

# Centre for the Study of Existential Risk



UNIVERSITY OF  
CAMBRIDGE

A report on recent activities

**NOVEMBER 2020**



**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).**

The last quarter has seen the CSER team sharing findings with peers, and recommendations and expertise with key policy-makers. Their efforts are ensuring attention is directed to key issues that impact existential risk, in order to reframe policy debate and mitigate the threats posed to humanity.

## **Contents**

Overview	3
1. Coronavirus update	4
2. Upcoming Conference programme and speakers	5
3. New Staff	6
4. Policy and industry engagement - impact	7
5. Public engagement – Field-building	8
6. Academic engagement – Field-building	10
7. 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, Risks from Artificial Intelligence, 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.

Our [last report](#) covered January – May 2020. This report covers the period May – September 2020 and outlines our activities and future plans. Highlights include:

- Contributing to **Cambridge University's decision to fully divest from fossil fuels**.
- Publication of **nine peer-reviewed papers** on: divestment and Cambridge; our second bioengineering horizon-scan; cross-cultural cooperation in AI governance; ethics with urgency in a crisis; two AI safety papers on paradigms and degrees of generality; capital theory and sustainability; famine dynamics; and existential risk assessment.
- Publication of **four written evidence submissions, and two policy briefs**, on premature procurement of AI defence systems; the EU's proposed AI regulation; proposals for the new Advanced Research Projects Agency (ARPA) UK funding body; how to improve the UK's risk management system; using AI to improve food supply chain security, and public mental health and COVID-19.
- Seán Ó hÉigartaigh joined the **Global Partnership on AI**.
- Co-organising two workshops **with the Institute for Public Policy Research (IPPR)** on catastrophic environmental breakdown.
- Extensive media coverage – including the **front page of *The Economist***.
- Welcoming six new team members.

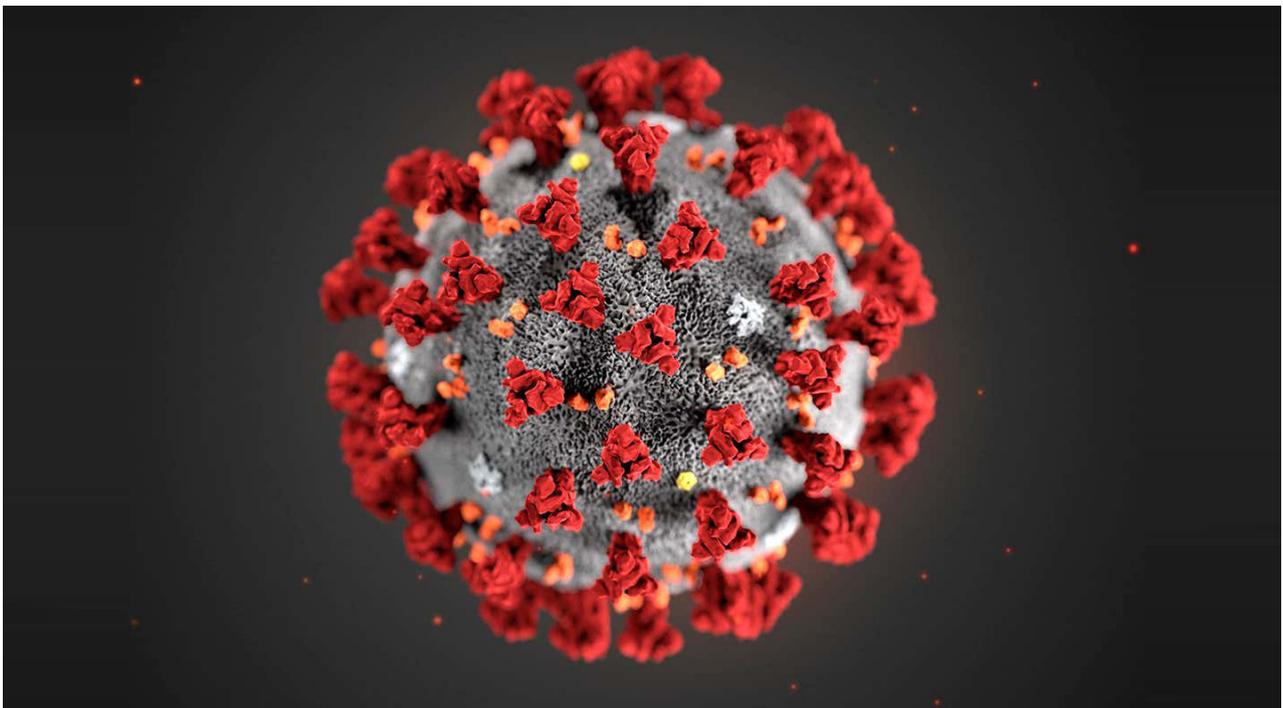
# 1. Coronavirus update

The team has successfully adapted to fully-remote work, with regular use of video conferencing and Slack. We are fortunate that much of our research can be done remotely. We are currently planning a physically distanced return to the office later in the year, subject to the situation and University advice.

We have also been doing some work to contribute to the response:

- A comment piece in *Nature* '[Artificial intelligence in a crisis needs ethics with urgency](#)' and contributions to the preprint '[Informing management of lockdowns and a phased return to normality: a Solution Scan of non-pharmaceutical options to reduce SARS-CoV-2 transmission](#)'. The preprint has received extensive media coverage. Shahar Edgerton Avin, Lara Mani, Jess Whittlestone, Tom Hobson are exploring a project on COVID Implementation Failures.
- Simon Beard spoke on a [COVID-19 panel](#) for ALDES (the Association of Liberal Democrat Engineers and Scientists) and the Young Liberals, and was the lead author on two policy briefs on [Public Mental Health and COVID-19](#) and [AI and Data Governance issues in responding to COVID-19](#).
- Our researchers have been interviewed about the pandemic for [Sky News](#), [BBC World Service](#), [Spear's](#), [the Guardian](#), [Politico](#) and [Newsweek](#).

Finally, we have passed on our thanks to the Big May Ball Appeal for Coronavirus for their £26,500 [donation](#), especially its student co-founders Jade Charles and Zehn Mehmood, and to all the students who generously donated their ticket refunds.



## 2. Upcoming Conference programme and speakers

The Cambridge Conference on Catastrophic Risk '2020 Hindsight: Our Final Century Revisited' ([CCCR 2020](#)), our third major international conference, will be held remotely 16-19 November 2020. It follows our 'CCCR' conferences in [2016](#) and [2018](#).

The programme will draw on key themes from Lord Martin Rees' seminal 2003 book [Our Final Century](#), to reflect on the growth and development of the field of global catastrophic risk research over the last decades. CCCR2020 is one of the few forums in the world discussing such issues, and we intend the conference to contribute to the further development of the community working on global risks, and its engagement with other relevant groups. To allow for in-depth collaboration, it is invite-only. The keynote will be by Lord Martin Rees and there will be a range of opportunities for networking and debate.

Sessions will include:

**1. Pre-2000: What global catastrophic risk researchers can learn from the history of global catastrophic risk**

Speakers: Francesco Calogero, Malcolm Potts  
Chair: Simon Beard

**2. 50/50: Assessing the chances of a global catastrophe**

Speakers: Nancy Connell, Karim Jebari  
Chair: Sabin Roman

**3. Are we nearly there yet? Taking the long and longer-term view of humanity**

Speakers: Anders Sandberg, Sheri Wells-Jensen  
Chair: Clarissa Rios Rojas

**4. Global justice and global catastrophic risk: Between error and terror**

Speakers: Bentley Allan, Ndidi Nwaneri  
Chair: Natalie Jones

**5. Threats without enemies: natural global disasters and their consequences**

Speakers: Doug Erwin, Lindley Johnson  
Chair: Lara Mani

**6. Governing science: Does it matter who is doing our scientific research and why?**

Speakers: Stuart Parkinson, Heather Roff  
Chair: Lalitha Sundaram

**7. Global catastrophic environmental risks: systemic collapse from anthropogenic environmental change**

Speakers: Jason Hickel, Veerabhadran Ramanathan, Alisha Graves, Salamatou Abdourahmane  
Chair: Luke Kemp

# 3. New Staff

Over the last few months we have welcomed five new postdoctoral researchers and one senior research associate:



[Dr Freya Jephcott](#) – focusing on the challenge of ‘unseen epidemics’ and the effectiveness of outbreak response systems. Freya also participates in work in the field, and policy advisory work on complex health emergencies with Médecins Sans Frontières (MSF) and the World Health Organization (WHO).



[John Burden](#) - focusing on the challenges of evaluating the capability and generality of AI systems. He is working on the project [Paradigms of Artificial General Intelligence and Their Associated Risks](#). He has a background in Computer Science and is in the final stages of completing his PhD from the University of York.



[Dr Tom Hobson](#) - focusing on the militarisation of emerging technologies, particularly biological technologies. Tom has a background in International Relations, Security Studies & STS, having completed his PhD within the Centre for War & Technology at the University of Bath.



[Dr Matthijs Maas](#) – focusing on adaptive global governance approaches for extreme technological risks, especially regimes for high-stakes or destabilising uses of AI technology, such as in military contexts. Matthijs has a PhD in Law from the University of Copenhagen.



[Dr Alex McLaughlin](#) - focusing on the moral questions that arise at the intersection of global justice and existential risk. He joins our [Global Justice and Global Catastrophic Risk](#) research area. He was previously a Leverhulme Doctoral Scholar in Climate Justice at the University of Reading



[Dr Jess Whittlestone](#) – focusing on AI ethics and policy, especially on what we can do today to ensure AI is safe and beneficial in the long-term. She joins us as a Senior Research Associate from our sister centre the Leverhulme Centre for the Future of Intelligence.

## 4. Policy and industry engagement – impact

We have met and engaged with policy makers and institutions across the world who are grappling with the challenge of global risks. Through these personal connections and institutional advice, we have had the opportunity to reframe key aspects of the policy debate. In addition, we continued our flourishing collaboration with corporate leaders. Extending our links improves our research and allows us to advise companies on more responsible practices.

- The Vice-Chancellor announced that the University of Cambridge will divest from all direct and indirect investments in fossil fuels by 2030. The £3.5 billion Cambridge University Endowment Fund – the biggest in Europe – intends to ramp up investments in renewable energy as it divests from fossil fuels. Our researcher Dr Ellen Quigley was [appointed](#) in May 2019 to work with the University of Cambridge's Chief Financial Officer on responsible investment. She is the lead author of a [report](#) on the advantages and disadvantages of divestment which the University released as part of the announcement, which explores divestment across moral, social, political, reputational, and financial dimensions. The news was covered on the [University website](#), the [Financial Times](#), the [Washington Post](#), the [Telegraph](#), the [Guardian](#) and elsewhere.
- Seán Ó hÉigeartaigh has joined the [Global Partnership on AI](#), an international and multi-stakeholder initiative to guide the responsible development and use of AI. The members are Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, the United States of America, and the European Union.
- Clarissa Rios Rojas advised the S20 on foresight policy recommendations. The S20 works with national science academies to support the G20 summit. She also contributed to a [collaborative exercise](#) with the Geneva Centre for Security Policy (GCSP) on the trends that are establishing themselves and accelerating as a result of the COVID-19 pandemic.
- Haydn Belfield, Amritha Jayanti and Shahar Avin submitted [written evidence](#) in May to the UK's Defence Committee on the risks of militaries procuring AI defence systems 'prematurely' – before they are fully de-risked, safe and secure.
- Haydn Belfield, José Hernández-Orallo, Seán Ó hÉigeartaigh, Matthijs M. Maas, Alexa Hagerty, Jess Whittlestone submitted a [Response to EU White Paper on AI](#) in June, which mainly focuses on mandatory conformity assessments for high-risk AI applications carried out by independent testing centres. Our key recommendation is to keep this proposed framework and not water it down.
- Nick Bostrom, Haydn Belfield and CSER Research Affiliate Sam Hilton submitted evidence on [UK ARPA – Key Recommendations](#) in September to the UK Parliament Science & Technology Committee, recommending a focus on high-impact risk research and global leadership in emerging technologies.
- Our researchers submitted evidence, which is not yet online, to the UK Parliament Science & Technology Committee on the UK's risk assessment and management system.
- 30 July, 6 August: CSER co-hosted two remote workshops with IPPR to explore implications of destabilisation resulting from environmental breakdown for efforts to realise more sustainable, equitable, resilient societies.
- Simon Beard spoke on a [COVID-19 panel](#) for ALDES (the Association of Liberal Democrat Engineers and Scientists) and the Young Liberals, and was the lead author on two policy briefs on [Public Mental Health and COVID-19](#) and [AI and Data Governance issues in responding to COVID-19](#).
- The APPG for Future Generations held several more evidence sessions as part of their Inquiry into long-termism in policy.

## 5. Public engagement – Field-building

- *The Economist* [front page](#). CSER's research – and catastrophic and existential risks more broadly – were featured in a front-page story in the Economist, pictured below. Haydn Belfield and CSER Research Affiliate Rumtin Sepasspour were interviewed for the story.
- *Telegraph* [opinion piece](#) 'This pandemic was entirely predictable. Why were we so poorly prepared?' by Lord Martin Rees and CSER Research Affiliates Angus Mercer and Sam Hilton.
- Clarissa Rios Rojas and Catherine Rhodes were interviewed for a [front cover article](#) in Chile's *La Segunda* about their work in existential risk.
- Clarissa Rios Rojas was interviewed on Peru's the [Sintesis Podcast](#), and for the [La Republica newspaper](#) about her recommendations in relation to science and policy during the COVID-19 crisis. She was also featured in the Global Young Academy 'Women in Science' Working Group video [Follow Your Dreams](#), and a UN Geneva [video](#) highlighting women that are helping to reduce the threat posed by bioweapons.
- Luke Kemp was interviewed in *Spear's*: '[COVID-19 is a black elephant](#)'.
- CNBC wrote a profile of the Centre for the Study of Existential Risk, the Leverhulme Centre for the Future of Intelligence and the Future of Humanity Institute: [How Britain's oldest universities are trying to protect humanity from risky A.I.](#)
- Sabin Roman and Luke Kemp were interviewed for a [Live Science](#) article on 'What could drive humans to extinction?'
- Simon Beard was interviewed on a podcast about [The Post Covid World?](#)
- Haydn Belfield spoke at a CogX event on '[Investing in systemic risk](#)' and was quoted in [Science|Business](#) on the EU's AI regulation.
- Chair of the CSER Management Board, Sir Partha Dasgupta, was interviewed for Sir David Attenborough's flagship new programme [Extinction: The Facts](#).
- Lord Martin Rees has a new [Twitter profile](#) and [website](#), developed by CSER research assistant Catherine Richards. He wrote an Aeon article on [How can scientists best address the problems of today and the future?](#). He also spoke on Being a Scientist at the Cheltenham Science Festival ([video](#), see also this [article](#)) and was interviewed on Sky News on [After The Pandemic: How will life and society change?](#)

We have also started a new Meet the Researcher series on our website. First up is Luke Kemp - [here](#).

*We're reaching more and more people with our research*

9,347



website visitors over the last two months

7,615



newsletter subscribers

10,450



Twitter followers

3,001



Facebook followers

# The Economist

Israel's annexation misjudgment

Commercial property in trouble

Why the world neglects the Uighurs

Fortnite as an Olympic sport

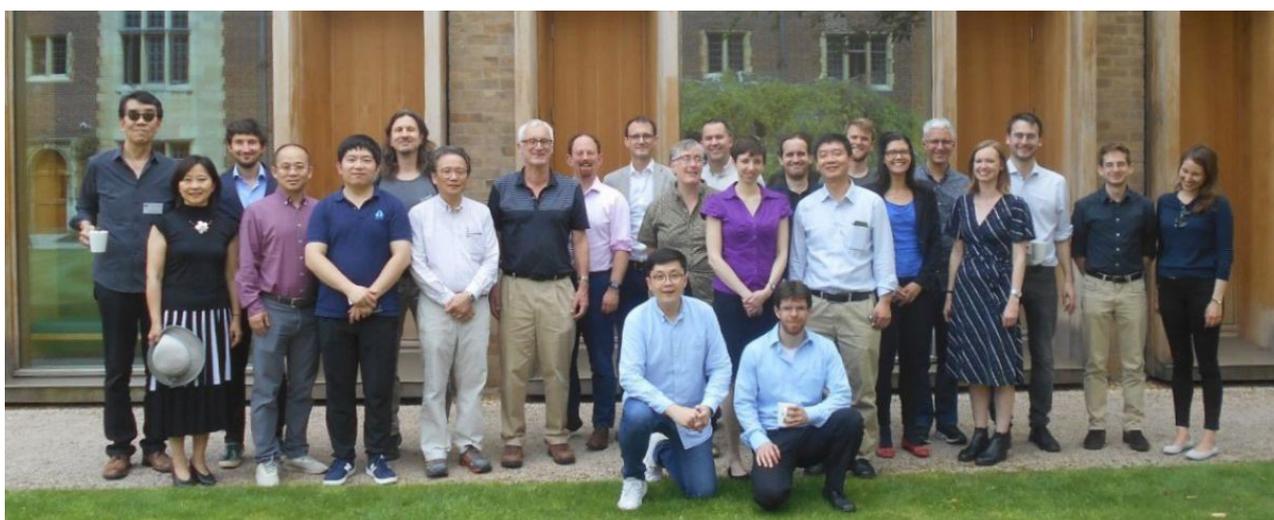
JUNE 27TH-JULY 3RD 2020

## The next catastrophe (and how to survive it)



## 6. Academic engagement – Field-building

- A workshop we co-organised with Chinese and Cambridge participants led to the publication of '[Overcoming Barriers to Cross-Cultural Cooperation in AI Ethics and Governance](#)' by a Cambridge-Beijing team, available in [English](#) and [Chinese](#). In order for AI to bring about global benefits, cross-cultural cooperation will be essential. Such cooperation will enable advances in one part of the world to be shared with other countries, and ensure that no part of society is neglected or disproportionately negatively impacted by AI. Without such cooperation, competitive pressures between countries may also lead to underinvestment in safe, ethical, and socially-beneficial AI development, increasing the global risks from AI. It was covered in CNBC on 4 June '[Academics call on nations to work together on A.I. and ensure it benefits all of humanity](#)'.
- Ellen Quigley was awarded a Global Research Alliance for Sustainable Finance and Investment [GRASFI Paper Prize](#) for 'Best Paper for Potential Impact on Sustainable Finance Practices' for her preprint paper 'Universal Ownership in Practice: A practical positive investment framework for asset owners'.
- '[Existential Risk of Climate Change](#)'. Luke Kemp spoke on an Oxford Climate Society panel with David Wallace-Wells, author of *The Uninhabitable Earth*.
- Clarissa Rios Rojas spoke on a Global Young Academy panel [COVID-19 in Latin America, perspectives of young scientists](#).
- [Webinar with Cambridge University Vice-Chancellor: AI and power](#). Professor Stephen J Toope moderated a discussion with Seán Ó hÉigeartaigh and Kanta Dihal, a Centre for the Future of Intelligence postdoctoral researcher. Together, they discussed the benefits and challenges arising from our increasing use of artificial intelligence, including its use in addressing the many threats posed by the current pandemic, but exploring more broadly the use of artificial intelligence that is integral to – and sometimes invisible in – our daily lives.
- Each month, The Existential Risk Research Assessment ([TERRA](#)) uses a unique machine-learning model to predict those publications most relevant to existential risk or global catastrophic risk. We provide these citations and abstracts as a service to aid other researchers in paper discovery.
  - 6 Recent Publications on Existential Risk ([Sep 2020 update](#))
  - 6 Recent Publications on Existential Risk ([Aug 2020 update](#))
  - 2 Recent Publications on Existential Risk ([July 2020 update](#))
  - 11 Recent Publications on Existential Risk ([June 2020 update](#))
  - 3 Recent Publications on Existential Risk ([May 2020 update](#))
  - 5 Recent Publications on Existential Risk ([April 2020 update](#))



Participants in the 'Overcoming Barriers to Cross-Cultural Cooperation in AI Ethics and Governance' workshop.

We continued our weekly work-in-progress series:

27 April	Lara Mani	<i>Results and Discussion of CSER Vision Staff Survey</i>
4 May	Natalie Jones	<i>Procedural Rights of Minority Groups and Indigenous Peoples in International Decision Making</i>
18 May	Nathan Sears	<i>Existential Security: Towards a Security Framework for the Survival of Humanity</i>
1 June		<i>Science of Global Risk updates</i>
8 June	Sabin Roman	<i>Chaotic Dynamics in the Chinese Dynasties</i>
15 June	<a href="#">Matthew Adler</a>	<i>Risk Regulation</i>
22 June	<a href="#">Andrew Futter</a>	<i>Towards a Third Nuclear Age</i>
29 June	Ellen Quigley	<i>Universal Ownership in the Age of COVID-19: Social Norms, Feedback Loops, and the Double Hermeneutic</i>
13 July	<a href="#">Tom Moynihan</a>	
20 July	Natalie Jones	<i>The SDGs and COVID-19: Reporting back from the UN High-Level Political Forum</i>
7 September	Shahar Edgerton Avin, Lara Mani, Jess Whittlestone, Tom Hobson	<i>COVID Implementation Failures project update</i>
21 September	Seth Baum	<i>The Global Catastrophic Risk Institute's recent work</i>

# 7. Publications

## Journal articles:

**Luke Kemp**, Laura Adam, **Christian R. Boehm**, Rainer Breitling, Rocco Casagrande, Malcolm Dando, Appolinaire Djikeng, **Nicholas Evans**, Richard Hammond, Kelly Hills, **Lauren Holt**, Todd Kuiken, Alemka Markotić, Piers Millett, Johnathan A Napier, Cassidy Nelson, **Seán Ó hÉigeartaigh**, Anne Osbourn, Megan Palmer, Nicola J Patron, Edward Perello, Wibool Piyawattanametha, Vanessa Restrepo-Schild, **Clarissa Rios Rojas**, **Catherine Rhodes**, Anna Roessing, Deborah Scott, Philip Shapira, Christopher Simuntala, Robert DJ Smith, **Lalitha Sundaram**, Eriko Takano, Gwyn Uttmark, **Bonnie Wintle**, Nadia B Zahra, **William Sutherland**. (2020). [Point of View: Bioengineering horizon scan 2020](#). *eLife*.

*“Horizon scanning is intended to identify the opportunities and threats associated with technological, regulatory and social change. In 2017 some of the present authors conducted a horizon scan for bioengineering (Wintle et al., 2017). Here we report the results of a new horizon scan that is based on inputs from a larger and more international group of 38 participants. The final list of 20 issues includes topics spanning from the political (the regulation of genomic data, increased philanthropic funding and malicious uses of neurochemicals) to the environmental (crops for changing climates and agricultural gene drives). The early identification of such issues is relevant to researchers, policy-makers and the wider public.”*

**Seán Ó hÉigeartaigh**, **Jess Whittlestone**, **Yang Liu**, Yi Zeng, Zhe Liu. (2020). [Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance](#). *Philosophy & Technology*.

*“Achieving the global benefits of artificial intelligence (AI) will require international cooperation on many areas of governance and ethical standards, while allowing for diverse cultural perspectives and priorities. There are many barriers to achieving this at present, including mistrust between cultures, and more practical challenges of coordinating across different locations. This paper focuses particularly on barriers to cooperation between Europe and North America on the one hand and East Asia on the other, as regions which currently have an outsized impact on the development of AI ethics and governance. We suggest that there is reason to be optimistic about achieving greater cross-cultural cooperation on AI ethics and governance. We argue that misunderstandings between cultures and regions play a more important role in undermining cross-cultural trust, relative to fundamental disagreements, than is often supposed. Even where fundamental differences exist, these may not necessarily prevent productive cross-cultural cooperation, for two reasons: (1) cooperation does not require achieving agreement on principles and standards for all areas of AI; and (2) it is sometimes possible to reach agreement on practical issues despite disagreement on more abstract values or principles. We believe that academia has a key role to play in promoting cross-cultural cooperation on AI ethics and governance, by building greater mutual understanding, and clarifying where different forms of agreement will be both necessary and possible. We make a number of recommendations for practical steps and initiatives, including translation and multilingual publication of key documents, researcher exchange programmes, and development of research agendas on cross-cultural topics.”*

**Asaf Tzachor**, **Jess Whittlestone**, **Lalitha Sundaram**, **Seán Ó hÉigeartaigh**. (2020). [Artificial intelligence in a crisis needs ethics with urgency](#). *Nature Machine Intelligence*.

*“Artificial intelligence tools can help save lives in a pandemic. However, the need to implement technological solutions rapidly raises challenging ethical issues. We need new approaches for ethics with urgency, to ensure AI can be safely and beneficially used in the COVID-19 response and beyond.”*

Jose Hernandez-Orallo, Fernando Martinez-Plumed, **Shahar Avin, Jess Whittlestone, Seán Ó hÉigeartaigh**. (2020). [AI Paradigms and AI Safety: Mapping Artefacts and Techniques to Safety Issues](#). **Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020)**.

*“AI safety often analyses a risk or safety issue, such as interruptibility, under a particular AI paradigm, such as reinforcement learning. But what is an AI paradigm and how does it affect the understanding and implications of the safety issue? Is AI safety research covering the most representative paradigms and the right combinations of paradigms with safety issues? Will current research directions in AI safety be able to anticipate more capable and powerful systems yet to come? In this paper we analyse these questions, introducing a distinction between two types of paradigms in AI: artefacts and techniques. We then use experimental data of research and media documents from AI Topics, an official publication of the AAIL, to examine how safety research is distributed across artefacts and techniques. We observe that AI safety research is not sufficiently anticipatory, and is heavily weighted towards certain research paradigms. We identify a need for AI safety to be more explicit about the artefacts and techniques for which a particular issue may be applicable, in order to identify gaps and cover a broader range of issues.”*

**John Burden, José Hernández-Orallo**. (2020). [Exploring AI Safety in Degrees: Generality, Capability and Control](#). **Proceedings of the Workshop on Artificial Intelligence Safety (SafeAI 2020) co-located with 34th AAAI Conference on Artificial Intelligence (AAAI 2020)**.

*“The landscape of AI safety is frequently explored differently by contrasting specialised AI versus general AI (or AGI), by analysing the short-term hazards of systems with limited capabilities against those more long-term risks posed by ‘superintelligence’, and by conceptualising sophisticated ways of bounding control an AI system has over its environment and itself (impact, harm to humans, self-harm, containment, etc.). In this position paper we reconsider these three aspects of AI safety as quantitative factors –generality, capability and control–, suggesting that by defining metrics for these dimensions, AI risks can be characterised and analysed more precisely. As an example, we illustrate how to define these metrics and their values for some simple agents in a toy scenario within a reinforcement learning setting.”*

**Asaf Tzachor**. (2020). [A capital theory approach should guide national sustainability policies](#). **Cambridge Journal of Science and Policy**.

*“The question of how to sustain human development in the current ecological and institutional landscape is arguably one of the utmost scientific and administratively challenging contemporary dilemmas. In response to this issue, the concept of Sustainable Development was proposed by the United Nations to inform policies for societal and human development. However, for national governments, the prevalent sustainability schemes summon more confusion than coherence. This is due to the frequent and inconsistent ways the concept of sustainability is put into practice, and consequently, difficulties in measuring and managing sustainability. The ability to evaluate how sustainable public projects are, will remain deficient if sustainability remains a notion open for interpretation. This perspective article maintains that the capital theory approach to sustainability stands out as the most rigorous framework on the topic. The capital theory is a state-centric system of ideas where national governments oversee a portfolio of capital stocks of four families: natural capital, economic capital, human capital, and social capital. It is the duty of the government to act on the capital flow between different stocks to allow sustainable long-term development. This perspective paper underscores several envisaged gains from the application of the capital theory approach in public policy. Considering these expected gains, policy makers should be encouraged to experiment with the approach.”*

**Asaf Tzachor**. (2020). [Famine Dynamics: The Self-undermining Structures of the Global Food System](#). **Global Relations Forum Young Academics Program Analysis Paper Series No.8**.

*“Civilization has steered itself onto a vicious spiral. The modern system of agriculture, upon which global food security hinges, devours the planet’s scarce supplies of fertile lands, fresh water, productive fisheries and forest ecosystems. 821 million lives hang in the balance, already suffering famine and all forms of malnutrition, while early signs of an even larger catastrophe begin to*

*transpire. Instead of perpetuating self-undermining dynamics, the international science and policy communities should radically reform the methods of food production and provision. Systems thinking should lend insight.”*

**Simon Beard**, Thomas Rowe, James Fox. (2020). [Existential Risk Assessment: A reply to Baum](#). **Futures**.

*“We welcome Seth Baum's reply to our paper. While we are in broad agreement with him on the range of topics covered, this particular field of research remains very young and undeveloped and we think that there are many points on which further reflection is needed. We briefly discuss three: the normative aspects of terms like 'existential catastrophe,' the opportunities for low hanging fruit in method selection and application and the importance of context when making probability claims.”*

## Reports and preprints:

**Ellen Quigley**, Emily Budgen, Anthony Odgers. (2020). [Divestment: Advantages and Disadvantages for the University of Cambridge](#). **University of Cambridge**.

*“The University of Cambridge holds assets of approximately £3.5 billion, the largest university endowment in Europe. Within the University there is broad agreement about the urgent need to reduce carbon emissions. However, whether full divestment of University funds from fossil fuel assets is the best way to make that happen has been the subject of intense debate. Based on a review of the academic literature, interviews and focus groups with relevant stakeholders inside and outside the University, records of University and college discussions, and some further primary data collection, this report explores the advantages and disadvantages of a policy of fossil fuel divestment across its moral, social, political, reputational, and financial dimensions, ending with a summary of costed divestment scenarios for the University.”*

**Haydn Belfield**, **Amritha Jayanti**, **Shahar Avin**. (2020). [Premature Procurement](#). **UK Parliament Defence Committee, Written Evidence - Defence industrial policy: procurement and prosperity**.

*“Our submission is on the risks of militaries procuring AI defence systems ‘prematurely’ – before they are fully derisked, safe and secure. We therefore make the following recommendations to protect against premature and/or unsafe procurement and deployment of ML-based systems:*

- Improve systemic risk assessment in defence procurement.*
- Ensure clear lines of responsibility so that senior officials can be held responsible for errors caused in the procurement chain and are therefore incentivised to reduce them;*
- Acknowledge potential shifts in international standards for autonomous systems, and build flexible procurement standards accordingly.*
- Update the MoD’s definition of lethal autonomous weapons - the Integrated Security, Defence and Foreign Policy Review provides an excellent opportunity to bring the UK in line with its allies.”*

**Haydn Belfield**, José Hernández-Orallo, **Seán Ó hÉigearthaigh**, Matthijs M. Maas, **Alexa Hagerty**, **Jess Whittlestone**. (2020). [Response to the European Commission’s consultation on AI](#). **European Commission**.

*“The submission mainly focuses on mandatory conformity assessments for high-risk AI applications carried out by independent testing centres. Our key recommendation is to keep this proposed framework and not water it down. In this submission we 1) support the Commission’s proposed structure, defend this approach on technical, policy, practical, and ethical grounds, and offer some considerations for future extensions; and 2) offer some specific recommendations for the mandatory requirements.”*

**Asaf Tzachor**. (2020). [Artificial intelligence for agricultural supply chain risk management: Constraints and potentials](#). **CGIAR Big Data Platform**.

*“Supply chains of staple crops, in developed and developing regions, are vulnerable to an array of disturbances and disruptions. These include biotic, abiotic and institutional risk factors. Artificial intelligence (AI) systems have the potential to mitigate some of these vulnerabilities across supply chains, and thereby improve the state of global food security. However, the particular properties of each supply chain phase, from “the farm to the fork,” might suggest that some phases are*

more vulnerable to risks than others. Furthermore, the social circumstances and technological environment of each phase may indicate that several phases of the supply chains will be more receptive to AI adoption and deployment than others. This research paper seeks to test these assumptions to inform the integration of AI in agricultural supply chains. It employs a supply chain risk management approach (SCRM) and draws on a mix-methods research design. In the qualitative component of the research, interviews are conducted with agricultural supply chain and food security experts from the Food and Agricultural Organization of the UN (FAO), the World Bank, CGIAR, the World Food Program (WFP) and the University of Cambridge. In the quantitative component of the paper, seventy-two scientists and researchers in the domains of digital agriculture, big data in agriculture and agricultural supply chains are surveyed. The survey is used to generate assessments of the vulnerability of different phases of supply chains to biotic, abiotic and institutional risks, and the ease of AI adoption and deployment in these phases. The findings show that respondents expect the vulnerability to risks of all but one supply chain phases to increase over the next ten years. Importantly, where the integration of AI systems will be most desirable, in highly vulnerable supply chain phases in developing countries, the potential for AI integration is likely to be limited. To the best of our knowledge, the methodical examination of AI through the prism of agricultural SCRM, drawing on expert insights, has never been conducted. This paper carries out a first assessment of this kind and provides preliminary prioritizations to benefit agricultural SCRM as well as to guide further research on AI for global food security.”

Nick Bostrom, **Haydn Belfield**, Sam Hilton. (2020). [UK ARPA – Key Recommendations](#). **UK Parliament Science & Technology Committee, Written Evidence - A new UK funding agency.**

*Recommends focussing on high-impact risk research and global leadership in emerging technologies. Also makes recommendations relating to structure and culture, mechanisms to identity and start projects, testing and evaluation, recommended tools and hiring.*

Sam Hilton, **Haydn Belfield**. (Not yet online). The UK Government’s risk management system. **UK Parliament Science & Technology Committee, Written Evidence - UK Science, Research and Technology Capability and Influence in Global Disease Outbreaks.**

*Recommends several improvements to the UK Government’s risk management system.*

**Simon Beard**, Kate Brierton, Paul Gilbert & Felicia Huppert. (2020). [Public Mental Health and COVID-19: a compassion based approach to recovery and resilience](#). **Association of Liberal Democrat Engineers and Scientists Policy Brief.**

*“The mental health impacts of COVID-19 are expected to be very significant and could be the single greatest health burden caused by this pandemic in the long term. However, at present they are not being assessed at the population level, while not enough is being done to study these impacts and how to respond to them. This briefing note explores the issues raised around mental health in the context of the COVID-19 pandemic, and encourages a genuine and robust compassionate response to COVID-19 which also requires ambitious and creative thinking in response to many social and economic problems.”*

**Simon Beard** & James Belchamber. (2020). [AI and Data Governance issues in responding to COVID-19](#). **Association of Liberal Democrat Engineers and Scientists Policy Brief.**

*“AI and data-driven technologies have the potential to support pandemic prevention and control at many stages, including improving response and recovery for COVID-19. As well as having direct benefits, the increased use of AI and data technologies could increase social understanding of their benefits and public trust in them. However, the speed at which new technologies are being considered and implemented raises a number of concerns. This briefing note discusses these concerns, including those specific to the NHS contact tracing app and also discusses some of the broader governance issues surrounding AI ethics.”*



## 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](mailto:daisy.ayre@admin.cam.ac.uk)

+44 (0)1223 764117

[www.cam.ac.uk/yourscambridge](http://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](mailto:hb492@cam.ac.uk)

+44 (0)770 9676 355

[www.cser.ac.uk](http://www.cser.ac.uk)



**UNIVERSITY OF  
CAMBRIDGE**

**Dear World...**

**Yours, Cambridge**

The campaign for the University  
and Colleges of Cambridge