



Centre for the Study of Existential Risk

A report on recent activities

FEBRUARY 2020



CENTRE FOR THE STUDY OF
EXISTENTIAL RISK



The University of Cambridge extends its sincere gratitude for the donation to support the work of the Centre for the Study of Existential Risk (CSER).

Thanks to this generosity, CSER is better equipped to tackle the most pressing challenges facing humanity. Through inter-disciplinary research, and engagement with industry and policy-makers, CSER is bringing together the brightest minds to mitigate existential threats and safeguard civilisation.

A report on CSER’s recent activities follows.

Contents

Introduction	3
1. Current and upcoming activities	4
1.1 Recruitment	4
1.2 Upcoming events	4
2. Review of recent activity	5
2.1 New staff	5
2.2 Policy and industry engagement - impact	6
2.3 Public engagement - field-building	7
2.4 Visiting scholars - field-building	8
2.5 Academic engagement - field-building	9
2.6 Publications	12

This report is confidential and only for the information of the intended recipient

An introduction from **Dr Catherine Rhodes** Executive Director of CSER

“

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.

Our research focuses on Global Catastrophic Biological Risks, Extreme Risks and the Global Environment, Artificial Intelligence and Global Risk, and Managing Extreme Technological Risks. Our work is shaped around three main goals:

- **Understanding:** we study existential risk.
- **Impact:** we develop collaborative strategies to reduce existential risk.
- **Field-building:** we foster a global community of academics, technologists and policy-makers working to tackle existential risk.

This report outlines our past and future plans and activities, as we continue to advance existential risk research and grow the field. Highlights of the past few months include:

- Publication of **six peer-reviewed papers** on: an automatically-updated global catastrophic risk bibliography; methods for estimating global catastrophic risk; the ethics of human extinction; the decline of the Roman Empire; disaster response policy; and responsible AI publication norms.
- Publication of **six policy reports** on: a ‘cartography’ of global catastrophic risk governance; assessing global catastrophic risk drivers; defining transformative AI; privacy and personalised targeting; our 2018 Conference; and reducing the fossil fuel production gap.
- Engagement with **decision-makers** at Google, DeepMind, Partnership on AI, the Singapore Government, Extinction Rebellion and the MoD’s think-tank Development, Concepts and Doctrine Centre.
- Our second biological engineering **horizon-scan workshop**, and many researcher exchanges, which encouraged new research strands and built deeper collaborations.
- Public engagement through several articles, media interviews, and **three Blavatnik Public Lectures**.
- Hiring **four new team members**, with recruitment in progress for four more.



1. Current and upcoming activities

1.1 Recruitment

We are currently recruiting for up to three Research Associates and one Senior Research Associate, from any relevant field of expertise or experience, who can contribute to our existing strands of work, or who might lead the development of work in additional areas of, or approaches to, existential or global catastrophic risk. These positions are funded by a grant which runs through to August 2024.

We also plan to start recruitment for additional administrative support soon.

Summer Visitors / Interns

Last summer we hosted four visiting researchers. This worked well, and we plan to run a similar scheme this year. Outputs from last year's visits included: **Ross Gruetzmacher** co-authored a paper with Shahar Avin that was accepted at the 2020 AIES conference, and a report with Jess Whittlestone, and contributed substantially to the development of Shahar's AI scenario exercise; **Amritha Jayanti** wrote a report with Shahar on 'accountability gaps' and lethal autonomous weapons, and secured a position at Harvard's Belfer Centre; **Nathaniel Cooke** contributed substantially to Luke Kemp's 'Rise and Fall' database; and **Siebe Rozendal** wrote a useful report on Eight high-level uncertainties about global catastrophic and existential risk.

1.2 Upcoming Events

- 3 March: **Blavatnik Public Lecture** – Rachel Bronson: The 2020 Doomsday Clock. President and CEO of the Bulletin of the Atomic Scientists.
- 23 March: **Blavatnik Public Lecture** – Nick Robins: New frontiers for sustainable finance: priorities for action in 2020. Professor in Practice for Sustainable Finance at the Grantham Research Institute.
- 6-7 April: CSER's next international **Cambridge Conference on Catastrophic Risk: 2020** Hindsight.
- 4 June: **Blavatnik Public Lecture** – Matthew Adler: Measuring Social Welfare. Richard A. Horvitz Professor of Law and Professor of Economics, Philosophy and Public Policy at Duke University. He will provide a systematic overview of the social-welfare-function framework, with particular reference to prioritarianism.

2. Review of recent activity

2.1 New staff

We have recruited for three positions for our new Templeton World Charity Foundation project A Science of Global Risk. Two of the new staff have started over the last couple of weeks, and the third will join us in early March.

Lara Mani - Postdoctoral Research Associate, Communication and Outreach for a Science of Global Risk – started January 2020

Lara will work towards building an empirical evidence base for a variety of outreach and communication techniques adopted to present global risk – and to understand how an improved knowledge of global risk can translate to action. Her PhD in Geoscience (University of Plymouth) involved developing video games for communicating volcano risk in the Eastern Caribbean, from which she has developed a strong interest and pioneering expertise in the evaluation of communications strategies and techniques.

Catherine Richards - Part-time Research Assistant for Lord Martin Rees – started January 2020

Catherine has an engineering and business background in the energy and water sectors, and is completing a PhD(Eng) as a John Monash Scholar at the University of Cambridge where she is working on Global Systems Engineering. She was selected as one of 30 women under 30 to watch by Forbes magazine. As well as contributing her own research experience and outputs to CSER's ongoing work Catherine will be supporting the Science of Global Risk programme and will be helping Martin Rees to leverage his expertise and network to better support CSER and our work.

Clarissa Rios Rojas - Postdoctoral Research Associate, Public Policy for Global Risk – starting March 2020

Clarissa is an Eisenhower Fellow and the Founder and Director of the Ekpa'palek NGO. She has worked at the Geneva Center for Security Policy, contributed to projects at the UN Biological Weapons Convention and UN Office for Disarmament Affairs, and is a member of the Global Young Academy. Clarissa has a PhD in Developmental Molecular Biology (University of Queensland). She also contributed to our 2019 Bioengineering Horizon Scan. Clarissa will be working to help us identify and address the many challenges of making public policy for global risk, with a focus on policy co-creation with key stakeholder groups in order to improve policy design and uptake.

Seán Ó hÉigeartaigh and Jose Hernandez Orallo have also recruited a Postdoctoral Research Associate for their Paradigms of AI grant:

John Burden - Postdoctoral Research Associate, Paradigms of Artificial General Intelligence and their Associated Risks – starting July 2020

John Burden has a Bachelors' and Masters' degree in Computer Science from the University of Oxford, and is completing a PhD at the University of York. His current research focuses on reinforcement learning, and his research interests include learning in AI systems, AI safety, and generality.



The CSER team

2.2 Policy and industry engagement – Impact

We have had the opportunity to speak directly with policymakers and institutions across the world who are grappling with the challenge of supporting the socially beneficial aspects of new technologies while mitigating their risks. Through advice and discussions, we have the opportunity to reframe key aspects of the policy debate. Researchers continued their extensive and deep collaboration with industry. Extending our links improves our research by exposing us to the cutting edge of industrial R&D, and helps to nudge powerful companies towards more responsible practices.

- 26-27 September: Seán Ó hÉigartaigh participated in the All Partners Meeting of the **Partnership on AI**.
- Haydn Belfield attended the **CSaP 10 Year Anniversary Reception** in Whitehall (17 Oct), and participated in Participatory Machine Learning, **Google's People and AI Research** (14 Nov).
- 18-20 October: Shahar Avin, Lauren Holt, Ross Gruetzemacher and Haydn Belfield presented at the **Effective Altruism Global London**.
- 21 November: Simon Beard attended a meeting at the **Grantham Institute** that brought together scientists and activists from the Extinction Rebellion movement to improve their understanding and communication of science.
- 3 December: Catherine Rhodes participated in a **Chatham House** workshop on Bridge Building in the Nuclear Disarmament Discourse.
- 3-4 December: Shahar Avin, Sabin Roman, Lauren Holt and Haydn Belfield presented at the annual conference of the **Development, Concepts and Doctrine Centre (DCDC)**, the “MoD’s think-tank”.
- 5 December: CSER group meeting with **Tom Countryman**, Chair of the Arms Control Association board of directors and distinguished US diplomat.
- 21 January: **Biosecurity event** in Parliament, organised by the All-Party Parliamentary Group for Future Generations and the Parliamentary Office for Science and Technology (POST). It was a closed briefing session to Members of Parliament about challenges to cross-government coordination to address biosecurity threats, and Catherine Rhodes and Des Browne were both invited speakers.
- The **APPG for Future Generations** also launched their Inquiry into ‘Long-termism in Policy-making’, advised on drafting of the Future Generations Bill, and secured funding for the next year.

2.3 Public engagement – Field-building

- Asaf Tzachor wrote a **guest post** ‘Down the Hunger Spiral: Pathways to the Disintegration of the Global Food System’ for the Council on Foreign Relations’ blog series Averting Global Catastrophe.
- Haydn Belfield and Gabriel Recchia wrote a **conversation article** ([here](#)) based on our recent report *Global Catastrophic Risks Policy 1 – Understand*, arguing that no country is sufficiently prepared for global risks.
- **Simon Beard was profiled in Vice** as “This Guy Studies the ‘Global Systems Death Spiral’ That Might End Humanity”.
- BBC Science Focus’ **front page cover-story** on Mass extinction: Can we stop it? quotes Lauren Holt and Simon Beard extensively.
- We announced our **fundraising success**. The five year grant from the Isaac Newton Trust, matching funds from generous philanthropists and foundations, new grants and funding renewals will allow the Centre to continue and expand its research programmes for the 2019-2024 period.
- **Lord Martin Rees lectured** at the Global Grand Challenges Summit 2019 ([video](#)) and the Perimeter Institute for Theoretical Physics ([video](#)) and was interviewed by the *Globe and Mail* newspaper, and by Brian Cox on the BBC ([video](#)).

We’re reaching more and more people with our research

12,202



website visitors over the last two months

7,000+



newsletter subscribers

8,542



Twitter followers

2,701



Facebook followers

2.4 Visiting scholars – Field-building

- **Dr Nick Evans**, University of Massachusetts, Lowell, visited in October-November 2019, working on a book project on scientific freedom as a factor in navigating dual-use risks in (but not limited to) biology. He also worked on a possible project with Lalitha Sundaram. He gave a work-in-progress seminar on 4 November.
- **Dr H Orri Stefansson** visited 13 – 17 October 2019, and gave a work-in-progress seminar on 14 October. Orri is currently a Pro Futura Fellow at the Institute for Future Studies in Stockholm, working mainly on decision theory and developed collaborations with Yang Liu and other CSER researchers during his visit. Dr Stefansson plans to spend a longer period at CSER as part of the Pro Futura Fellowship within the next couple of years.
- **Jaime Sevilla** is visiting until April 2020, working on forecasting the impact of new technologies – such as Artificial Intelligence and Quantum Computing – and understanding key mathematical considerations to improve decision-making – such as extreme value theory, causal inference and decision theory.



Cover image from BBC Science Focus issue which quotes Laura Holt and Simon Beard



Promotional image for Professor Brian Cox's interview with Professor Martin Rees which was broadcast by the BBC

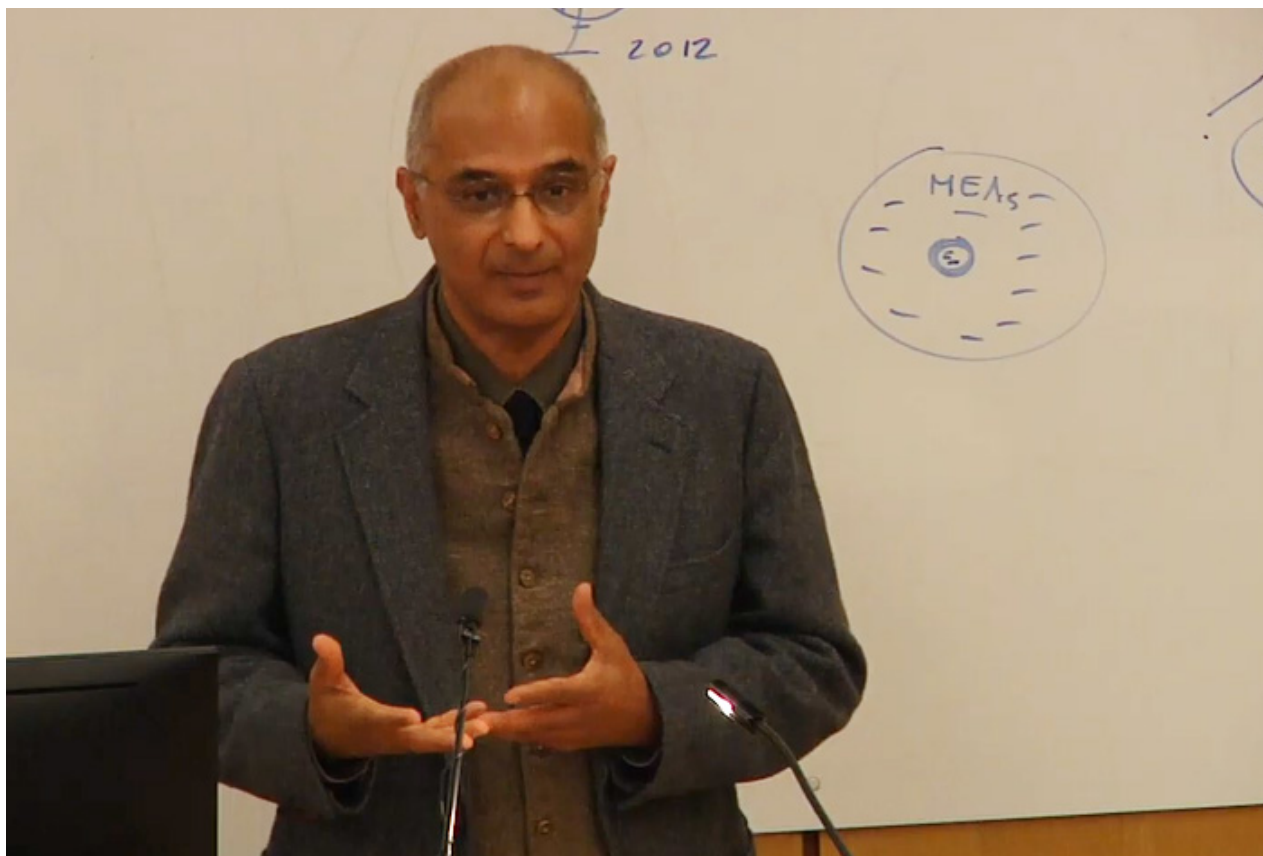
- **Shivam Patel** is also visiting until April 2020, and is working on modelling the AI development ecosystem with an agent-based approach. Shivam is an undergraduate student from Nirma University in India, and is working under the supervision of Jess Whittlestone and Shahar Avin.
- **Nick Robins**, Professor in Practice for Sustainable Finance at the Grantham Research Institute, will be visiting for a few days around the time of his Public Lecture on 23 March, and will be working with Ellen Quigley on a co-authored paper on Universal Ownership and the Just Tradition. **Trude Myklebust**, from the University of Oslo, will also be visiting during this period (around 16-27 March), and co-authoring a paper with Ellen on Universal Ownership Theory.
- **Edward Elliot**, from the Australian Department of Foreign Affairs and Trade will visit for a few months over the summer. He will be investigating policy frameworks in Australia, the UK and the US for managing global catastrophic biological risks.



2.5 Academic engagement – Field-building

As an interdisciplinary research centre within Cambridge University, we seek to grow the academic field of existential risk research, so that it receives the rigorous and detailed attention it deserves.

- 19 September: Catherine Rhodes presented on 'Power, Trust and distrust in the Governance of (bio)Technologies' at the **Cambridge Trust and Technology Initiative 2019 Symposium**.
- 30 September - 4 October: Simon Beard and Luke Kemp visited the **Institute for Futures Studies** in Stockholm, including meeting with the Stockholm Resilience Centre and the Global Challenges Foundation.
- 1 October: **Blavatnik Public Lecture – Jason Matheny**. Founding Director of the Center for Security and Emerging Technology at Georgetown University. Previously he was Assistant Director of National Intelligence and Director of IARPA. He is a member of the National Security Commission on Artificial Intelligence and was named one of Foreign Policy's "Top 50 Global Thinkers."
- 9 October: **Biological Engineering Horizon-Scanning Workshop** (led by Luke Kemp). This is our second horizon-scan, building on the first which was welcomed by academics and presented at the Biological Weapons Convention.
- Ellen Quigley gave a lecture in Stockholm (20 September), presented at Cambridge's **Carbon Neutral Futures Initiative Town Hall Meeting** (24 September), and gave a guest lecture at Copenhagen Business School (21 October).
- 30 October: **Blavatnik Public Lecture – Zia Mian**. Co-director of Princeton University's Program on Science and Global Security. Received the 2014 Linus Pauling Legacy Award for "his accomplishments as a scientist and as a peace activist in contributing to the global effort for nuclear disarmament". [Video](#).



Zia Mian delivering the Blavatnik Public Lecture 'A Conceivable Horizon of Horror' (30 Oct 2019)

- 1 November: Lalitha Sundaram lectured at the **Imperial MRes**.
- 12 November: **visit from Dr Toby Ord**, Future of Humanity Institute, Oxford University.
- 18 November: **Blavatnik Public Lecture – Grethe Helene Evjen**. Senior advisor at Norwegian Ministry of Agriculture and Food and key to the implementation and coordination of the Svalbard Global Seed Vault. Video.
- 19 November: Simon Beard gave a talk at the **University of Kent** on the subject 'Is Extinction Imminent?'
- 20 – 22 November: Lauren Holt participated in the **Hamburg Insecurity Sessions** conference on 'Uncancelling the Future'.
- 28 November: **visit from Professor Hiski Haukkala**, University of Tampere and former Foreign Policy Advisor to the Finnish President.
- 5-6 December: Haydn Belfield presented on CSER research at the **Future of Humanity Institute**, Oxford University.
- 8-14 December: At NeurIPS, the world's leading AI conference, Alexa Hagerty ran a **workshop on 'Minding the Gap: Between Fairness and Ethics'**, and Jess Whittlestone presented a paper.
- 14 January: CSER group meeting with Cat Tully (Director) and Andrew Curry (Director of Futures) from the **School of International Futures**, to discuss approaches and potential collaborations on futures work.
- 15 January: CSER group meeting with foresight group from **Cambridge Display Technologies**.
- 16 January: visit from Jenn Chubb and Prof Peter Cowling from **York University's AI Futures** initiative.



Grethe Helene Evjen delivering the Blavatnik Public Lecture 'Svalbard Global Seed Vault - Saving Seeds for Eternity' (18 Nov 2019)

We continued our weekly 'Work-in-Progress' series:

- 16 September: **Des Browne**
- 23 September: **Asaf Tzachor**: AI for Global Food Security
- 1 October: Roundtable with **Jason Matheny**
- 7 October: **Shahar Avin** – updates on 'Intelligence Rising' AI Scenario Role Play, Trust in AI development, Epistemic Security, and AI impacts on Strategic Stability.
- 14 October: **H. Orri Stefánsson** - Is the Precautionary Principle plausible for existential risk management?
- 21 October: **Haydn Belfield** – Effective Altruism introduction and overview
- 28 October: **Simon Beard** led a group discussion on our approach to politics and power, including a briefing from Steffan Hasselwimmer of Cambridgeshire Climate Emergency
- 30 October: Roundtable with **Zia Mian**
- 4 November: **Nick Evans** – Scientific Freedom
- 18 November: Roundtable with **Grethe Helene Evjen**
- 25 November: **Sabin Roman** - Societal Collapse: What it is, how it occurs and why?
- 2 December: **Ellen Quigley** – Cambridge University Responsible Investment Report
- 9 December: **Olaf Corry** - Geoengineering Governance and International Relations
- 20 January: **Anders Sandberg** – Macrostrategy Research Agenda



The launch of the campaign for a Future Generations Bill, based on CSER research and enabled by the APPG for Future Generations CSER founded.

2.6 Publications

Peer-reviewed papers

Gorm Shackelford*, Luke Kemp*, Catherine Rhodes*, Lalitha Sundaram*, Seán Ó hÉigeartaigh*, Simon Beard*, Haydn Belfield*, Julius Weitzdörfer*, Shahar Avin*, Dag Sørenbø, Elliot M. Jones, John B. Hume, David Price, David Pyle, Daniel Hurt, Theodore Stone, Harry Watkins, Lydia Collas, and William Sutherland*. (2019). **Accumulating evidence using crowdsourcing and machine learning: a living bibliography about existential risk and global catastrophic risk.** *Futures*.

“The study of existential risk—the risk of human extinction or the collapse of human civilization—has only recently emerged as an integrated field of research, and yet an overwhelming volume of relevant research has already been published. To provide an evidence base for policy and risk analysis, this research should be systematically reviewed. In a systematic review, one of many time-consuming tasks is to read the titles and abstracts of research publications, to see if they meet the inclusion criteria. We show how this task can be shared between multiple people (using crowdsourcing) and partially automated (using machine learning), as methods of handling an overwhelming volume of research. We used these methods to create The Existential Risk Research Assessment (TERRA), which is a living bibliography of relevant publications that gets updated each month (www.x-risk.net). We present the results from the first ten months of TERRA, in which 10,001 abstracts were screened by 51 participants. Several challenges need to be met before these methods can be used in systematic reviews. However, we suggest that collaborative and cumulative methods such as these will need to be used in systematic reviews as the volume of research increases.” [Link to paper here.](#)

Simon Beard*, Thomas Rowe, and James Fox. (2019). **An Analysis and Evaluation of Methods Currently Used to Quantify the Likelihood of Existential Hazards.** *Futures*.

“This paper examines and evaluates the range of methods that have been used to make quantified claims about the likelihood of Existential Hazards. In doing so, it draws on a comprehensive literature review of such claims that we present in an appendix. The paper uses an informal evaluative framework to consider the relative merits of these methods regarding their rigour, ability to handle uncertainty, accessibility for researchers with limited resources and utility for communication and policy purposes. We conclude that while there is no uniquely best way to quantify Existential Risk, different methods have their own merits and challenges, suggesting that some may be more suited to particular purposes than others. More importantly, however, we find that, in many cases, claims based on poor implementations of each method are still frequently invoked by the Existential Risk community, despite the existence of better ones. We call for a more critical approach to methodology and the use of quantified claims by people aiming to contribute research to the management of Existential Risk, and argue that a greater awareness of the diverse methods available to these researchers should form an important part of this.” [Link to paper here.](#)

Patrick Kaczmarek* and Simon Beard*. (2019). **Human Extinction and Our Obligations to the Past.** *Utilitas*.

“On certain plausible views, if humanity were to unanimously decide to cause its own extinction, this would not be wrong, since there is no one whom this act would wrong. We argue this is incorrect. Causing human extinction would still wrong someone; namely, our forebears who sacrificed life, limb and livelihood for the good of posterity, and whose sacrifices would be made less morally worthwhile by this heinous act.” [Link to paper here.](#)

*indicates the author is a member of CSER



Sabin Roman* and Erika Palmer. (2019). **The Growth and Decline of the Western Roman Empire: Quantifying the Dynamics of Army Size, Territory, and Coinage.** *Clodynamics*, 10(2).

“We model the Western Roman Empire from 500 BCE to 500 CE, aiming to understand the interdependent dynamics of army size, conquered territory and the production and debasement of coins within the empire. The relationships are represented through feedback relationships and modelled mathematically via a dynamical system, specified as a set of ordinary differential equations. We analyze the stability of a subsystem and determine that it is neutrally stable. Based on this, we find that to prevent decline, the optimal policy was to stop debasement and reduce the army size and territory during the rule of Marcus Aurelius. Given the nature of the stability of the system and the kind of policies necessary to prevent decline, we argue that a high degree of centralized control was necessary, in line with basic tenets of structural-demographic theory.”
[Link to paper here.](#)

Julius Weitzdörfer* and Simon Beard*. (2019). **Law and Policy Responses to Disaster-Induced Financial Distress**, in Akiko Kamesaka and Franz Waldenberger (Eds.) *Governance, Risk and Financial Impact of Mega Disasters: Lessons from Japan*. Springer, Singapore.

“This chapter treats disaster response policies directed at the economic recovery of private households. First, we examine problems of disaster-induced financial distress from a legal and economic perspective. We do this both qualitatively and quantitatively, and focussing on residential loans, using the victims of the 11 March 2011 tsunami as our example. Then, using doctrinal and systematic analysis, we set out the broad array of law and policy solutions tackling disaster-induced debt launched by the Japanese Government. On this basis, we assess the strengths and weaknesses of these measures in terms of their practical adequacy to prevent and mitigate financial hardship and examine them against multiple dimensions of disaster justice. We conclude with suggestions for improving financial disaster recovery by taking a prospective approach, preventing the snowballing of disaster-related losses, which we argue represents an equitable and effective way forward in allocating resources following future mega disasters.”
[Link to paper here.](#)

Jess Whittlestone* and Aviv Ovadya. (2019). **The tension between openness and prudence in responsible AI research.** *NeurIPS 2019 Joint Workshop on AI for Social Good*.

“This paper explores the tension between openness and prudence in AI research, evident in two core principles of the Montréal Declaration for Responsible AI. While the AI community has strong norms around open sharing of research, concerns about the potential harms arising from misuse of research are growing, prompting some to consider whether the field of AI needs to reconsider publication norms. We discuss how different beliefs and values can lead to differing perspectives on how the AI community should manage this tension, and explore implications for what responsible publication norms in AI research might look like in practice.” [Link to paper here.](#)

Reports

Luke Kemp* and Catherine Rhodes*. (2020). **The Cartography of Global Catastrophic Governance.** Global Challenges Foundation, Stockholm.

“The international governance of global catastrophic risks (GCRs) is fragmented and insufficient. This report provides an overview of the international governance arrangements for eight different GCR hazards and two drivers. We find that there are clusters of dedicated regulation and action, including in nuclear warfare, climate change and pandemics, biological and chemical warfare. Despite these concentrations of governance their effectiveness is often questionable. For others, such as catastrophic uses of AI, asteroid impacts, solar geoengineering, unknown risks, super-volcanic eruptions, inequality and many areas of ecological collapse, the legal landscape is littered more with gaps than effective policy.” [Link to report here.](#)





Simon Beard* and Phil Torres. (2020). **Identifying and Assessing the Drivers of Global Catastrophic Risk**. Global Challenges Foundation, Stockholm.

“This report highlights how the body of emerging GCR research has failed to produce sufficient progress towards establishing a unified methodological framework for studying these risks. Key steps that will help to produce such a framework include moving away from a hazard-focused conception of risk, typified by the majority of quantitative risk assessments that we analyse, and toward a more sophisticated approach built on a mature understanding of risk assessment and disaster risk reduction and preparedness. We further suggest that a key barrier to the development of a mature science capable of comprehensively assessing the drivers of GCRs has been the political, philosophical, and economic context within which the field has arisen, as demonstrated by five distinct “waves” of GCR research. We believe that a suitably committed funder that transcends these contextual boundaries could have a transformative impact on the discipline, and with it our understanding of GCRs and their drivers. We propose that the Global Challenges Foundation is in a uniquely strong position to play this role.”

Ross Gruetzmacher and Jess Whittlestone*. (2019). **Defining and Unpacking Transformative AI**. arXiv:1912.00747.

“Recently the concept of transformative AI (TAI) has begun to receive attention in the AI policy space. TAI is often framed as an alternative formulation to notions of strong AI (e.g. artificial general intelligence or superintelligence) and reflects increasing consensus that advanced AI which does not fit these definitions may nonetheless have extreme and long-lasting impacts on society. However, the term TAI is poorly defined and often used ambiguously. Some use the notion of TAI to describe levels of societal transformation associated with previous ‘general purpose technologies’ (GPTs) such as electricity or the internal combustion engine. Others use the term to refer to more drastic levels of transformation comparable to the agricultural or industrial revolutions. The notion has also been used much more loosely, with some implying that current AI systems are already having a transformative impact on society. This paper unpacks and analyses the notion of TAI, proposing a distinction between TAI and radically transformative AI (RTAI), roughly corresponding to societal change on the level of the agricultural or industrial revolutions. We describe some relevant dimensions associated with each and discuss what kinds of advances in capabilities they might require. We further consider the relationship between TAI and RTAI and whether we should necessarily expect a period of TAI to precede the emergence of RTAI. This analysis is important as it can help guide discussions among AI policy researchers about how to allocate resources towards mitigating the most extreme impacts of AI and it can bring attention to negative TAI scenarios that are currently neglected.” [Link to report here](#).

Karina Vold and Jess Whittlestone*. (2019). **Privacy, autonomy and personalised targeting: rethinking how personal data is used**. IE Report on Data, Privacy and the Individual.

“Technological advances are bringing new light to privacy issues and changing the reasons for why privacy is important. These advances have changed not only the kind of personal data that is available to be collected, but also how that personal data can be used by those who have access to it. We are particularly concerned with how information about personal attributes inferred from collected data (such as online behaviour), can be used to tailor messages and services to specific individuals or groups. This kind of ‘personalised targeting’ has the potential to influence individuals’ perceptions, attitudes, and choices in unprecedented ways. In this paper, we argue that because it is becoming easier for companies to use collected data for influence, threats to privacy are increasingly also threats to personal autonomy—an individual’s ability to reflect on and decide freely about their values, actions, and behaviour, and to act on those choices. While increasing attention is directed to the ethics of how personal data is collected, we make the case that a new ethics of privacy needs to also think more rigorously about how personal data may be used, and its potential impact on personal autonomy.” [Link to report here](#).





Simon Beard* and Catherine Rhodes*. (2020). **Conference Report: Cambridge Conference on Catastrophic Risk** (CCCR2018). Centre for the Study of Existential Risk, University of Cambridge.

“The second of the Centre for the Study of Existential Risk’s international conferences provided a timely opportunity for the Centre, along with the wide communities working on existential and global catastrophic risks and in related fields, to reflect on our work so far and to deepen and broaden our learning from other disciplines. This allowed us to both focus on some of the practical challenges of the task that we have set ourselves and identify what we are doing well. Hence, the Conference served not only to address important issues facing the existential risk research community, but also to establish and maintain the connections with other communities that have important contributions to make to the developing ‘science’ of existential risk research.”
[Link to report here.](#)

Ploy Achakulwisut, Harro van Asselt, Peter Christoff, Richard Denniss, Fergus Green, Natalie Jones*, Georgia Piggot, Oliver Sartor, Fernando Tudela, Cleo Verkuijl, Oscar Widerberg. (2019). **Policy options to close the production gap**, in The Production Gap: 2019 report.

“Countries can begin to close the production gap by aligning their energy and climate plans. Governments have a range of policy options to regulate fossil fuel supply, including limits on new exploration and extraction and removal of subsidies for production. Some countries are already demonstrating leadership: Belize, Costa Rica, Denmark, France, and New Zealand have all enacted partial or total bans on oil and gas exploration and extraction. Germany and Spain are phasing out coal extraction. Non-state actors and subnational governments can also help facilitate a transition away from fossil fuels, by mobilizing constituencies and shifting investment to lower carbon options. Individuals and institutions have already pledged to divest over USD 11 trillion from fossil fuel holdings. Several governments are planning for a “just transition” that aims to minimize disruption for affected workers and communities.” [Link to report here.](#)



The CSER team pictured in Summer 2019.





Contact

Daisy Ayre

**Senior Associate Director
Arts and Humanities**

University of Cambridge Development
and Alumni Relations

1 Quayside, Bridge Street

Cambridge

CB5 8AB

daisy.ayre@admin.cam.ac.uk

+44 (0)1223 764117

www.cam.ac.uk/yourscambridge

Haydn Belfield

Academic Project Manager

Centre for the Study of
Existential Risk

16 Mill Lane

Cambridge

CB2 1SB

hb492@cam.ac.uk

+44 (0)770 9676 355

www.cser.ac.uk



**UNIVERSITY OF
CAMBRIDGE**

Dear World...

Yours, Cambridge

