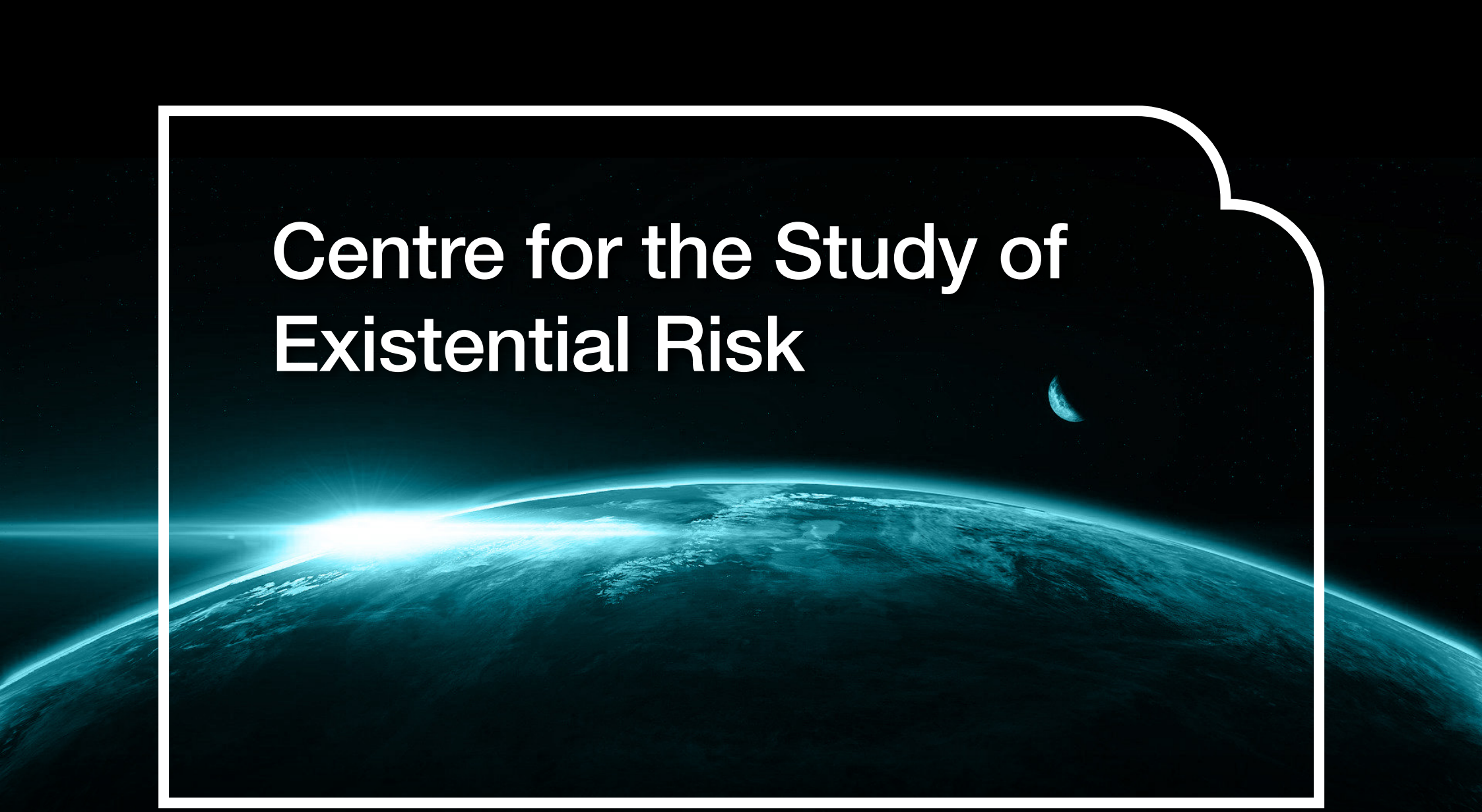
Centre for the Study of Existential Risk



UNIVERSITY OF CAMBRIDGE

A report prepared for CSER supporters

MARCH 2022

The University of Cambridge extends its sincere thanks for your support of the activities of the Centre for the Study of Existential Risk (CSER).

Supported by your generosity, the work of CSER researchers is increasing our understanding of, and preparedness for, existential threats to our world.

Contents

Navigation

Scroll through the document, or click on the relevant section in the table of contents to go directly to that section. To return to the contents list, click the page number at the bottom of the page.

Click to return to contents

Reuben Scholars at Cambridge December 2021

An introduction from **Seán Ó hÉigeartaigh**

Seán Ó hÉigeartaigh

Executive Director, Centre for the Study of Existential Risk

The Centre for the Study of Existential Risk (CSER) is an interdisciplinary research centre within the University of Cambridge dedicated to the study and mitigation of risks that could lead to civilizational collapse or human extinction. We are developing a science of global risk that we apply to Global Catastrophic Biological Risks, Extreme Risks and the Global Environment, and Risks from Artificial Intelligence. Our work is shaped around three main goals:

- Understanding: we study existential and global catastrophic risk.
- Impact: we develop collaborative strategies to reduce these risks.
- Field-building: we foster a global community of academics, technologists and policy-makers who share our goals.

Our last report covered May–September 2021. This report covers the period October–December 2021 and outlines our activities and future plans. Highlights of the last three months include:

- **We published ten papers – including one in *Science***, on future foods, geoengineering, representation of indigenous peoples in international law, AI and competition law, the impact of COVID-19 on research, generality in AI, trustworthy AI development, zoonosis, urban resilience, and the transformative potential of AI.
- We presented and participated in numerous events. A particular highlight being a **summit of universal owners** (large institutional investors) co-hosted with Jesus College ahead of COP26, which saw investor managers with responsibility for billions of dollars engaging with global catastrophic risks and how to reduce them.
- **Five policy reports**, including a horizon scan of issues in dual use research of concern co-produced with the World Health Organization, and the Cambridge Principles for systemic stewardship by universal owners coming out of our summit. CSER's work was also prominently featured in the UK House of Lords' "Preparing for Extreme Risks: Building a Resilient Society" and in the UN Disaster Risk Reduction's report "A Framework for Global Science In Support of Risk-informed Sustainable Development and Planetary Health".
- We welcomed a second Academic Programme Manager, **Paul Ingram**, to help coordinate our increasingly ambitious and diverse research programmes and build a stronger future for the Centre.
- Our researchers have appeared on podcasts highlighting their research to a **wide and engaged audience**, including Russell Brand's *Under the Skin*, the BBC's *Naked Reflections*, and *21st Talks*, while our work has been featured in the *Financial Times* and BBC Future.

1. Update

While the temporary relaxation of COVID-19 regulations allowed for a partial return to office working for some during the autumn, we continue to operate hybrid working for most of our staff members while some remain fully remote in line with the Government and University guidance. We are fortunate that most of our work can be done remotely. The return to office working has not only helped to build new connections, especially with staff members who joined CSER during the pandemic, but has also allowed us to both host and participate in a greater number and wider range of events. In November and December we hosted two away days, which provided a much needed opportunity to host Centre-wide discussions and brain storming sessions about the Centre and our future.

2. People

2.1 New Staff

Over the last few months we have welcomed a new Academic Programme Manager.

Paul Ingram has several decades experience leading diverse and multicultural teams to impact decisions on existential threats, particularly nuclear war. He was the Executive Director of the transatlantic British American Security Information Council (BASIC) 2007–19, focusing on nuclear deterrence and disarmament issues in the US, Europe, the Middle East and Asia. Since 2019, he has worked closely with the Swedish Foreign Ministry crafting the Stepping Stones Approach. The associated 16-nation Stockholm Initiative for Nuclear Disarmament has become a widely-acknowledged glimmer of hope for the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) Review process. Paul recently talked more about his work for our meet the researcher series.

2.2 Visiting Scholars

We have welcomed two new visitors to CSER during this period.

Cecil Abungu (September 2021–February 2022). Cecil works on one project touching on how AI could lead to extreme inequality and power concentration and another connected to mapping adaptive governance regimes for changes in AI capabilities. He has a background in law, including as an Open Philanthropy research grantee and a research fellow at the Legal Priorities Project. He holds an undergraduate law degree from Strathmore Law School in Nairobi and a Master's in Law degree from Harvard Law School.

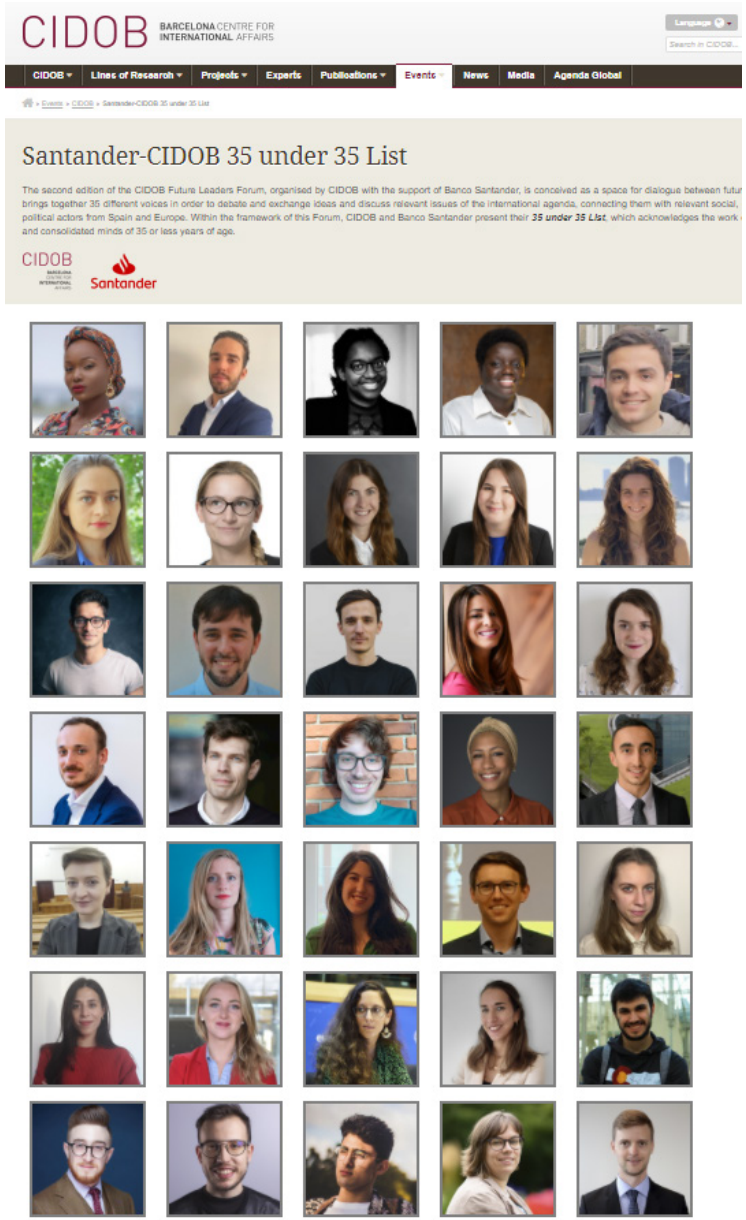
Tegan Maharaj (October 2021–October 2022): Tegan's research seeks to understand AI systems in realistic settings, in order to ensure their responsible use. At CSER, Tegan examines the use of AI in forecasting and mitigating risk. Tegan is an Assistant Professor in the Faculty of Information at the University of Toronto, where she is an affiliate of the Schwartz-Reisman and Vector institutes. Her background is in deep representation learning, as a PhD student at Mila (the Quebec AI Institute) and Ecole Polytechnique in Montreal.

2.3 Researcher Updates

Matthijs Maas has been selected as a new member of the Santander-CIDOB 35 under 35 List. The list acknowledges the work of 35 young people with the potential to become future leaders in their different areas of expertise and activity.

Lara Mani has been appointed as the Outreach and Engagement representative on the committee of The Volcanic and Magmatic Studies Group. The Group aims to facilitate discussion, in an inclusive environment, amongst researchers with interests in volcanology, igneous petrology, geochemistry and allied fields.

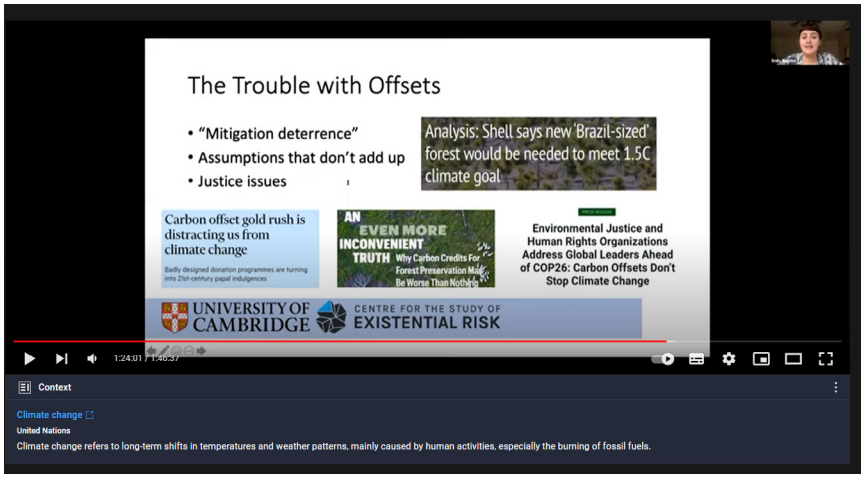
SJ Beard and Clarissa Rios Rojas have been awarded the 2021 The Borysiewicz Interdisciplinary Fellowship. The Fellowship is awarded annually to ten exceptional researchers working at the University of Cambridge and consists of a programme with a unique mixture of purposeful and outcome-driven engagements for developing leaders that focus on global challenges.



3. Events, Engagement and Outreach

3.1 Academic Engagement

- 7 September: Matthijs Maas spoke at the **European Society for International Law** Interest Group on International Law and Technology workshop about the effects of AI systems and legal automation on the coherence of international law.
- 7 September: long-time CSER visitor Emily Bugden took part in a panel on Toward a Sociology of Climate Risk hosted by the London School of Economics.
- 17 September: Lara Mani presented at the discussion meeting for the 2020–2021 eruption of La Soufrière a talk on “Evaluating the crisis communications campaign during the La Soufriere eruption”.
- 23 September: Clarissa Rios Rojas took part in a panel on Strategy Foresight for the Future Normal of Young Scientists at the Tsukuba Conference.
- 29 September: Luke Kemp co-curated a panel on "Extreme Climate Risks: What Are the Worst-Case Scenarios" at the Climate Risk Summit – a COP26 Univerities Network Virtual Conference. Ellen Quigley appeared as one of the panellists.
- 1 October: Ellen Quigley appeared on the panel Deep Dive 7: Canadian banks and insurers: climate risk stress testing and scenario analysis hosted by Responsible Investor.
- 15 October: Luke Kemp organized a panel on Extreme Climate Risks: What Are the Worst-Case Scenarios at the Cambridge Zero Climate Change Festival 2021. While Natalie Jones spoke on panels "Can leaders turn Paris ambition into Glasgow action at COP26? Cambridge Talks" and "Cambridge Zero Research Symposium: Law & Policy" at the same event.
- 3 November: Emily Bugden gave a talk on “benchmarking the benchmarkers” at the Cambridge Zero research symposium (pictured below).
- 4 December: Sean O hEigeartaigh co-organised a workshop on “Cross-cultural cooperation in AI” at Tsinghua University.



3.2 Policy Engagement

- 31 September: Ellen Quigley and Akaraseth Puranasamriddi contributed to the **World Economic Forum white paper** Reshaping Risk Mitigation: The Impact of Non-financial Levers.
- 12 October: Clarissa Rios Rojas spoke at the Organisation for Economic Co-operation and Development's OECD Government Foresight Community Annual Meeting 2021 providing a Keynote on Longtermism as a Global Policy Priority.
- 18 October: CSER researchers submitted evidence to the UK Government's National Resilience Strategy.
- 19 October: Ellen Quigley appeared on a panel on "The role of finance in addressing climate change" organised by the Centre for Competition Policy.
- 11 November: Partha Dasgupta, Chair of the CSER Management Board, was on a panel with Nick Stern at COP26 in Glasgow, titled Putting the Environment at the Heart of Economic Decision-Making: Dasgupta and Stern in Conversation.
- 23 November: Paul Ingram moderated the Middle East Treaty Organization-Geneva Centre for Security Policy Panel Discussion: Achieving Success and the Path Forward – UN Conference on a WMD-Free Zone in the Middle East.
- 3 December: The House of Lords' Select Committee report on Risk Assessment and Risk Planning was released, and **cited written and oral evidence** from Centre for the Study of Existential Risk scholars.
- 14 December: The United Nations Disaster Risk Reduction's report "A Framework for Global Science In Support of Risk-informed Sustainable Development and Planetary Health", which Clarissa Rios Rojas contributed to as an expert, prominently featured existential risks and CSER's work.
- 23 December: CSER Academic Programme Manager, Paul Ingram, published a report for BASIC on achieving progress on disarmament, drawing on his in-depth experience in nuclear disarmament.

3.3 Public Engagement

- 9 September: Monash University hosted a public lecture by CSER co-founder and Academic Director, Professor Huw Price, titled The Future of Artificial Intelligence: Academia's Role in Getting it Right.
- 9 September: The Geneva Science Policy Interface reported on our event at the Biological Weapons Convention about Decision-making in the face of global catastrophic biological risks.
- 18 September: SJ Beard appeared on the *21st Talks* podcast to talk about the History of Existential Risk.
- 2 October: Freya Jephcott appeared on the *Naked Reflections* podcast episode Wet Markets and Leaky Labs: Covid-19 origins.
- 12 October: Freya Jephcott and Charlotte Hammer gave a talk on Surveillance Systems and Epidemiological Obfuscation hosted by the Cambridge University Centre for Research in the Arts, Social Sciences, and Humanities.
- 13 October: Lara Mani co-authored an article on the *Effective Altruism Forum* - On the assessment of volcanic eruptions as global catastrophic or existential risks.
- 28 October: Luke Kemp wrote about who is causing planetary destruction and why for BBC Future.
- 28 October: Matthijs Maas was interviewed by Zetland about the use of DeepFake AI voice systems in financial crime.
- 29 October: Luke Kemp was a guest of Russell Brand's podcast *Under the Skin* to talk about the use and abuse of Emergency Powers.
- 11 November: Partha Dasgupta, Chair of the CSER Management Board, was interviewed by the ***Financial Times*** about the economics of biodiversity.
- 1 December: CSER's research was featured in the first of Stewart Russel's BBC *Reith Lectures* The Biggest Event in Human History, after which Paul Ingram was invited to ask a question from the floor.
- 4 December: Luke Kemp and Catherine Rhodes appeared on the *21st Talks* podcast to talk about Governing Global Risks.
- 7 December: Luke Kemp interviewed Dr Andrew Leigh MP on his latest book, *What's the Worst That Could Happen?: Existential Risk and Extreme Politics* about the greatest challenges facing humanity and the role of politics in responding to them in an event organised by The Commission for the Human Future.
- 9 December: CSER's research on developing trustworthy AI was featured in an article on *Gizmodo* – Hackers Could Make Dangerous AI Safer.
- 10 December: Paul Ingram was featured in a Press TV broadcast Iran: Talks to proceed faster if West display initiative.

3.4 Events

- 6 September: Lalitha Sundaram supported The Simon Institute for Longterm Governance in hosting an in-person lunch session on the topic of long-term decision-making in the face of extreme biological risks, complementing the **Biological Weapons Convention** Meeting of Experts in Geneva, Switzerland.
- 7–10 September: Alex McLaughlin co-organised a workshop at the MANCEPT workshops on Political Resistance and Contemporary Injustice, at which he presented a paper on Climate Resistance and the Far Future.

- 21 and 28 October: Clarissa Rios Rojas and the Simon Institute for Longterm Governance hosted two workshops to co-create a science-policy agenda on Global Catastrophic Risks.
- 26 to 28 October: Ellen Quigley hosted a **Universal Owners Summit** ahead of COP26.
- 8 to 26 November: Lara Mani and Tom Hobson organised a ParEvo participatory foresight activity with experts in the field of Biosecurity and Biotechnology.
- 15 December: CSER and the University of Tokyo hosted a joint panel titled Resilience and Global Risks as part of the UTokyo-Cambridge Voices series.

4. Publications

4.1 Papers

"Future Foods for Urban Food Production" in *The Palgrave Encyclopedia of Urban and Regional Futures* 4 October 2021 by Asaf Tzachor, Catherine Richards.

Urban food security, a global concern for over four billion city dwellers, currently relies on traditional staple foods, underpinned by conventional systems of food production and provision. These systems and supply chains are vulnerable to a litany of biotic and abiotic risks, and thereby to yield failures. Future foods, including microalgae, macroalgae, bivalve mollusks, mycoprotein, insect larvae, and cultured meat, may provide nutritious and sustainable alternatives to customary food sources. Furthermore, against the backdrop of risks to conventional food systems, future foods may be cultivated in state-of-the-art, closed-environment configurations that mitigate exposure to external hazards. Such configurations provide a risk-resilient supply of safe and nutritious foods through a decentralized and modular architecture of discrete production units. Additionally, these novel farming systems may be integrated into urban landscapes and mixed-use buildings, allowing localization of food supply chains, consequently facilitating the rise of compact cities and prosumer innovations in food cultivation techniques and end use products, as well as realizing circular economy co-benefits, such as waste recycling. As advanced controlled-environment agriculture technologies, future foods production systems are well suited for smart city and Agriculture 4.0 applications. Mainly, sensor and automation technologies may be used to optimize internal physical, chemical, and biological cultivation processes and parameters. Big data infrastructure, such as the Internet of Things, participatory sensing, cloud computing, and data mining, can take full advantage of consumer insights and connectivity for efficient management of urban food systems of the future.

A Fate Worse Than Warming? Stratospheric Aerosol Injection and Global Catastrophic Risk in Frontiers in *Climate* 19 November 2021 by Aaron Tang, Luke Kemp.

Injecting particles into atmosphere to reflect sunlight, stratospheric aerosol injection (SAI), represents a potential technological solution to the threat of climate change. But could the cure be worse than the disease? Understanding low probability, yet plausible, high-impact cases is critical to prudent climate risk management and SAI deliberation. But analyses of such high impact outcomes are lacking in SAI research. This paper helps resolve this gap by investigating SAI's contributions to global catastrophic risk. We split SAI's contributions to catastrophic risk into four interrelated dimensions:

1. Acting as a direct catastrophic risk through potentially unforeseen ecological blowback.
2. Interacting with other globally catastrophic hazards like nuclear war.
3. Exacerbating systemic risk (risks that cascade and amplify across different systems);

4. Acting as a latent risk (risk that is dormant but can later be triggered).

The potential for major unforeseen environmental consequences seems highly unlikely but is ultimately unknown. SAI plausibly interacts with other catastrophic calamities, most notably by potentially exacerbating the impacts of nuclear war or an extreme space weather event. SAI could contribute to systemic risk by introducing stressors into critical systems such as agriculture. SAI's systemic stressors, and risks of systemic cascades and synchronous failures, are highly understudied. SAI deployment more tightly couples different ecological, economic, and political systems. This creates a precarious condition of latent risk, the largest cause for concern. Thicker SAI masking extreme warming could create a planetary Sword of Damocles. That is, if SAI were removed but underlying greenhouse gas concentrations not reduced, there would be extreme warming in a very short timeframe. Sufficiently large global shocks could force SAI termination and trigger SAI's latent risk, compounding disasters and catastrophic risks. Across all these dimensions, the specific SAI deployment, and associated governance, is critical. A well-coordinated use of a small amount of SAI would incur negligible risks, but this is an optimistic scenario. Conversely, larger use of SAI used in an uncoordinated manner poses many potential dangers. We cannot equivocally determine whether SAI will be worse than warming. For now, a heavy reliance on SAI seems an imprudent policy response.

Self-Determination and the Right of Peoples to Participate in International Law-Making in *The British Yearbook of International Law* 23 November 2021 by Natalie Jones

In recent decades, peoples have frequently asserted a right to participate in the international legal order, and have participated in various international law-making and regulatory processes carried out under the auspices of intergovernmental organizations. However, while self-determination has long been understood to encompass a right to participation at the national level, the case for this legal right at the international level has not been comprehensively discussed. I present an account of the development of the law of self-determination and argue that by its logic, the law of self-determination can be interpreted as justifying a collective right of peoples to participate in international law-making. The article proposes a formulation of the right including its corresponding duties. The article's key contribution is to offer an understanding of how practice by states and international organizations to grant indigenous peoples enhanced status in intergovernmental fora may be seen to logically follow from the law of self-determination.

AI & Antitrust: Reconciling Tensions Between Competition Law and Cooperative AI Development in the *Yale Journal of Law and Technology* 23 November 2021 by Shin-Shin Hua, Haydn Belfield

Cooperation between companies developing artificial intelligence (AI) can help them create AI systems that are safe, secure, and with broadly shared benefits. Researchers have proposed a range of cooperation strategies, ranging from redistributing "windfall" profits to assistance to address the harmful dynamics of a competitive race for technological superiority. A critical tension arises, however, between cooperation and the goal of competition law, which is to protect the very process of competition between rival companies. Whilst these potential conflicts are significant, they are currently underexplored in the literature. This paper examines the relationship between proposed forms of AI cooperation and competition law, focusing on the competition law of the European Union (EU). EU competition law governs the behavior of the world's largest AI companies, though many are based abroad, especially in the US. Its jurisdiction can extend to any foreign company that is active in the EU. Scrutiny of US "Big Tech" is also an area of strategic focus for the European Commission (EC). This paper seeks to reconcile the cooperative AI development and competition law. It examines fourteen forms of AI cooperation, both those that are applicable today and

longer-term strategies that will apply when AI development is more advanced. Where we identify potential tensions with EU competition law, we suggest mitigation steps. Our aim is to ensure the long-term sustainability of these important safeguards to the responsible and beneficial development of AI.

General intelligence disentangled via a generality metric for natural and artificial intelligence in *Nature Scientific Reports* 11(1) by José Hernández-Orallo, Bao Sheng Loe, Lucy Cheke, Fernando Martínez-Plumed, Seán Ó hÉigeartaigh

Success in all sorts of situations is the most classical interpretation of general intelligence. Under limited resources, however, the capability of an agent must necessarily be limited too, and generality needs to be understood as comprehensive performance up to a level of difficulty. The degree of generality then refers to the way an agent's capability is distributed as a function of task difficulty. This dissects the notion of general intelligence into two non-populational measures, generality and capability, which we apply to individuals and groups of humans, other animals and AI systems, on several cognitive and perceptual tests. Our results indicate that generality and capability can decouple at the individual level: very specialised agents can show high capability and vice versa. The metrics also decouple at the population level, and we rarely see diminishing returns in generality for those groups of high capability. We relate the individual measure of generality to traditional notions of general intelligence and cognitive efficiency in humans, collectives, non-human animals and machines. The choice of the difficulty function now plays a prominent role in this new conception of generality, which brings a quantitative tool for shedding light on long-standing questions about the evolution of general intelligence and the evaluation of progress in Artificial General Intelligence.

"Mitigating losses: how scientific organisations can help address the impact of the COVID-19 pandemic on early-career researchers" in *Humanities and Social Sciences Communications*, 8(1) 24 November by Sandra López-Vergès, Bernardo Urbani, David Fernández Rivas, Sandeep Kaur-Ghumaan, Anna K. Coussens, Felix Moronta-Barrios, Suraj Bhattarai, Leila Niamir, Velia Siciliano, Andreea Molnar, Amanda Weltman, Meghnath Dhimal, Shalini S. Arya, Karen J. Cloete, Almas Taj Awan, Stefan Kohler, Chandra Shekhar Sharma, Clarissa Rios Rojas, Yoko Shimpuku, John Ganle, Maryam M. Matin, Justine G. Nzweundji, Abdeslam Badre, Paulina Carmona-Mora

Scientific collaborations among nations to address common problems and to build international partnerships as part of science diplomacy is a well-established notion. The international flow of people and ideas has played an important role in the advancement of the 'Sciences' and the current pandemic scenario has drawn attention towards the genuine need for a stronger role of science diplomacy, science advice and science communication. In dealing with the COVID-19 pandemic, visible interactions across science, policy, science communication to the public and diplomacy worldwide have promptly emerged. These interactions have benefited primarily the disciplines of knowledge that are directly informing the pandemic response, while other scientific fields have been relegated. The effects of the COVID-19 pandemic on scientists of all disciplines and from all world regions are discussed here, with a focus on early-career researchers (ECRs), as a vulnerable population in the research system. Young academies and ECR-driven organisations could suggest ECR-powered solutions and actions that could have the potential to mitigate these effects on ECRs working on disciplines not related to the pandemic response. In relation with governments and other scientific organisations, they can have an impact on strengthening and creating fairer scientific systems for ECRs at the national, regional, and global level.

Filling gaps in trustworthy development of AI in *Science* 10 December 2021 by Shahar Avin, Haydn Belfield, Miles Brundage, Gretchen Krueger, Jasmine Wang, Adrian Weller, Markus Anderljung, Igor Krawczuk, David Krueger, Jonathan Lebensold, Tegan Maharaj, Noa Zilberman

The range of application of artificial intelligence (AI) is vast, as is the potential for harm. Growing awareness of potential risks from AI systems has spurred action to address those risks while eroding confidence in AI systems and the organizations that develop them. A 2019 study (1) found more than 80 organizations that have published and adopted "AI ethics principles," and more have joined since. But the principles often leave a gap between the "what" and the "how" of trustworthy AI development. Such gaps have enabled questionable or ethically dubious behavior, which casts doubts on the trustworthiness of specific organizations, and the field more broadly. There is thus an urgent need for concrete methods that both enable AI developers to prevent harm and allow them to demonstrate their trustworthiness through verifiable behavior. Below, we explore mechanisms [drawn from (2)] for creating an ecosystem where AI developers can earn trust—if they are trustworthy (see the figure). Better assessment of developer trustworthiness could inform user choice, employee actions, investment decisions, legal recourse, and emerging governance regimes.

Disaster Displacement and Zoonotic Disease Dynamics: The Impact of Structural and Chronic Drivers in Sindh, Pakistan in PLOS Global Public Health 13 December by Dorien H. Braam, Rafiq Chandio, Freya Jephcott, Alex Tasker, James L. N. Wood

Pathways for infectious disease transmission including zoonoses, diseases transmitted between animals and humans, are complex and non-linear. While forced migration is considered an important driver for the spread of zoonoses, actual disease dynamics remain under researched. This paper presents the findings of a case study investigating how disaster displacement affected zoonotic disease transmission risk following the 2010 'superfloods' in Sindh province, Pakistan. We interviewed 30 key informants and 17 household members across 6 rural communities between March and November 2019, supported by observational studies and a review of secondary data. Results were analysed using the ecosocial theoretical framework. Buffalo, cattle and goats were often the only moveable asset, therefore livestock was an important consideration in determining displacement modality and destination location, and crowded locations were avoided to protect human and animal health. Meanwhile however, livestock was rarely included in the humanitarian response, resulting in communities and households fragmenting according to the availability of livestock provisions. We found that rather than a driver for disease, displacement acted as a process affecting community, household and individual zoonotic disease risk dynamics, based on available resources and social networks before, during and after displacement, rooted in the historical, political and socio-economic context. We conclude that in rural Sindh, disaster displaced populations' risk of zoonoses is the result of changes in dynamics rooted in pre-existing structural and chronic inequalities, making people more or less vulnerable to disease through multiple interlinked pathways. Our findings have implications for policy makers and humanitarian responders assisting displaced populations dependent on livestock, with a call to integrate livestock support in humanitarian policies and responses for health, survival and recovery.

Six novel interdisciplinary resilience principles emerging from interdisciplinary exchange around post-COVID-19 centres and peripheries in *Biodiversity* 14 December 2021 by Meredith Root-Bernstein, Alexa Hagerty, François Chiron, Jianguo Liu, Lara Mani, Harini Nagendra, Catharine Ward Thompson

Urban centres and peripheries shape the texture and quality of everyday life, our ability to coexist with and benefit from nature, the risks we face, and our resiliency (UN 2019). Here

we report on lessons we draw as organizers of and presenters at an interdisciplinary, international, and (necessarily) virtual symposium, 'Centres and Peripheries: Reconfiguring Post-COVID-19 Landscapes', which took place on 5 February 2021, and was funded by the UK Embassy to France in the context of the upcoming UN Climate Change Conference 2021 (COP 26). The symposium examined the dynamics emerging between urban and rural, centre and periphery, at global, regional, and landscape scales. Within an international comparative context, the symposium identified solutions and approaches beyond a narrow focus on the 'smart green city' to bring new focus on the systems and social, ecological, and physical infrastructures of human habitation.

The Transformative Potential of Artificial Intelligence in *Futures* 17 December 2021 by Ross Gruetzemacher, Jess Whittlestone

The terms 'human-level artificial intelligence' and 'artificial general intelligence' are widely used to refer to the possibility of advanced artificial intelligence (AI) with potentially extreme impacts on society. These terms are poorly defined and do not necessarily indicate what is most important with respect to future societal impacts. We suggest that the term 'transformative AI' is a helpful alternative, reflecting the possibility that advanced AI systems could have very large impacts on society without reaching human-level cognitive abilities. To be most useful, however, more analysis of what it means for AI to be 'transformative' is needed. In this paper, we propose three different levels on which AI might be said to be transformative, associated with different levels of societal change. We suggest that these distinctions would improve conversations between policy makers and decision makers concerning the mid- to long-term impacts of advances in AI. Further, we feel this would have a positive effect on strategic foresight efforts involving advanced AI, which we expect to illuminate paths to alternative futures. We conclude with a discussion of the benefits of our new framework and by highlighting directions for future work in this area.

4.2 Reports

Reconfiguring Resilience for Existential Risk: Submission of Evidence to the Cabinet Office on the new UK National Resilience Strategy by Matthijs Maas, Diane Cooke, Tom Hobson, Lalitha Sundaram, Haydn Belfield, Lara Mani, Jess Whittlestone, Seán Ó hÉigeartaigh Published on 15 October 2021

This document presents CSER experts' collected responses to the UK Cabinet Office's Call for Evidence on the National Resilience Strategy, as submitted on September 27th, 2021 (response ID: ANON-7FMB-F6JK-W).

"Emerging technologies and dual-use concerns: a horizon scan for global public health" published by the World Health Organization on 28 October by Luke Kemp, Kai Ilchmann, Soatiana Rajatonirina, and Anna Laura Ross.

Dual-use research of concern (DURC) is life science research that is intended for benefit but which might be misapplied to do harm. Such research has increased substantially in the past two decades. It includes, for instance, synthesis of the poliovirus, modification of the mousepox virus, production of mammal-transmissible strains of H5N1 avian influenza and, more recently, de-novo synthesis of the horsepox virus. Dual-use issues can arise in a range of disciplines, beyond gain-of-function experiments. This report presents the results of an international horizon scanning exercise, organized by the WHO to ensure foresight of such risks. The group of experts, from a range of disciplines, undertook a broad examination of scientific and technological developments that could give rise to concern over the next two decades and identified 15 priorities. These fall largely into three overlapping areas: new technologies; issues associated with the convergence of technical areas; and governance of the life sciences.

The Critical Role of Transparency in Addressing the Production Gap Report by Elisa Arond, Harro van Asselt, Siân Bradley, Chandra Bushan, Fergus Green, Ipek Gençsü, Aarti Gupta, Patrick Heller, Robin Hocquet, Natalie Jones, Angela Picciariello, David Manley, Georgia Piggot, Robert Schuwerk, Romain Weikmans Published by the Stockholm Environment Institute on 28 October 2021

The public disclosure of verifiable and comparable information by governments and corporations is key to addressing the fossil fuel production gap. Such information can reveal the extent to which governments are supporting fossil fuel production, and provide insights into how countries can wind down production in light of the Paris Agreement's goals.

Government Support and Policies For Fossil Fuel Production Report by Chandra Bushan, Lucile Dufour, Lisa Fischer, Ipek Gençsü, Patrick Heller, Robin Hocquet, Natalie Jones, David Manley, Miquel Muñoz Cabré, Greg Muttitt, Angela Picciariello, Luma Ramos, Leo Roberts, Bronwen Tucker Published by the Stockholm Environment Institute on 28 October 2021

Governments have injected trillions of US dollars into the economy to respond to the consequences of the COVID-19 pandemic. Many governments have committed to using some of these funds to "build back better," including through public investment in low-carbon development, high-quality clean energy jobs, and a just transition for all. However, the policies, investments, and measures adopted so far have yet to match up with this "build back better" commitment.

The Cambridge Principles: System Stewardship for Universal Owners by Ellen Quigley published on 3 December 2021

Universal owners – including pension funds, sovereign wealth funds, endowments, and other long-term institutional investors – exist to serve their beneficiaries. Because these funds are widely diversified and their returns depend primarily upon the performance of the financial market as a whole (beta), they can best serve beneficiaries and satisfy their legal duties by preserving the health of the whole economy and the environmental and social systems on which it depends.

Contact

Elizabeth Brent
Associate Director — Arts and Humanities
University of Cambridge Development and Alumni Relations
1 Quayside, Bridge Street
Cambridge
CB5 8AB
Elizabeth.Brent@admin.cam.ac.uk
+44 (0)1223 762891
www.cam.ac.uk/yourscambridge

Dr S.J. Beard
Academic Programme Manager and Senior Research Associate
Centre for the Study of Existential Risk
16 Mill Lane
Cambridge
CB2 1SB
sjb316@cam.ac.uk
(+44) 01223 766838
www.cser.ac.uk

UNIVERSITY OF CAMBRIDGE

Dear World...
Yours, Cambridge

The campaign for the University and Colleges of Cambridge