. She is working on several survey projects, including a large-scale cross-cultural public opinion survey and two new AI researcher surveys. She holds a DPhil in Experimental Psychology from Oxford.

---





## **Jonas Schuett**

### **Research Fellow**

Jonas spent 3 months on secondment to the UK's Cabinet Office last year before joining GovAI full-time as a Research Fellow. His area of expertise is in law and corporate governance. He is developing a research agenda on corporate governance interventions which could help mitigate the risks from advanced AI systems.

---



## **Lennart Heim**

### **Research Fellow**

Lennart joined GovAI's policy team last year as our first Research Scholar. Lennart focuses on compute governance, including the compute supply chain, the role of compute in AI development, hardware security, and technological forecasting. He's also a member of the OECD.AI Expert Group on AI Compute and Climate and helped set up Epoch, a new organisation focussed on researching strategic questions around advanced AI. He has a background in Computer Engineering. Previously Lennart worked as a consultant to the OECD and as a researcher at ETH Zürich.

---



## **Sam Clarke**

### **Strategy Researcher**

Sam researches actionable questions related to AI governance field-building strategy. He previously worked as a researcher at the University of Cambridge's Centre for the Study of Existential Risk and holds an MSc in Computer Science from the University of Oxford.

---



## **David McCaffary**

### **Survey Researcher**

David works as a data scientist and supports GovAI's survey work. His main areas of focus are survey analysis, forecasting, and statistical methods. His background is in computational neuroscience and machine learning, which he studied at the University of Oxford.

---

# Research Scholars

Last year, we hired our first cohort of Research Scholars. These are one-year visiting positions, intended to support the career development of particularly promising AI governance researchers.

## Elizabeth Seger

### Research Scholar

Elizabeth joined GovAI last autumn as part of our first cohort of Research Scholars. She studies responsible model release practices, the aims and effects of efforts to democratise AI, and the implications of AI progress for epistemic security. Elizabeth holds a PhD and MPhil in Philosophy of Science from the University of Cambridge, where she remains a research affiliate with the Centre for the Study of Existential Risk. She holds a BSc in Bioethics from UCLA.

---

## Fynn Heide

### Research Scholar

Fynn studies technical AI progress and AI policy in the People's Republic of China. He currently focuses on advances at the country's frontier AI research institutions, their emerging approaches to AI safety, and their relationships with the Chinese Party-State. He previously researched Chinese data and technology policy at Sinolytics, Trivium China, and the Mercator Institute for China Studies. Fynn holds a BA in Politics, Philosophy, and Economics from the University of Warwick.

---

## **Eoghan Stafford**

### **Research Scholar**

Eoghan's research explores the impact of AI on autocracy, as well as risks from technological competition between great powers. Prior to GovAI, he was a research fellow at the Harvard Kennedy School's Middle East Initiative. He holds a PhD in Political Science from UCLA, an MSc in Political Theory from LSE, and an AB in Social Studies from Harvard.

---

## **Guive Assadi**

### **Research Scholar**

Guive was part of our Summer Fellow cohort last year and joined GovAI as a Research Scholar after the completion of the Fellowship. He currently works on value erosion due to competitive dynamics in AI development. Previously, he researched the history of nuclear espionage and its potential lessons for preventing AI espionage. Guive holds a master's in history from Cambridge University and a bachelor's from UC Berkeley.

---

# Non-Research Staff

## **Georg Arndt** **Chief of Staff**

Georg supports Ben in day-to-day decision-making for the organisation, retains a high-level overview of GovAI programmes, and manages GovAI's non-research staff.

---

## **Anne le Roux** **Head of Partnerships**

Anne transitioned from Head of Operations to Head of Partnerships last July. She leads collaborations with other organisations working on long-term AI governance and safety.

---

## **Emma Bluemke** **Research Manager**

Emma is responsible for tracking ongoing research projects and ensuring that researchers can complete and communicate their research. She is also responsible for GovAI's Summer and Winter Fellowship Programme.

---

## **Gina Moss** **Executive Assistant**

Gina is Ben's executive assistant. She manages Ben's time and communication channels, while also supporting Ben in the prioritisation and organisation of his work. In addition, Gina takes on small administrative tasks for other team members.

---



## **Paul Harding**

### **Operations Manager**

Paul is responsible for executing GovAI's priorities and programmes and running the day-to-day operations of the organisation. He manages our two Operations Associates, Aquila and Hunter.

---

## **Hunter Muir**

### **Operations Associate**

Hunter helps ensure the smooth running of GovAI's communications, recruitment, and HR processes.

---

## **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.

---

## Programmes

Since spinning out from the University of Oxford in November of 2021, GovAI has established or expanded a range of programmes that help us to achieve our mission.



## Summer and Winter Fellowships

Twice a year, this programme brings a cohort of aspiring or early-career AI governance researchers to Oxford for three months. They work on a research project, receive substantial supervision and guidance, and participate in Q&A sessions with established experts in the space. We aim for everyone to leave the programme having produced a significant research sample. You can read more about the programme [here](#).

We successfully completed our first Summer Fellowship with 12 Fellows from June to August 2022. For a brief overview of what our 2022 Summer Fellows worked on, please see [here](#). GovAI gratefully thanks all of the supervisors for dedicating their time to developing the next generation of researchers.

## Research Scholar Programme

This newly-launched programme offers one-year visiting positions to promising AI governance researchers and policy practitioners. Participants are given both mentorship and significant research freedom. The programme is primarily intended to support the career development of people who are relatively new to AI governance and are hoping to learn,

skill up, make connections, build their profiles, and clarify their plans and directions. For some Summer/Winter Fellows, applying to the Research Scholar Programme may also offer a valuable opportunity to stay at GovAI past the end of their fellowship.

We launched the programme in 2022 and are currently hosting four Research Scholars. We are likely to take on several more Research Scholars this year. We also plan to experiment with adding further resources and structure to increase the programme's impact on the careers of participants.

## GovAI Policy Programme

This eight-week, part-time programme will help current and future U.S. policy practitioners learn about AI, its risks, and policy tools for mitigating these risks. The programme is structured around guided self-study, workshops, and seminars. It aims to give participants legible expertise on topics that are currently central to AI policy, while also making them aware of risks and issues that may grow in prominence over time.

We began preparation for this course in 2022 and plan to trial it this year. We will decide whether to continue, expand, or drop the programme on the basis of the initial trial.

## Policy Team

We have organised a portion of our research staff into a policy team. The team collaborates to develop, improve, and support the implementation of proposals for what influential institutions can do in the next few years to prepare the world for advanced AI capabilities. The team is led by Markus Anderljung, GovAI's Head of Policy. It currently hosts a workstream on corporate governance, led by Jonas Schuett; a workstream on compute governance, led by Lennart Heim; and a workstream on AI regulation and the external scrutiny of large models, led by Markus.

We formed this team in 2022, in an effort to shift GovAI's research portfolio in a more applied direction and pursue new advising opportunities. One of the team's significant activities in 2022 was to advise the UK government on its approach to regulation; during this project, Markus and Jonas were seconded to the Cabinet Office for three months.

## Seminar Series

In 2022, GovAI began hosting a series of public seminars featuring leading AI governance experts. Going forward, in 2023, this series will likely increase in frequency to monthly events, which will be distributed to the public online. You can view past events on our website [here](#).

## Surveys

GovAI conducts surveys of public and expert opinion, with the aim of informing and positively influencing AI governance discussions. Ongoing projects include an annual survey of the beliefs of AI researchers on a range of technical and ethical questions, a survey of leading economists' views on AI and economic growth, a survey of public opinion across several countries, and a survey of AI researchers' views on AI sentience. Noemi Dreksler leads our survey workstream.

In late 2022, we took on David McCaffary as a survey researcher to increase our ability to produce and share results quickly. David may also help to build a platform that will make it easier for our researchers to explore and visualise survey findings.

## Research Output

In 2022, GovAI researchers authored or contributed to over 30 public research outputs.



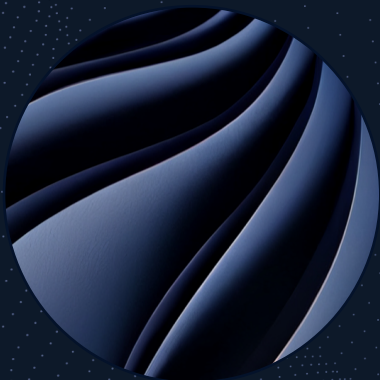


## Highlighted Research



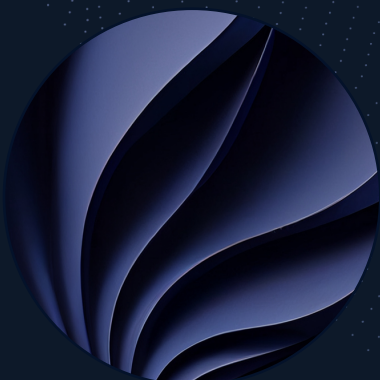
### **The Artefacts of Intelligence: Governing Scientists' Contribution to AI Proliferation** **Toby Shevlane**

Within the AI community, there is substantial debate about how responsible AI developers should share systems that can bring both benefits and harms. This dissertation clarifies the debate by drawing on interviews and case studies, particularly OpenAI's high-profile release of GPT-2 and GPT-3. It analyses structured access as a new approach to sharing AI systems.



### **The Brussels Effect and Artificial Intelligence** **Markus Anderljung and Charlotte Siegmann**

The European Union has a track record of developing regulations that affect how technologies are developed and used even outside its jurisdiction (producing a so-called "Brussels effect"). This report argues that forthcoming EU AI regulations could have similarly consequential effects abroad.



### **Forecasting AI Progress: Evidence from a Survey of Machine Learning Researchers** **Baobao Zhang, Noemi Dreksler, Markus Anderljung, Lauren Kahn, Charlie Giattino, Allan Dafoe, Michael C. Horowitz**

Forecasts of AI development could help improve policy- and decision-making in areas ranging from transportation, health care, science, finance, to the military. We conducted a survey of AI researchers to examine their attitudes on their beliefs about progress in AI. The AI researchers surveyed placed a 50% likelihood of human-level machine intelligence being achieved by 2060, and exhibited significant optimism about how human-level machine intelligence will impact society.

## **Publications, Reports, and Working Papers**

### **The Impact of Artificial Intelligence: A Historical Perspective**

Ben Garfinkel

---

### **Lessons from the Development of the Atomic Bomb**

Toby Ord

---

### **Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning**

Pablo Villalobos, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn, and Anson Ho

---

### **Emerging Technologies, Prestige Motivations and the Dynamics of International Competition**

Joslyn Barnhart

---

### **Safety Not Guaranteed: International Strategic Dynamics of Risky Technology Races**

Robert Trager, Eoghan Stafford, and Allan Dafoe

---

### **The Security Governance Challenge of Emerging Technologies**

Robert Trager

---

### **AI Challenges for Society and Ethics**

Jess Whittlestone and Sam Clarke

---

### **Structures Access: An Emerging Paradigm for Safe AI Development**

Toby Shevlane

---

### **Three Lines of Defense Against Risks From AI**

Jonas Schuett

---

---

## **Why We Need a New Agency to Regulate Advanced Artificial Intelligence: Lessons on AI Control from the Facebook Files**

Anton Korinek

---

## **Aligned With Whom? Direct and Social Goals for AI Systems**

Anton Korinek and Avital Balwit

---

## **Robert Jervis and the Social Dilemmas of Technological Innovation**

Robert Trager

---

## **Information Hazards in Races for Advanced Artificial Intelligence**

Nicholas Emery, Andrew Park, and Robert Trager

---

## **Red-Teaming the Stable Diffusion Safety Filter**

Javier Rando, Daniel Paleka, David Lindner, Lennart Heim, and Florian Tramèr

---

## **The IAEA Solution: Knowledge Sharing to Prevent Dangerous Technology Races**

Robert Trager and Eoghan Stafford

---

## **AI Ethics Statements: Analysis and Lessons Learnt From NeurIPS Broader Impact Statements**

Carolyn Ashurst, Emmie Hine, Paul Sedille, and Alexis Carlier

---

## **Machine Learning Model Sizes and the Parameter Gap**

Pablo Villalobos, Jaime Sevilla, Tamay Besiroglu, Lennart Heim, Anson Ho, and Marius Hobbhahn

---

## **Preparing for the (Non-Existent?) Future of Work**

Anton Korinek and Megan Juelfs

---

## Opinion Articles

### **Lethal Autonomous Weapons Need to be Regulated - But Not the Way Advocates Say**

Robert Trager

---

### **@Elonmusk and @twitter: The Problem With Social Media is Misaligned Recommendation Systems, Not Free Speech**

Justin Bullock and Anton Korinek

---

### **Exploring Epistemic Security: The Catastrophic Risk of Epistemic Insecurity in a Technologically Advanced World**

Elizabeth Seger

---

## Public Advisory Contributions

### **Submission to the Request for Information (RFI) on Implementing Initial Findings and Recommendations of the NAIRR Task Force**

Lennart Heim and Markus Anderljung

---

### **GovAI Response to the Future of Compute Review - Call for Evidence**

Lennart Heim and Markus Anderljung

---

### **Submission to the NIST AI Risk Management Framework**

Jonas Schuett and Markus Anderljung

---

## Blog Posts

### **Compute Funds and Pre-trained Models: The US National AI Research Resource Should Provide Structured Access to Models, Not Just Data and Compute**

Markus Anderljung, Lennart Heim, and Toby Shevlane

---

### **Safety-Performance Tradeoff Model Web App**

Robert Trager, Nicholas Emery, Eoghan Stafford, Paolo Bova, and Allan Dafoe

---

### **How Technical Safety Standards Could Promote TAI Safety**

Cullen O’Keefe, Jade Leung, and Markus Anderljung

---

### **Sharing Powerful AI Models: Structured Access**

Toby Shevlane

---

## Miscellaneous

### **Deliberating Autonomous Weapons**

Robert Trager

---

### **‘Economics of AI’ Open Online Course**

Anton Korinek

---

### **Technological Progress and Artificial Intelligence**

Anton Korinek, Martin Schindler, and Joseph E Stiglitz

---





© 2023 CENTRE FOR THE GOVERNANCE OF AI  
VISIT US AT [GOVERNANCE.AI](https://www.governance.ai)