

# Risk-sensitive innovation: leveraging interactions between technologies to navigate technology risks

Jonas B. Sandbrink<sup>1,2,\*</sup>, Hamish Hobbs<sup>1</sup>, Jacob L. Swett<sup>3</sup>, Allan Dafoe<sup>4,5</sup>, Anders Sandberg<sup>1</sup>

<sup>1</sup>Future of Humanity Institute, University of Oxford, Trajan House, Mill Street, Oxford, Oxfordshire OX2 0DJ, United Kingdom

<sup>2</sup>Nuffield Department of Medicine, University of Oxford, Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7FZ, United Kingdom

<sup>3</sup>Blueprint Biosecurity, 1333 New Hampshire Avenue NW fl 2, Washington, DC 20036, United States

<sup>4</sup>Google DeepMind, 14-18 Handyside Street, London N1C4DN, United Kingdom

<sup>5</sup>The Centre for the Governance of AI, Trajan House, Mill Street, Oxford OX2 0DJ, United Kingdom

\*Corresponding author. Future of Humanity Institute, University of Oxford, Trajan House, Mill Street, Oxford, Oxfordshire OX2 0DJ, United Kingdom.  
Email: [jonas.sandbrink@trinity.ox.ac.uk](mailto:jonas.sandbrink@trinity.ox.ac.uk)

The advancement of low-emission technologies to curb climate change demonstrates how strategically managing the timing of technological developments can mitigate negative societal outcomes. We propose that a similar approach, which we term ‘risk-sensitive innovation’, can also address risks from emerging technologies like synthetic biology and artificial intelligence. Risk-sensitive innovation (RSI) leverages interactions between technologies to reduce specific risks across a technology portfolio. For instance, it may be beneficial to delay or halt risk-increasing technologies and preferentially advance risk-reducing defensive, safety, or substitute technologies. To be of practical use, RSI requires the ability to anticipate or identify impacts and intervene in the relative timing of technological advancements. Our analysis suggests that both are sometimes viable. Consideration of RSI may inform government research funding and technology regulation, as well as philanthropic research and development funders and corporate social responsibility measures.

**Keywords:** technology policy; innovation governance; differential technological development; risk management; responsible innovation; emerging technology; technology foresight; artificial intelligence; synthetic biology.

## 1. Introduction

The global response to climate change has highlighted that a diverse set of actors can play a role in shifting the direction of innovation to manage a particular risk. Decarbonization interventions have demonstrated that it is possible to actively restructure the technology portfolio at a societal scale and preferentially advance technologies to mitigate climate risk (Mowery, Nelson and Martin 2010). Certain technologies, like carbon capture, decrease the negative impacts of complementary high-emission technologies. Other technologies, like clean energy sources, serve as low-risk substitutes for high-emission technologies. This portfolio-based approach may be useful more generally for mitigating other risks from emerging technologies.

Artificial intelligence (AI) provides another example of the potential to direct innovation to manage technological risks. Rapid developments in AI, while producing many benefits, are also creating or exacerbating certain risks. Risks from AI systems include AI-enabled disinformation, hacking, and discrimination (Bengio 2024). Directing innovation towards certain risk-reducing technologies is a promising approach to manage these risks. Technologies for watermarking, tracing, or detecting the provenance of authentic or AI-generated content can help mitigate disinformation risks (Partnership on AI 2023). AI-enabled cybersecurity tools can ensure that AI-enabled hacking does not outpace cybersecurity capabilities (US Cybersecurity & Infrastructure Security Agency 2024).

Interpretable and explainable AI techniques can ensure that bias or discrimination is identified and addressed (Linardatos, Papastefanopoulos, and Kotsiantis 2021). Governments and other relevant actors can support innovation in each of these risk-reducing aspects of AI innovation to help manage the risks associated with broader progress in AI.

These same types of interactions between risk-increasing and risk-reducing innovations exist for a wide variety of technologies and can be leveraged to mitigate technology risks. DNA synthesis screening, vaccines, and diagnostic technologies can mitigate the dual-use risks of synthetic biology. Advances in technologies for transparency and verification can support international arms control measures and mitigate risks from weapons technologies (Vaynman 2021). Antinausea drugs can mitigate negative impacts of chemotherapy (Jordan, Jahn, and Aapro 2015). Seatbelts and airbags can mitigate risks from vehicle crashes (Cummings et al. 2002).

### 1.1 Risk-sensitive innovation

In this paper, we propose an innovation governance consideration called ‘risk-sensitive innovation’, which explicitly leverages risk-reducing interactions between technologies through their relative timing. Governments and other relevant actors may consider technology interactions to preferentially advance technologies that reduce specific risks. This approach is not a substitute for broader responsible innovation governance; rather, it serves as a perspective that can help

generate governance options, which can then undergo more holistic consideration.

In the first half of this article, we introduce the risk-sensitive innovation (RSI) approach and provide definitions and context. First, we analyse the existing responsible innovation efforts (Section 1.2). Then, we define RSI and suggest its application for risk mitigation (Section 2). We then outline how we define risk-increasing and risk-reducing technologies relative to specific risk (Section 3). In the second part of this article, we examine when it is possible to anticipate or identify the impacts of technologies (Section 4) and how different actors can delay or preferentially advance specific technologies (Section 5).

## 1.2 A history of shaping technology

Attempts to control the pace and direction of technological progress are not new. In the 1960s, interest in the societal impacts of technologies gave rise to the field of technology assessment, which sought to analyse the short- and long-term consequences of the applications of specific new technologies (Banta 2009). A crucial goal of technology assessment was to shape natural science research and engineering efforts based on broader societal inputs (Guston and Sarewitz 2002). For instance, starting in the 1980s, the Danish Board of Technology held consensus conferences to consult the public on technology developments (Fisher, Mahajan, and Mitcham 2006).

A variety of efforts in the 1990s and 2000s explored the assessment and shaping of technology in a societally beneficial way. Constructive technology assessment built on other technology assessment approaches to not only assess but also modulate ongoing technology development, through strategies like technology forcing and strategic niche management (Schot and Rip 1997). Value-sensitive design creates a system for considering human values in a given technological design process (Friedman et al. 2013). The study of ethical, legal, and social aspects (ELSA) of science and technology has been particularly influential in considering societal issues alongside research projects (Zwart, Landeweerd, and van Rooij 2014). Research on ELSA has often accompanied major life sciences research investments and has tracked those specific research areas, such as in the case of the US National Institutes of Health's Human Genome Project in 1990 (Zwart, Landeweerd, and van Rooij 2014). Similarly, the US National Nanotechnology Initiative reinvigorated the field of technology assessment in the early 2000s (Guston 2014).

In the 2010s, the concept of responsible research and innovation (RRI) began to receive substantial attention and became the centrepiece of European science and innovation policies (Owen, Macnaghten, and Stilgoe 2012). Through responsible innovation, policymakers and academics seek to increase the extent to which research and innovation produce beneficial outcomes for society. Compared to ELSA, the framing and ambition of RRI have moved more towards collaboration with industry and a focus on the socio-economic benefits of technology (Zwart, Landeweerd, and van Rooij 2014).

Socio-technical transitions research is another area that considers the impacts of technologies and their interplay with society. Out of this field, the multi-level perspective (MLP) has emerged as a core framework for studying sustainability transitions (Geels 2019). The MLP investigates changes

at different levels (individual technology niches, governance regimes, and the overall socio-technological landscape) that lead to socio-technical transitions, particularly in relation to sustainability.

Climate change and the need to decarbonize the economy have driven technologies and policies that leverage interactions across the technology portfolio to improve societal and environmental outcomes. The main goal of these efforts has been the preferential advancement of low-carbon technologies and the substitution of fossil fuels (Farmer et al. 2019). To this end, interventions such as carbon pricing and feed-in tariffs for photovoltaics have been used (Haegel et al. 2017; Green 2021). These policies have incorporated technology assessment and modulation approaches by using strategies like technology forcing and strategic niche management. Acemoglu (2002) developed an economic theory of directed technical change, which has also been applied to the decarbonization of the economy (Aghion et al. 2016).

Arms control has been a key tool by which states have shaped the development of technology. In Cold War literature, Schelling and Halperin link arms control closely to strategic stability, with a focus on agreements between states to limit the number of weapons systems with the goal of preventing arms race dynamics and mitigating inadvertent escalation by promoting transparency (Schelling and Halperin 1961). Traditionally, arms control features agreements between consenting parties to restrict the development or deployment of weapons technologies (Croft 1996). More recently, scholars have argued for broadening arms control beyond legally binding treaties (Kreps 2018) and the historical focus on counting weapons systems to a more holistic set of policy instruments (Gallagher 2015).

In this paper, we propose a concept of 'RSI' to consider governance options relating to particular risks. At the heart of this concept lies the consideration that the relative timing of technological developments and their interactions shape their impact relative to a risk. This approach seeks to complement existing innovation governance approaches by providing additional considerations to generate risk management options.

## 2. Defining RSI

### 2.1 Definition

Innovation can not only contribute to a range of new or evolving risks but can also help to mitigate these risks. We propose RSI as a concept to help relevant actors (such as governments) manage these dual impacts of innovation on a given risk. Concretely, this concept calls on relevant actors seeking to mitigate a specific risk to consider preferentially advancing technologies that are risk-reducing relative to that specific risk and delay risk-increasing technologies:

Risk-sensitive innovation (RSI): To mitigate a specific risk, consider options to leverage risk-reducing interactions between technologies by affecting their relative timing.

RSI is characterized by its focus on the interactions between technologies. Advancing technologies that reduce a specific risk before or soon after the advent of risk-increasing technologies would improve societal outcomes. For example, printing technologies and anti-counterfeiting technologies, cars and seatbelts, and anticoagulants and anticoagulant

reversal agents all share simple interactions that can be leveraged to reduce specific technology risks.

Relevant actors, such as government research funding agencies could consider RSI for specific high-priority technology risks (such as those targeted by existing mission-oriented innovation policies) to guide their overall grantmaking strategy and prioritize between different innovation objectives. This would not substitute for other approaches to balancing the complex risks and benefits of technologies, such as broader responsible research and innovation practices, but would provide agencies with an additional perspective to help generate policy options.

RSI is still possible in the early or even late stages of the diffusion of a technology. The risk of global warming was not appreciated until long after the diffusion of fossil fuel technologies, but interventions to speed the adoption of clean energy technologies have nonetheless been central to mitigating the resulting harms.

This perspective can be seen as a way to implement the concept of differential technological development, which was first articulated by [Bostrom \(2014\)](#). Bostrom introduces 'differential technological development' to explore how the timing of advanced AI relative to other technologies may impact associated risks. We propose an application of this approach within the broader context of innovation governance.

## 2.2 Application to catastrophic risk mitigation

It is already well understood that preventing catastrophic risks from future technologies should be an important priority for responsible innovation efforts. As the power of technologies increases, they can pose catastrophic risks. Advances in synthetic biology and virology create a risk for engineered pandemics much worse than COVID-19 ([Schoch-Spana et al. 2017](#)). While AI promises benefits across many aspects of society, AI-enabled disinformation attacks are already being used to sow disinformation and destabilize democracies and pose risks of destabilizing military balances of power ([Horowitz 2018](#); [National Security Commission on Artificial Intelligence 2021](#)). This paper explores the potential of RSI to reduce such risks.

## 3. Risk-reducing interactions between technologies

In this section, we discuss how we define risk-increasing and risk-reducing technologies relative to a specific risk. This helps to illustrate the importance of the sequence, which different technologies are developed for mitigating risks. These categories serve as a starting point to illustrate opportunities and practical applications of RSI.

Risk itself is a highly contested concept, and perceptions of risk are influenced by a range of social and cultural factors ([Dake 1991](#)). As a result, it is not possible to assess risk without discussing underlying worldviews and values. To address this, contemporary risk management approaches seek to integrate evidence about the world with value-based evaluation of risks ([Aven 2016](#)). These approaches are continually being iterated in an attempt to address the complex interplay between risk assessment, worldviews, and values. While it is not possible to remove worldviews and value decisions from risk assessment, attempts can be made to ensure

that these values are acknowledged and made explicit ([Wandall 2004](#)). More generally, there are a range of theoretical approaches to best identify values that are incorporated into responsible innovation efforts ([Boenink and Kudina 2020](#)). Given these complexities, we do not seek to prescribe a specific risk identification methodology or the best approaches to agree on guiding values. Instead, we suggest that once risks have been identified through larger holistic processes, RSI may provide a useful consideration to generate options to address these risks. These options will then also need to be considered as part of a value-based process, including broader impacts beyond the specific risk being considered.

In each case below, technologies are defined with respect to a specific risk. For the purposes of this discussion, the risk is one that has been identified and agreed as a priority for mitigation through value-based processes that are beyond the scope of this article.

### 3.1 Risk-increasing technologies

'Risk-increasing technologies' are defined relative to a specific risk. This risk may be of a sudden onset event (such as a cyber-attack), gradual onset (such as a pandemic), or be steady-state risk (such as poverty). High-carbon emission technologies insidiously drive global warming and cause harmful air pollution, while the development of nuclear weapons has created the threat of nuclear war and subsequent nuclear winter.

Some risk-increasing technologies, such as landmines, are purely offensive and have no civilian use. To help mitigate harm from these risk-increasing technologies, we can invest in mitigations such as technologies to detect and remove landmines ([Dorn 2019](#)). Most risk-increasing technologies feature upsides that drive their development and adoption despite possible negative societal impacts. Certain technologies may, for instance, be 'dual-use', featuring both beneficial and harmful applications. This dual-use dilemma is particularly pronounced not only for synthetic biology ([Atlas and Dando 2006](#)) and AI ([Brundage et al. 2018](#)) but it also applies to many other technologies. For instance, additive manufacturing not only has many beneficial applications but also facilitates building nuclear weapons ([Volpe 2019](#)).

Other technologies may be able to reduce the risks posed by these risk-increasing, frequently dual-use technologies or achieve the same benefits without creating risks.

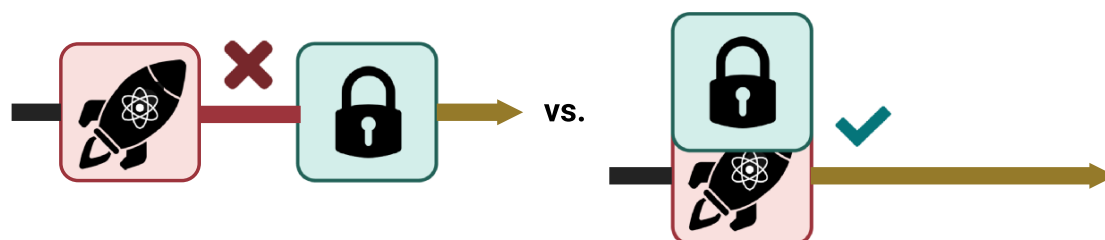
### 3.2 Risk-reducing technologies

**3.2.1 Safety technologies.** 'Safety technologies' reduce the severity or likelihood of a particular risk by modifying risk-increasing technologies. For instance, carbon capture and sequestration may reduce the emissions of coal-fired power plants ([Chu 2009](#)). Seat belts and airbags make driving a car safer for the passengers ([Cummins et al. 2011](#)).

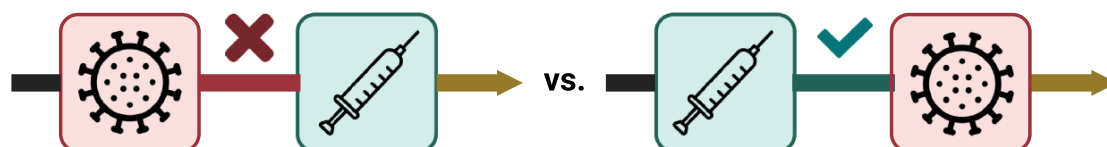
For dual-use technologies, safety technologies may reduce the possibility of accidental or deliberate misuse. The development of electronic locks for nuclear weapons, as permissive action links (PALs), in the 1960s has reduced the risk of accidental or unauthorized launches ([Caldwell 1987](#)). DNA synthesis screening technology may prevent the harmful application of DNA synthesis machines to generate genetic materials for the illicit creation of pathogens ([Esvelt 2018](#)).

Minimizing the time between developing a risk-increasing technology and a relevant safety technology can mitigate expected societal impacts ([Fig. 1a](#)). For instance, developing

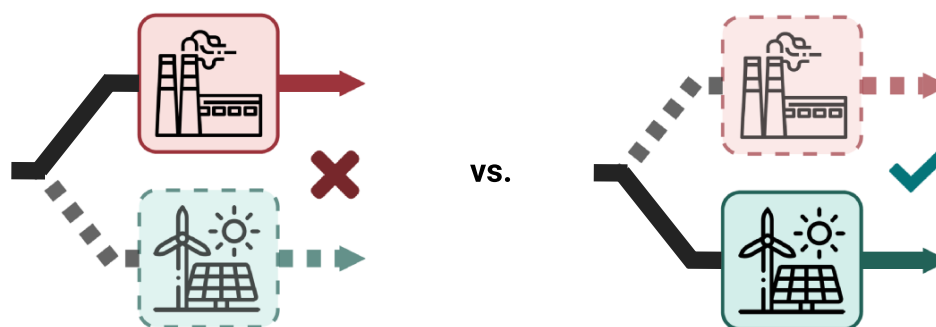
a) **Safety technologies** sooner relative to risk-increasing technologies



b) **Defensive technologies** sooner relative to risk-increasing technologies



c) **Substitute technologies** instead of risk-increasing technologies



**Figure 1.** Mechanisms by which risk-sensitive innovation can reduce negative societal impacts. (a) Developing safety technologies before or soon after risk-increasing technologies can reduce negative societal impacts; e.g. electronic locks can prevent the unauthorized use of nuclear weapons. (b) Developing defensive technologies before or soon after risk-increasing technologies reduce a window of vulnerability; e.g. society should first develop ways to prevent a pandemic from a particular virus, for instance, by vaccinating the population, before disseminating its blueprint for its synthesis for biotechnology applications. (c) Low-risk substitute technologies can replace risk-increasing technologies or make their development unnecessary; e.g. renewable energy sources are replacing the environmentally harmful use of coal for energy production.

PALs for nuclear weapons as part of the Manhattan Project instead of two decades later would have reduced the window of vulnerability for catastrophic misuse of nuclear weapons over this period.

**3.2.2 Defensive technologies.** ‘Defensive technologies’ decrease the severity or likelihood of a particular risk from a risk-increasing technology or set of technologies without modifying these technologies. For example, mRNA vaccines, a novel vaccine platform technology that can be quickly adapted to different pathogens, significantly contributed to curbing the COVID-19 pandemic (Sandbrink and Shattock 2020). mRNA vaccines and similar vaccine platforms will be crucial to reduce the societal impact of future pandemics of any origin—whether natural or caused by the accidental or deliberate misuse of synthetic biology.

Advancing defensive technologies before relevant risk-increasing technologies prevent a window of vulnerability and thus leads to better expected societal outcomes (Fig. 1b). Concretely, if government research funding agencies prioritize pandemic prevention technologies like pathogen detection and platform vaccines before advancing research that

risks increasing the ability to create pandemic pathogens, this could lead to less societal harm from accidental or deliberate pandemics.

**3.2.3 Substitute and low-risk alternative technologies.** ‘Substitute technologies’ achieve similar benefits as risk-increasing technology while contributing less to specific risks. One prominent example is clean energy technologies, like wind turbines or photovoltaics, which can replace carbon-intensive fossil fuels. Another example was the development of substitute technologies for the phase-out of ozone-depleting substances, estimated to have otherwise caused two million additional cases of skin cancer each year in 2030 (Van Dijk et al. 2013; McKenzie et al. 2019). These substitutions are by no means simple, requiring a weighing of a complex set of other risks and benefits, but consideration of differential technological development can highlight them as options for deeper consideration.

There may also be substitutes for dual-use technologies that offer the same benefits with less potential for misuse. For instance, instead of learning to engineer and fine-tune viruses for the delivery of vaccines and therapeutics, non-viral

delivery methods could be advanced (Sandbrink *et al.* 2021). Where viral delivery methods feature substantial advantages, such as for specific gene therapy applications, non-heritable methods of viral modification should be preferentially investigated (Sandbrink and Koblentz 2022). These non-heritable methods may provide the same benefits but cannot be used to effectively enhance transmissible viruses.

Preferential investigation of low-risk technologies to solve a given challenge would lead to less risk across our technology portfolio (Fig. 1c). When risk-increasing technologies are already in use, the advancement of substitute technologies may help to reduce negative societal impacts.

### 3.3 Shaping individual technologies

Risk-sensitive innovation not only focuses on leveraging interactions between different technologies but it is also compatible with related efforts to shape individual technologies. While the framing of differential technological development focuses on technology interactions and timing, this does not imply that particular technologies and associated design choices are inevitable. Instead, different design choices for a given technology can be integrated into a RSI analysis by considering these design choices as constituting potential substitutable technologies. For instance, it is beneficial to substitute away from refrigeration systems that use ozone-depleting substances to refrigeration systems that do not. Indeed, attempts to implement risk-sensitive innovation may often benefit from this approach. This is especially true for general purpose and platform technologies (Bresnahan and Trajtenberg 1995), which may have both defensive and offensive applications. Some risk-reducing technologies may feature dual-use potential themselves, highlighting the need for holistic assessment beyond a single risk (Sandbrink and Koblentz 2022). For instance, metagenomic sequencing-based pathogen detection will likely be a crucial defensive technology for preventing pandemics (Nucleic Acid Observatory Consortium 2021). However, if developed in a way that does not prevent privacy infringements, genomic sequencing may be misused for identifying and tracking individuals or ethnic populations.

Responsible innovation scholars have proposed various strategies for shaping individual technologies through technology assessment, eliciting societal inputs, and concrete interventions (Schot and Rip 1997; Guston and Sarewitz 2002; Friedman *et al.* 2013; Stilgoe, Owen and Macnaghten 2013). One mechanism applicable to both risk-increasing and risk-reducing technology review is stage-gating. Stage-gating breaks up a larger project into a series of steps with associated decision points (Stilgoe, Owen, and Macnaghten 2013). Thus, stage-gating may help identify risks and opportunities for safety technologies or other forms of interventions to enhance defence bias. Stage-gating was applied to the UK geoengineering SPICE experiment at different stages of preparation, leading initially to its postponement and later its cancellation in 2011 (Stilgoe, Owen, and Macnaghten 2013). Methods and practices similar to stage-gating need to be advanced to assess the dual-use potential of individual technologies and shape their development.

### 3.4 RSI and arms control

Arms control is one field in which the shaping of innovation trajectories to mitigate risks is already a core focus. Arms control seeks to prevent mutually harmful, inadvertent escalation

associated with weapons technologies (Schelling and Halperin 1961). Risk-informed innovation can be complementary to arms control. For example, ‘lack of transparency’ is one key factor for inadvertent escalation. In this case, an RSI concept would suggest the preferential advancement of technologies that increase transparency.

Certain types of RSI could be complementary to arms control agreements. For instance, RSI can help to mitigate the risk of inadvertent escalation and accidents associated with weapons technologies. As discussed in Section 3.2.1, in the case of nuclear weapons, PALs are a safety technology that may have reduced the risk of unintended launches. Technologies that enable verification, such as satellites and other novel sensors, may also be seen as safety or defensive technologies to be advanced through RSI measures, especially where they open the possibility for new arms control agreements on the basis of improved verification effectiveness. While the Biological Weapons Convention upholds international norms against state development and use of biological weapons, the advancement of safety technologies like DNA synthesis screening can prevent terrorists from acquiring and misusing similar technologies. Furthermore, both arms control and RSI approaches seem important for mitigating risks from AI. The Secretary-General of the United Nations (UNODA 2024) has called for a ban on lethal autonomous weapons systems and experts have even suggested the creation of an IAEA-like organization for global governance of general-purpose AI (Ho *et al.* 2023; OpenAI 2023). At the same time, RSI offers ways to enable the governance of AI systems and foster transparency, for instance through advancing evaluation research for dangerous capabilities of AI systems (Shevlane *et al.* 2023) and creating technologies for on-chip monitoring and verification of computational activity (Aarne, Fist, and Whithers 2024).

RSI can also serve as an alternative where arms control agreements are difficult. This may in particular be the case for dual-use technologies, which can be used as weapons but also have many civilian uses and may be very diffusely spread (Vaynman and Volpe 2023). In the case of synthetic biology and biological weapons, while Biological Weapons Convention effectively bans biological weapons, states parties were not able to agree on a verification regime due to the dual-use nature of biological research and its diffuse integration into civilian industry (Koblentz 2009). In this instance, RSI like defensive vaccine technologies or safety technologies like DNA synthesis screening can help to reduce risks. In the case of cyber technologies, the lack of an arms control agreement that limits state use of these systems may be partially offset through RSI in the form of cyber security systems and infrastructure resilient to cyber-attacks. Lastly, scholars have noted that rising geopolitical tensions (Wisotzki and Kühn 2021) and the steadily increasing number of technologies that could threaten strategic stability (Williams 2019) may make bilateral and multilateral arms control agreements more difficult. RSI approaches, many of which can be implemented unilaterally, may provide an alternative pathway to at least partially addressing risks.

## 4. Anticipating negative societal impacts

For consideration of RSI to be useful, it must be viable in practice. This requires the ability to: (1) anticipate or observe

how technologies might interact to produce a particular risk, and (2) intervene in the relative timing of technology development. The possibility of anticipating technology impacts is the subject of this section, while the question of intervention is the subject of [Section 5](#). Both of these requirements inherently involve making complex and highly contestable decisions about the values guiding technological development. Governments, NGOs, and firms face ongoing fundamental challenges in ensuring these decisions are reflexive, responsive, and inclusive ([Stilgoe, Owen, and Macnaghten 2013](#)). Differential technological development provides just one framing which can help to highlight where value-based choices must be made, raising governance options that then require value-based decisions. This analysis must be paired with legitimate and inclusive processes to arrive at governance responses.

#### 4.1 When is anticipation possible?

Collingridge's dilemma of control sums up the challenge of anticipating the impacts of technologies: 'When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult, and time-consuming' ([Collingridge 1980](#)). However, the history of technological development suggests that this dilemma is not insurmountable. The following sections outline the conditions under which anticipation of the impact of technologies may be possible.

There are notable historical examples of pioneers in a field anticipating technological impacts. One of the most important historical examples of technology anticipation is that of Leo Szilard ([Lanouette 1992](#)). After discovering the possibility of a nuclear chain reaction in 1933, Szilard anticipated the possibility of a nuclear bomb. Initially, Szilard kept this knowledge hidden. However, in 1939, Szilard's fear of Germany developing nuclear weapons drove him to write about the possibility of the atomic bomb to US President Franklin D. Roosevelt. This led to the commencement of the Manhattan Project 2 years later. In the wake of the German surrender, Szilard unsuccessfully worked to prevent the use of the bomb in Japan and the subsequent nuclear arms race. Szilard's story highlights how researchers may anticipate the potential future impacts of technologies. At the same time, it highlights the complex values decisions that are required to act upon this anticipatory intelligence.

However, actionable anticipation of societal impacts of technologies is also possible without reliance upon a limited set of technology developers. Actionable anticipation may generally be easier: (1) when risks are linked to defining features of a technology, (2) when technology improvements are incremental, or (3) the slower or more controlled technology diffusion is.

**4.1.1 Risks linked to defining features.** While Szilard could not be confident that a bomb based on the nuclear chain reaction would be feasible, he could predict that if such a weapon was produced, it would be far more energy dense than conventional weapons and thus capable of producing larger explosions. Thus, negative societal impacts of a technology can be anticipated when they are tightly linked to a defining feature of the technology—even if it is radically novel. For example, while we may not know the exact mechanism by which a novel anticoagulant medication will function, we can reliably anticipate that it will increase bleeding risk. Similarly,

reviewers of funding proposals can use the goals of scientific experiments to identify associated risks. For instance, gain-of-function experiments to enhance the transmissibility of potential pandemic pathogens can be anticipated to create risks of accidents and misuse ([Duprex et al. 2015](#)). As biosecurity risks are commonly linked to defining technology features, the World Health Organization was able to map emerging biotechnologies with dual-use potential ([World Health Organization 2021](#)).

Attempts of modular innovation may have relatively predictable impacts because they target an established dominant design and simply change one of its components. For example, the development of DNA benchtop synthesis machines will predictably conserve the core DNA synthesis function of industrial DNA synthesizers but increase distributed accessibility and portability. Distributed access to synthetic DNA features significant upsides for life sciences research but can also be anticipated to provide challenges for preventing the illicit synthesis of pathogens.

**4.1.2 Incremental improvements.** Based on experiences with precursor technologies, the impacts of incrementally improved technologies may be relatively straightforward to assess. The development of increasingly powerful natural language processing machine learning models can be expected to translate into growing usefulness for scams or disinformation campaigns ([Caldwell et al. 2020](#); [Seeger et al., 2020](#)).

For incremental technology improvements, it is often possible to observe trends for how improvements affect functionality. The most common example is Moore's Law, which accurately predicted that the number of transistors on a microprocessor chip would double every two years since the 1960s ([Schaller 1997](#)). Moore's Law has informed industrial strategy in the semiconductor industry since the 1990s ([Waldrop 2016](#)). DNA sequencing costs have decreased dramatically, at a greater rate than Moore's Law. While sequencing a whole-human genome cost \$10 million USD in 2007, in 2021, it was less than \$1,000 USD ([Wetterstrand, 2022](#)). With the fall in sequencing costs, we can anticipate the use of not only sequencing for personalized medicine ([Brittain, Scott, and Thomas 2017](#)) but also increasing accessibility by less well-resourced actors to use for tracking and identifying ethnic groups or individuals. Scaling laws for advanced AI systems, such as large language models, can predict the performance of future systems based on the proposed size of the AI model and the amount of computational power and data used to train it ([Kaplan et al. 2020](#)). These scaling laws have been sufficiently reliable to inform strategic decisions of leading AI companies.

**4.1.3 Before diffusion of a technology.** The window between the invention and widespread adoption of a novel technology may constitute a sensitive intervention point ([Farmer et al. 2019](#)) for mitigating the negative societal impacts of novel technologies. Such an intervention may include advancing complementary risk-reducing technologies, particularly safety technologies. For instance, while it would have been useful to have PALs with the very first nuclear weapon in 1945, it would have been almost as good to develop them before the manufacturing of additional bombs or their proliferation to other countries. This would have been possible given a single actor, the US Government,

initially developed the technology, thus controlling initial diffusion. The USA could have shared PAL technology as soon as other countries attained nuclear bombs. Indeed, 20 years later, once it had developed PALs, the US Government deliberately leaked related information to the Soviet Union (Nye 1987).

Even uncontrolled diffusion of novel technologies often takes decades and thus offers a window of opportunity to assess impacts and intervene. Potential negative impacts of technologies often correlate with their diffusion and can thus also take a long time to peak. In car safety, for example, the Ford Model T was available from 1908, but it was not until 1937 that road fatalities per capita peaked (Federal Highway Administration, Office of Highway Information Management 2009). Similarly, more than half of CO<sub>2</sub> emissions have occurred after 1988, long after technologies reliant upon fossil fuel energy sources were developed during the industrial revolution (Our World in Data 2022). Of course, as the production and consumption of a technology increases, so does inertia in its trajectory. It becomes integrated into society and different actors become invested in its current uses and its future trajectory. Nevertheless, technology diffusion is generally a gradual process and no strict lock-ins occur at the moment of invention or commercialization.

Thus, the window between invention and diffusion of a technology may be useful for reflecting on and studying risks, including offering an opportunity for a moratorium to assess impacts and develop safety technologies (see Section 5.1.2 for discussion of moratoria). In practice, this could include deliberate investment by governments to identify early impacts of novel technologies.

**4.1.4 After diffusion of a technology.** RSI is also still possible after the diffusion of a technology, once downside risks have been identified. Interventions to advance renewable energy sources and electric cars demonstrate the possibility of reducing negative impacts despite the existing widespread use of high emissions technologies. If governments had a specific policy and regular practice of RSI, they could potentially have acted more rapidly to mitigate global warming. Thus, successful anticipation of technology impacts prior to widespread diffusion is not necessary for RSI, even though it is generally beneficial.

## 4.2 When is anticipation difficult?

There are limits to anticipation, and certain areas where anticipating societal impacts of technologies may be more challenging.

**4.2.1 Higher-order effects.** Firstly, higher-order effects can be difficult to predict. This includes the second-order effects of risk-reducing technologies. For instance, the risk-reducing effects of safety technologies depend on their adaptation and related social factors. Even after the development of PALs to prevent the unauthorized launch of nuclear weapons, the US Air Force kept these electronic keys set to only zeros for decades (Ellsberg 2017). This practice was driven by the military's desire to retain decentralized retaliatory capacity if cut off from central command. Similarly, developing defensive technologies may have second-order risk-increasing effects in adversarial situations. For instance, the development

of a US anti-ballistic missile system was considered dangerous because it might upset the careful balance of mutual deterrence, potentially sparking incentives for a pre-emptive strike or inducing additional investments into nuclear delivery systems (Maas 2019).

**4.2.2 General-purpose technologies.** Secondly, the effects of general-purpose technologies that intersect with a wide range of other technologies can be difficult. This is true for the impact of technologies like the computer, internet, or AI technologies, which continue to transform human civilization. However, even in the face of great uncertainty around the effects of general-purpose technology intersections, RSI may still be employed for certain applications of such general-purpose technologies. For instance, even if an artificial intelligence system's future impacts are not yet predictable, we can still deploy application programming interfaces to safeguard language models from misuse (Shevlane 2024).

**4.2.3 Breakthrough discoveries.** Lastly, breakthrough discoveries from basic science research are difficult to anticipate. However, as the story of Szilard demonstrates, security-minded researchers have the potential to reflect on how to share their insights responsibly. Furthermore, depending on what basic science research is conducted, it might be possible to predict what insights may be found. For instance, screening novel bacteria for functional components might identify novel tools for molecular biology, similar to how the discovery of thermostable bacteria led to the 1983 invention of Polymerase Chain Reaction (PCR) (Dove 2018). Similarly, training an AI system with data on certain phenomena, if successful, is likely to improve the ability to model those phenomena. For instance, training an AI system on data related to protein sequences and structures produced a transformative breakthrough in the prediction of protein structures (Jumper *et al.* 2021; Tunyasuvunakool *et al.* 2021).

## 4.3 Overcoming costs of anticipation

Although there are many examples where predicting societal impacts of technology has been helpful, anticipation may have costs. Predictions may be incorrect, thus providing false warnings or inspiring the advancement of technologies incorrectly judged to feature little potential for misuse. Furthermore, widespread practices of anticipation and associated consideration of risks might slow the adoption of beneficial novel technologies. These costs are real and merit substantial weight in policy decisions. Nevertheless, the upsides of anticipation and related RSI practices may outweigh the costs when managing extreme risks, given the potential harms involved. Further investigation may find other goals where the costs of anticipation are worth paying.

The cost of anticipation may decrease as it is done more systematically and better methodologies and tools are developed. Multiple methods for anticipating impacts have been proposed: analogical case studies, scenario exercises, research program mapping, identifying possible negative impacts of technologies and initial products, and researcher interviews (Guston and Sarewitz 2002). Programs such as the IARPA FUSE program have sought to advance new methodologies for technology foresight and assessment ('IARPA - FUSE', 2010). Recent work by Zhou *et al.* (2020) used a deep neural network classifier to forecast emerging technologies.

It was able to predict which patented innovations would become classified as emerging technologies with an accuracy of 77 per cent, one year before these technologies were recognized in Gartner's annual graphical summary of emerging technology (Zhou et al. 2020). More generally, in the last decade, novel forecasting techniques have been developed that can help to bound uncertainty and improve insight into the future (Tetlock and Gardner 2016). Similar advances for predicting possible societal impacts of novel technologies seem achievable.

## 5. Intervening in the timing of technology development

To apply RSI, actors involved in technology development such as government research funding agencies, government regulatory agencies, civil society, and technology companies need to be able to intervene in the pace of development of specific technologies. To leverage risk-reducing interactions between technologies, interventions can aim to slow risk-increasing technologies or speed the development of risk-reducing technologies. The following sections discuss how governments and other relevant actors can use different intervention strategies to implement the principle of RSI. These sections are not intended as a comprehensive assessment of potential strategies but as an initial indication of the diverse set of potentially promising options for implementation.

### 5.1 Strategies for delaying risk-increasing technologies

Governments and other actors can sometimes delay the development of specific risk-increasing technologies. Delaying technologies may seem infeasible due to coordination issues or too controversial if it involves slowing down increases in wealth and health. However, this is not necessarily the case. Policy decisions delay the development of technologies all the time. With every decision to advance one technology, resources are diverted from an alternative that could have been developed. This is most apparent in academic funding, where funding is awarded to a small subset of grant proposals.

It can be difficult to see the missing technologies, the paths not taken. For example, policy decisions have led to less innovation in technologies for geoengineering and human genetic engineering but an alternate decision could have led in a very different direction (Stilgoe, Owen and Macnaghten 2013; Sykora and Caplan 2017).

As resources can frequently be channelled into alternative technologies, delaying select risk-increasing technologies does not necessarily slow down innovation (Mahdi, Nightingale, and Berkhout 2002)—and importantly, this practice can be expected to reduce risks and thus safeguard innovation in the long term. Concrete strategies to delay risk-increasing technologies until their risks have been assessed and managed are presented below and summarized in Table 1.

**5.1.1 Restraint and defunding.** Restraint in developing or defunding specific technologies can be a straightforward approach to slowing their development, but can be complicated by interactions with other actors. Funding for research and development generally leads to more rapid development of related technologies, so reduced funding can directly impact the pace of development (Bolívar-Ramos 2017). The situation

is often complicated by the fact that other developers may continue their efforts, and other actors may step in to fill gaps. For example, former US President Bush restricted the federal funding of stem cell research in 2001 (Taylor 2005). However, in 2004, the California state government stepped in to fund stem cell research while this research also continued in other countries. Nevertheless, given its significant role in science funding, limited US federal funding of stem cell research likely still slowed its development. In situations where development and research happen in a distributed manner and are not driven by a single actor, slowing the development of a technology may be possible unilaterally or may require substantial coordination across multiple stakeholders.

**5.1.2 Moratoria to assess and address risks.** Moratoria of possible risk-increasing technologies may be used to assess risks and advance risk-mitigating measures. In 1974, the international community of researchers working with recombinant DNA engaged in a voluntary moratorium and gathered for the Asilomar conference to evaluate possible public health risks (Berg and Singer 1995). After the conference, researchers resumed recombinant DNA work with agreed safety measures. Next to such voluntary moratoria, governments and funding bodies can also impose moratoria on certain research. The US National Institutes of Health's 2014–2017 moratorium on viral gain-of-function research delayed high-risk research until a new oversight policy was created (Reardon 2017). Moratoria might be especially useful to delay the development of risk-increasing technologies until safety technologies are in place. One example application might be an industry-led moratorium on the distribution of benchtop DNA synthesis machines until effective DNA synthesis screening has been integrated into all devices. This could prevent the irreversible proliferation of devices that allow illicit access to the materials for the creation of any pathogen. Similarly, discussions are underway to ensure that any AI systems that pose 'intolerable' risks are not developed. These include Frontier AI Safety Commitments announced at the AI Seoul Summit in 2024, in which leading AI companies committed to develop risk thresholds and 'not to develop or deploy a model or system at all, if mitigations cannot be applied to keep risks below the thresholds' (UK Department for Science, Innovation and Technology 2024). In the future, such requirements could be codified in domestic or international law.

**5.1.3 Bans.** In certain cases, banning a technology or its development may be feasible. Bans at the international level require substantial coordination and political will; thus, they are only feasible for technologies that are seen as overwhelmingly negative and not critical for national interests. The Chemical Weapons Convention, Biological Weapons Convention, and Nuclear Non-Proliferation Treaty are examples of international treaties that aim to prevent the development and proliferation of certain military technologies. While enforcement is challenging, each of these treaties has slowed the proliferation of these technologies and has also likely slowed the development of more advanced chemical and biological weapons (Wheelis 2006; The power of treaties, 2013; Fuhrmann and Lupu 2016).

Another example is the Montreal Protocol of 1987, which successfully banned ozone-depleting substances. Global consumption of these substances was reduced by 98.5 per cent

**Table 1.** Strategies to delay and assess risk-increasing technologies and research.

| Strategy          | Actors                                     | Example                                                                                                  | Reference                                                                                        |
|-------------------|--------------------------------------------|----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Defunding         | Government<br>Industry<br>Philanthropy     | Reduction in government funding impeded stem cell research                                               | <a href="#">Taylor (2005)</a>                                                                    |
| Moratoria         | Academia<br>Industry<br>Government         | 1974 voluntary moratorium on recombinant DNA research                                                    | <a href="#">Herzog and Parson (2016)</a>                                                         |
| Stage-gating      | Government<br>Industry<br>Academia<br>NGOs | UK geoengineering SPICE experiment was subject to stage-gated review; first postponed and then cancelled | <a href="#">Stilgoe, Owen, and Macnaghten (2013)</a>                                             |
| Bans              | International organizations<br>Government  | Montreal protocol reduced production of ozone-depleting substances                                       | <a href="#">European Environment Agency (2021)</a> ,<br><a href="#">Fuhrmann and Lupu (2016)</a> |
| Regulations       | Government                                 | Less innovation in high-emissions technologies through taxation of fossil fuels                          | <a href="#">Aghion et al. (2016)</a>                                                             |
| Social norms      | Academia<br>NGOs<br>Activists              | Social norms against the recreation of eradicated smallpox virus                                         | <a href="#">Kupferschmidt (2017)</a>                                                             |
| Advocacy          | NGOs<br>Academia<br>Industry               | Letter signed by 50 NGOs led to the termination of the UK SPICE Geoengineering project                   | <a href="#">Stilgoe, Owen, and Macnaghten (2013)</a>                                             |
| Divestment        | NGOs<br>Industry                           | Fossil fuel divestment has had a significant impact on discourse and norms                               | <a href="#">Bergman (2018)</a>                                                                   |
| Information loops | Government<br>NGOs<br>Industry             | Toxic Release Inventory 1986 requiring public reporting of pollutants substantially reduced emissions.   | <a href="#">Meadows (1999)</a>                                                                   |

between 1986 and 2018 ([European Environment Agency 2021](#)). While the ban focused on the production of ozone-depleting substances, it also removed the incentive to continue developing technologies reliant upon ozone-depleting substances and produced a new incentive to develop substitute technologies.

**5.1.4 Shifting the technology portfolio: regulation, norms, and information loops.** Incentives may be used to encourage a shift away from developing risk-increasing technologies. Governments can use regulatory interventions this way, described by Rip and Kemp as ‘weed pulling’ ([Rip and Kemp 1998](#)). An example is carbon taxation. Countries with higher fuel prices (a proxy for greenhouse gas emissions tax level) have lower levels of innovation in high-emissions technologies and higher levels of innovation in green technologies ([Aghion et al. 2016](#)). Similarly, paperwork-heavy regulations may disincentivize relevant research, as is the case for research on dangerous pathogens in the USA ([Evans et al. 2021](#)).

Social norms and pressure may be sufficient to generate incentives to prevent or move away from risk-increasing efforts. Strong social norms against the recreation of dangerous smallpox led to a strong outcry over synthesizing the related horsepox virus, despite this not being banned by existing regulations. Climate activism has driven divestment from fossil fuels. While the direct effects of fossil fuel divestment are small, indirect impacts on societal discourse and norms have been substantial ([Bergman 2018](#)).

Indeed, simply awareness of risks may reduce the development of risk-increasing technologies. To highlight the power of such information loops, Meadows tells the story of the US Toxic Release Inventory ([Meadows 1999](#)). When introduced in 1986, this regulation forced companies to report the release of pollutants publicly. After four years, emissions had

dropped by 40 per cent—without any fines, bans, or other interventions ([Meadows 1999](#)). Information loops may similarly be used to encourage the pursuit of less risky lines of life sciences research. Consideration of Replacement, Reduction, and Refinement of animal use in experiments (‘the 3Rs’) is required in many grant applications to ensure researchers are minimizing animal harm. A similar strategy could be deployed for the safety and security risks of life sciences research. If all proposed research projects were categorized by risk level at the funding stage, this might lead to less risky research being funded and conducted. In the absence of a superior promise of one project over another, researchers and funders should choose the least risky avenue of research. Only if all highly promising low-risk avenues are saturated should researchers and funders engage in more risky projects. Importantly, creating information loops does not rely on government mandates. Non-governmental actors like companies or non-profits may create information loops based on publicly available data.

## 5.2 Strategies for advancing the development of risk-reducing technologies

Governments, firms, philanthropic funders, and other relevant actors are frequently able to advance the development of risk-reducing technologies. Increasing funding or other resources provides a direct pathway toward advancing a specific technology. However, other strategies may also be effective, such as regulation ([Table 2](#)). Advancement of risk-reducing technologies may require less coordination than delaying risk-increasing technologies and may be a viable approach for a range of actors.

**5.2.1 Funding, prizes, advance purchase commitments.** Governments or other technology funders can accelerate specific technologies by prioritizing the funding of related

**Table 2.** Strategies for the preferential advancement of risk-reducing technologies.

| Strategy                                   | Actor                                  | Example                                                                                                            | Reference                   |
|--------------------------------------------|----------------------------------------|--------------------------------------------------------------------------------------------------------------------|-----------------------------|
| Funding and direct development             | Government<br>Industry<br>Philanthropy | ARPA-style efforts led to creation of internet, synthetic biology, carbon nanotubes, and clean energy technologies | Mervis (2016)               |
| Prizes                                     | Government<br>Industry<br>Philanthropy | Genetic engineering attribution challenge winning entries significantly advanced state-of-the-art capabilities     | Crook et al. (2021)         |
| Advanced market commitments                | Government<br>Industry<br>Philanthropy | Advance market commitments have been used to induce development of pneumococcal vaccines                           | Cernuschi et al. (2011)     |
| Technology forcing                         | Government                             | Clean air standards in California in 1988 requiring 2% of car sales to be zero-emission vehicles                   | Schot and Rip (1997)        |
| Niche management                           | Government                             | Feed-in tariffs and subsidies for photovoltaics until commercial viability                                         | Haegel et al. (2017)        |
| Tax incentives                             | Government                             | Tax incentives for carbon sequestration to advance carbon capture and storage technologies                         | Anderson et al., (2021)     |
| Regulation of risk-increasing technologies | Government                             | Innovation in low-emissions technologies through taxation of fossil fuels                                          | Aghion et al. (2016)        |
| Coordination                               | Industry<br>NGOs<br>Government         | Voluntary DNA synthesis screening of companies part of International Gene Synthesis Consortium                     | Diggans and Leproust (2019) |

research and development. For example, when firms and universities increase their research spending, this increases their patent output (Ernst 1998; Bolívar-Ramos 2017).

Specific strategies may be particularly suitable to advance certain technologies. The US Defense Advanced Research Projects Agency (DARPA) has successfully developed specific, application-ready technologies from basic research. The agency is credited with key advances in the creation of the internet, synthetic biology, and carbon nanotubes (Mervis 2016). Focused research organizations (FROs) may be particularly well-suited to develop risk-reducing technologies that constitute public goods (Marblestone et al. 2022).

Prizes may be another tool to encourage the development of specific risk-reducing technologies, especially those relying on interdisciplinary research or with unclear solutions. Genetic engineering attribution, the ability to computationally identify the laboratory origins of engineered DNA sequences, is a defensive technology that may deter the release of a biological weapon. In 2020, a prize, the Genetic Engineering Attribution Challenge, induced the development of new tools that significantly outperformed state-of-the-art approaches (Crook et al. 2021). Similarly, bug bounty programs where individuals outside a software company are rewarded for identifying and reporting security vulnerabilities in that company's software, have a track record of identifying vulnerabilities and enabling the development of more secure software (Walshe and Simpson 2020).

Advance market commitments may be a useful tool for the development of crucial defensive technologies (Monrad, Sandbrink, and Cherian 2021). There may be little incentive to develop vaccines or personal protective equipment in the absence of an ongoing emergency. Nevertheless, development and stockpiling are essential for pandemic preparedness. In such cases, commitments to purchase a select number of units or guarantee a market for a product may be used to induce advances. For instance, advanced market commitments have been used to encourage the development of pneumococcal vaccines (Cernuschi et al. 2011).

### 5.2.2 Regulations can advance risk-reducing technologies.

'Regulation is the mother of invention', as Ruth Rutenberg noted (Rip and Kemp 1998). Governments may use regulations to force the development of risk-reducing technologies. This is frequently the case for safety technologies. For instance, regulations on how to handle pathogens reduce research risks through laboratory safety technologies (Pastorino, de Lamballerie, and Charrel 2017). Similarly, governments could induce the development and adoption of security-sensitive application programming interfaces by mandating their use for machine learning models with significant potential for misuse (Shevlane 2024).

Strategies for advancing substitute technologies have been extensively studied and explored for clean energy and low emissions technologies. A combination of strategies to increase the momentum of niche innovations, weaken existing systems, and strengthen exogenous pressures may help to replace risk-increasing technologies (Geels et al. 2017). Regulations of risk-increasing technologies discussed in Section 5.1.4, including carbon taxation, incentivize the advancement of substitute technologies (Aghion et al. 2016).

Governments may induce the development of specific technologies through technology-forcing. Clean air standards passed in California in 1988 demanded that 2 per cent of car sales must be zero-emission vehicles (Schot and Rip 1997). These vehicles did not exist at the time. Thus, this regulation likely contributed to the advancement of electric vehicles.

Niche management strategies, such as feed-in tariffs that enable the competitiveness of substitutes before their commercial viability, can act as building blocks for larger sociotechnical transformations to manage risks (Schot and Geels 2008). For example, subsidies have successfully helped advance photovoltaics (Haegel et al. 2017). Tax incentives for carbon sequestration have likely induced advances in carbon capture and sequestration technologies (Anderson et al., 2021).

**5.2.3 Coordination and pre-competitive consortia.** Industry coordination may advance risk-reducing technologies and practices. Basic DNA synthesis screening of 80 per cent of the market exists because of voluntary coordination in the International Gene Synthesis Consortium (International Gene Synthesis Consortium 2017; Diggans and Leproust 2019). Pre-competitive consortia are a promising method for facilitating cross-sector coordination in biomedicine (Mittleman, Neil, and Cutcher-Gershenfeld 2013). Such pre-competitive consortia may be used to develop risk-reducing technologies. In AI, a range of coordination mechanisms, such as the Frontier Model Forum and Partnership on Artificial Intelligence, seeks to bring together leading AI companies to share best practices and advance the state of the art on safe and responsible AI development (Frontier Model Forum 2024; Partnership on AI 2024).

### 5.3 Challenges for timing interventions

**5.3.1 Wrestling deterministic paths.** Competition between groups of actors to gain a technological advantage can cause the development of technologies to proceed along more deterministic paths (Dafoe 2015). In cases of competition to win a market or international competition, the choices of individual actors may be constrained by the competitive dynamic. For example, the US development of nuclear weapons was prompted by fears that Nazi Germany would develop them first. If progress in a specific technology is not led by a single or a handful of coordinating actors, delaying or making technologies safer may be difficult. Coordination through international organizations may have the greatest leverage to delay or shape technologies with global power implications. For other technologies, national governments may be able to break competitive dynamics across industry and academia that drive the development of risk-increasing technologies. Next to competition, various other dynamics can also lead to differing degrees of path dependence in technological development, including first-mover advantages, institutional persistence, structural inertia, limits to absorptive capacity, and natural monopolies (Vergne and Durand 2010).

The more resources are already committed to the development of a specific technology, the more difficult it is to have an impact on the timing of its development. This is also true for the advancement of risk-reducing technologies. Small actors can have significant effects if no or few other actors are making similar investments. In such cases, a small investment may have a prolonged effect on the level of technology available. Historically, this has to be proven particularly relevant in cases with little market incentive to develop a publicly beneficial technology. An example is vaccines for neglected or possible future diseases (Monrad, Sandbrink, and Cherian 2021).

Interventions to promote the development of technology can sometimes result in the technology passing a threshold of commercial viability, thereby producing a positive feedback loop that begets further and continued development of the technology by other actors. Government support has successfully led to photovoltaics becoming commercially viable (McDonald and Schrattenholzer 2001; Haegel *et al.* 2017), triggering additional investments and reductions in cost. Such feedback loops may render the replacement of marginally risk-increasing technologies cost-effective. The advancement of RNA vaccines and their success against COVID-19 may induce a shift away from viral vector vaccines, which are

associated with the development of capabilities with greater potential to be misused (Sandbrink and Koblentz 2022). Similarly, non-viral delivery methods for gene therapy may eventually become superior to viral delivery methods and thus lead to their replacement.

**5.3.2 Directing change.** Ensuring that interventions in the development of specific technologies have the desired effect can be difficult. While it generally appears to be possible to advance the development of risk-reducing technologies, this may inadvertently speed up related risk-increasing technologies or create interactions with other technologies that result in negative societal impacts. For example, civilian nuclear power technologies are difficult to develop in a way that does not advance nuclear weapons development.

One part of the difficulty in directing change is the challenge of anticipating higher-order effects of risk-reducing technologies. As discussed in Section 4.2, anticipating the effects of general-purpose technologies may be particularly difficult. Advancing the use of metagenomic sequencing for pathogen detection may inadvertently lead to advanced sequencing applications in areas of biotechnology that increase the pandemic risks that pathogen detection hopes to reduce. Interventions to advance or delay specific technologies must be closely coupled with reviewing potential downstream risks and appropriately weighing these risks against the potential benefits.

Another part of the difficulty in directing change is predicting the effects of interventions. For instance, increasing taxation on a risk-increasing technology may inadvertently lead to alternatives with even more or different undesirable consequences. Additional funding for a risk-reducing technology may inadvertently displace other funders. These challenges, though real, are not specific to risk-sensitive innovation.

**5.3.3 Balancing different values.** Risk-sensitive innovation requires governments, philanthropists, or other relevant actors to pursue interventions in pursuit of a goal, such as mitigating the potential negative impact of a technology. However, any given goal needs to be balanced with other values or goals. Global security needs to be balanced with the values of scientific freedom and openness. In certain cases, reducing catastrophic risks may be traded against short-term gains.

Risk-sensitive innovation does not need to dominate every governance decision to be successfully implemented. Despite the largely universal assumption that a sustainable economy is desirable, this does not mean that adopting a goal of sustainable development necessitates every single policy decision to move in this direction. Rather, it is a cross-cutting consideration of this goal and the aggregate of many decisions that allow such goals to be implemented. Public engagement is needed to inform how to manage the trade-offs involved in implementation. This is partially ensured through political representatives for government actors but may also call for the engagement of the broader public, including experts from different backgrounds and areas of expertise.

**5.3.4 International coordination.** Various strategies for implementing risk-sensitive innovation benefit from the ability to coordinate between governments at an international level. However, RSI does not require perfect international,

or even national, coordination. Just as there is no single regime to mitigate climate change but rather a complex system composed of different international, national, and other societal initiatives (Keohane and Victor 2011), a combination of initiatives may be sufficient to put RSI into practice. Indeed, a more fragmented system may ensure continuous adaptation to evolving circumstances and values, including consideration of local values and culture (Alter and Raustiala 2018).

## 6. Conclusion

As novel technologies are becoming increasingly powerful, there is a greater urgency for responsible innovation practices to mitigate associated risks. The interaction and relative sequence of different technologies are crucial levers for improving societal outcomes. Thus, actors interested in improving societal outcomes, such as government research funding agencies, government regulators, philanthropic funders, and firms engaged in corporate social responsibility, could include consideration of RSI as part of a broader holistic assessment of priorities for allocating resources to research and technology development.

In particular, these actors should consider RSI to manage potential extreme or global risks from technologies. Some technologies or issues appear potentially well suited to the application of this principle for risk reduction, including pandemic prevention investments, DNA synthesis screening, trustworthy machine learning models, and consideration of risks when funding science. The engagement of experts from different fields would likely illuminate further opportunities to usefully use RSI as a concept to mitigate specific technology risks.

## Acknowledgements

We are grateful to Michael Aird, Markus Anderljung, Jan Ole Ernst, Ben Garfinkel, Sihao Huang, Matthijs Maas, Cassidy Nelson, and James Wagstaff for their useful discussions and comments on the manuscript. Furthermore, we are also grateful for feedback from participants of work-in-progress meetings of the Future of Humanity Institute and Centre for the Governance of AI. We thank Shrestha Rath for help with formatting and organizing references.

## Author contributions

Jonas B. Sandbrink (Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Hamish Hobbs (Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Jacob L. Swett (Conceptualization, Writing—review & editing), Allan Dafoe (Conceptualization, Writing—review & editing), and Anders Sandberg (Conceptualization, Writing—review & editing).

*Conflict of interest statement.* None declared.

## Funding

The doctoral research of J.B.S. is funded by Open Philanthropy. H.H.'s contribution to the paper largely occurred while being funded as a Research Scholar at the Future of Humanity Institute.

## Data Availability

Not Applicable.

## References

- Aarne, O., Fist, T., and Withers, C. (2024) 'Secure, Governable Chips', Center for a New American Security.
- Acemoglu, D. (2002) 'Directed Technical Change', *The Review of Economic Studies*, 69: 781–809.
- Aghion, P. et al. (2016) 'Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry', *Journal of Political Economy*, 124: 1–51.
- Alter, K. J., and Raustiala, K. (2018) 'The Rise of International Regime Complexity', *Annual Review of Law and Social Science*, 14: 329–49.
- Anderson, J. J. et al. (2021) 'A Techno-economic Assessment of Carbon-sequestration Tax Incentives in the U.S. Power Sector', *International Journal of Greenhouse Gas Control*, 111: 103450.
- Atlas, R. M., and Dando, M. (2006) 'The Dual-use Dilemma for the Life Sciences: Perspectives, Conundrums, and Global Solutions', *Biosecurity and Biodefense: Biodefense Strategy, Practice, and Science*, 4: 276–86.
- Aven, T. (2016) 'Risk Assessment and Risk Management: Review of Recent Advances on Their Foundation', *European Journal of Operational Research*, 253: 1–13.
- Banta, D. (2009) 'What Is Technology Assessment?', *International Journal of Technology Assessment in Health Care*, 25: 7–9.
- Bengio, Y. (2024) 'International Scientific Report on the Safety of Advanced AI', Interim Report.
- Berg, P., and Singer, M. (1995) 'The Recombinant DNA Controversy: Twenty Years Later', *Nature Biotechnology*, 13: 1132–4.
- Bergman, N. (2018) 'Impacts of the Fossil Fuel Divestment Movement: Effects on Finance, Policy and Public Discourse', *Sustainability*, 10: 2529.
- Boenink, M., and Kudina, O. (2020) 'Values in Responsible Research and Innovation: From Entities to Practices', *Journal of Responsible Innovation*, 7: 450–70.
- Bolívar-Ramos, M. T. (2017) 'The Relation between R&D Spending and Patents: The Moderating Effect of Collaboration Networks', *Journal of Engineering and Technology Management*, 46: 26–38.
- Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies*, Illustrated edn. Oxford: OUP.
- Bresnahan, T. F., and Trajtenberg, M. (1995) 'General Purpose Technologies 'Engines of Growth'', *Journal of Econometrics*, 65: 83–108.
- Brittain, H. K., Scott, R., and Thomas, E. (2017) 'The Rise of the Genome and Personalised Medicine', *Clinical Medicine London*, 17: 545–51.
- Brundage, M. et al. (2018) 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation'.
- Caldwell, D. (1987) 'Permissive Action Links: A Description and Proposal', *Survival*, 29: 224–38.
- Caldwell, M. et al. (2020) 'AI-enabled Future Crime', *Crime Science*, 9: 14.
- Cernuschi, T. et al. (2011) 'Advance Market Commitment for Pneumococcal Vaccines: Putting Theory into Practice', *Bulletin World Health Organisation*, 89: 913–8.
- Chu, S. (2009) 'Carbon Capture and Sequestration', *Science*, 325: 1599–1599.
- Collingridge, D. (1980) *The Social Control of Technology*. London: Frances Pinter.
- Croft, S. (1996) *Strategies of Arms Control: A History and Typology*. Manchester: Manchester University Press.
- Crook, O. M. et al. (2021) 'Analysis of the First Genetic Engineering Attribution Challenge', *Nature Communications*, 13: 7374.
- Cummings, P. et al. (2002) 'Association of Driver Air Bags with Driver Fatality: A Matched Cohort Study', *BMJ*, 324: 1119–22.
- Cummins, J. et al. (2011) 'Do Seat Belts and Air Bags Reduce Mortality and Injury Severity after Car Accidents?', *American Journal of Orthopedics*, 40: E26–9.

- Dafoe, A. (2015) 'On Technological Determinism: A Typology, Scope Conditions, and A Mechanism', *Science, Technology, & Human Values*, 40: 1047–76.
- Dake, K. (1991) 'Orienting Dispositions in the Perception of Risk: An Analysis of Contemporary Worldviews and Cultural Biases', *Journal of Cross-Cultural Psychology*, 22: 61–82.
- Diggins, J., and Leproust, E. (2019) 'Next Steps for Access to Safe, Secure DNA Synthesis', *Frontiers in Bioengineering and Biotechnology*, 7: 86.
- Dorn, A. W. (2019) 'Eliminating Hidden Killers: How Can Technology Help Humanitarian Demining?', *Stability: International Journal of Security and Development*, 8: 5.
- Dove, A. (2018) 'PCR: Thirty-five Years and Counting', Technology Feature.
- Duprex, W. P. *et al.* (2015) 'Gain-of-function Experiments: Time for a Real Debate', *Nature Reviews, Microbiology*, 13: 58–64.
- Ellsberg, D. (2017) *The Doomsday Machine: Confessions of a Nuclear War Planner*, 1st edn. New York: Bloomsbury.
- Ernst, H. (1998) 'Industrial Research as a Source of Important Patents', *Research Policy*, 27: 1–15.
- Esvelt, K. M. (2018) 'Inoculating Science against Potential Pandemics and Information Hazards', *PLoS Pathogen*, 14: e1007286.
- European Environment Agency (2021) 'Consumption of Ozone-depleting Substances [WWW Document]', European Environment Agency. <https://www.eea.europa.eu/ims/consumption-of-ozone-depleting-substances>, accessed 8 Sept. 2022.
- Evans, S. W. *et al.* (2021) 'Stakeholder Engagement Workshop on the Implementation of the United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern: Workshop Report (SSRN Scholarly Paper No. ID 3955051)', Rochester: Social Science Research Network.
- Farmer, J. D. *et al.* (2019) 'Sensitive Intervention Points in the Post-carbon Transition', *Science*, 364: 132–4.
- Federal Highway Administration, Office of Highway Information Management (2009) 'Motor Vehicle Traffic Fatalities 1900–2007', Washington: Federal Highway Administration.
- Fisher, E., Mahajan, R. L., and Mitcham, C. (2006) 'Midstream Modulation of Technology: Governance from Within', *Bulletin of Science, Technology & Society*, 26: 485–96.
- Friedman, B. *et al.* (2013) 'Value Sensitive Design and Information Systems', in N. Doorn, D. Schuurbijs, I. van de Poel *et al.* (eds) *Early Engagement and New Technologies: Opening up the Laboratory, Philosophy of Engineering and Technology*, pp. 55–95. Dordrecht: Springer.
- Frontier Model Forum (2024) 'About Us', Frontier Model Forum. <https://www.frontiermodelforum.org/about-us/>, accessed 14 Jun. 2024.
- Fuhrmann, M., and Lupu, Y. (2016) 'Do Arms Control Treaties Work? Assessing the Effectiveness of the Nuclear Nonproliferation Treaty', *International Studies Quarterly*, 60: 530–9.
- Gallagher, N. W. (2015) 'Re-thinking the Unthinkable: Arms Control in the Twenty-First Century', *The Nonproliferation Review*, 22: 469–98.
- Geels, F. W. (2019) 'Socio-technical Transitions to Sustainability: A Review of Criticisms and Elaborations of the Multi-Level Perspective', *Current Opinion in Environmental Sustainability*, 39: 187–201.
- Geels, F. W. *et al.* (2017) 'Sociotechnical Transitions for Deep Decarbonization', *Science*, 357: 1242–4.
- Green, J. F. (2021) 'Does Carbon Pricing Reduce Emissions? A Review of Ex-post Analyses', *Environmental Research Letters*, 16: 043004.
- Guston, D. H. (2014) 'Understanding 'Anticipatory Governance'', *Social Studies of Science*, 44: 218–42.
- Guston, D. H., and Sarewitz, D. (2002) 'Real-time Technology Assessment', *Technology in Society, American Perspectives on Science and Technology Policy*, 24: 93–109.
- Haegel, N. M. *et al.* (2017) 'Terawatt-scale Photovoltaics: Trajectories and Challenges', *Science*, 356: 141–3.
- Herzog, M. M., and Parson, E. A. (2016) 'Moratoria for Global Governance and Contested Technology: The Case of Climate Engineering', UCLA Public Law & Legal Theory, Series Open Access Policy Deposits, UCLA: School of Law.
- Ho, L. *et al.* (2023) 'International Institutions for Advanced AI'.
- Horowitz, M. C. (2018) 'Artificial Intelligence, International Competition, and the Balance of Power', *Texas National Security Review*, 1: 36–57.
- IARPA - FUSE [WWW Document] (2010) FUSE: Foresight and Understanding from Scientific Exposition <https://www.iarpa.gov/research-programs/fuse>, accessed 10 May 2022.
- International Gene Synthesis Consortium (2017) 'Harmonized Screening Protocol V2.0', International Gene Synthesis Corporation.
- Jordan, K., Jahn, F., and Aapro, M. (2015) 'Recent Developments in the Prevention of Chemotherapy-induced Nausea and Vomiting (CINV): A Comprehensive Review', *Annals of Oncology*, 26: 1081–90.
- Jumper, J. *et al.* (2021) 'Highly Accurate Protein Structure Prediction with AlphaFold', *Nature*, 596: 583–9.
- Kaplan, J. *et al.* (2020) 'Scaling Laws for Neural Language Models', OpenAI.
- Keohane, R. O., and Victor, D. G. (2011) 'The Regime Complex for Climate Change', *Perspectives on Politics*, 9: 7–23.
- Koblentz, G. D. (2009) *Living Weapons: Biological Warfare and International Security*. Ithaca: CUP.
- Kreps, S. E. (2018) 'The Institutional Design of Arms Control Agreements', *Foreign Policy Analysis*, 14: 127–47.
- Kupferschmidt, K. (2017) 'How Canadian Researchers Reconstituted an Extinct Poxvirus for \$100,000 Using Mail-order DNA', *Science Insider*.
- Lanouette, W. (1992) *Genius in the Shadows: A Biography of Leo Szilard: The Man behind the Bomb*. Oxford; New York: Charles Scribner's Sons; Maxwell Macmillan International.
- Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2021) 'Explainable AI: A Review of Machine Learning Interpretability Methods', *Entropy*, 23: 18.
- Maas, M. M. (2019) 'How Viable Is International Arms Control for Military Artificial Intelligence? Three Lessons from Nuclear Weapons', *Contemporary Security Policy*, 40: 285–311.
- Mahdi, S., Nightingale, P., and Berkhout, F. (2002) 'A Review of the Impact of Regulation on the Chemical Industry', University of Sussex.
- Marblestone, A. *et al.* (2022) 'Unblock Research Bottlenecks with Non-profit Start-ups', *Nature*, 601: 188–90.
- McDonald, A. and Schrattenholzer, L. (2001) 'Learning Rates for Energy Technologies', *Energy Policy*, 29: 255–61.
- McKenzie, R. *et al.* (2019) 'Success of Montreal Protocol Demonstrated by Comparing High-Quality UV Measurements with "World Avoided" Calculations from Two Chemistry-Climate Models', *Scientific Reports*, 9: 12332.
- Meadows, D. (1999) 'Leverage Points: Places to Intervene in a System', Sustainability Institute.
- Mervis, J. (2016) 'What Makes DARPA Tick?' *Science*, 351: 549–53.
- Mittleman, B., Neil, G., and Cutcher-Gershenfeld, J. (2013) 'Precompetitive Consortia in Biomedicine—How are We Doing?' *Nature Biotechnology*, 31: 979–85.
- Monrad, J. T., Sandbrink, J. B., and Cherian, N. G. (2021) 'Promoting Versatile Vaccine Development for Emerging Pandemics', *NPJ Vaccines*, 6: 1–7.
- Mowery, D. C., Nelson, R. R., and Martin, B. R. (2010) 'Technology Policy and Global Warming: Why New Policy Models are Needed (Or Why Putting New Wine in Old Bottles Won't Work)', *Research Policy*, 39: 1011–23.
- National Security Commission on Artificial Intelligence (2021) 'Final Report', Washington.
- Nucleic Acid Observatory Consortium (2021) 'A Global Nucleic Acid Observatory for Biodefense and Planetary Health', *arXiv*, <https://doi.org/10.48550/arXiv.2108.02678>.

- Nye, J. S. (1987) 'Nuclear Learning and U.S.–Soviet Security Regimes', *International Organization*, 41: 371–402.
- OpenAI (2023) 'Governance of Superintelligence [WWW Document]', <https://openai.com/index/governance-of-superintelligence/>, accessed 14 Jun. 2024.
- Our World in Data (2022) 'Cumulative CO<sub>2</sub> Emissions by World Region [WWW Document]', Our World in Data. <https://ourworldindata.org/grapher/cumulative-co2-emissions-region>, accessed 17 Aug. 2022.
- Owen, R., Macnaghten, P., and Stilgoe, J. (2012) 'Responsible Research and Innovation: From Science in Society to Science for Society, with Society', *Science and Public Policy*, 39: 751–60.
- Partnership on AI (2024) 'About Us [WWW Document]', Partnership on AI. <https://partnershiponai.org/about/>, accessed 14 Jun. 2024.
- Partnership on AI (2023) 'PAI's Responsible Practices for Synthetic Media'.
- Pastorino, B., de Lamballerie, X., and Charrel, R. (2017) 'Biosafety and Biosecurity in European Containment Level 3 Laboratories: Focus on French Recent Progress and Essential Requirements', *Frontiers in Public Health*, 5: 121.
- Reardon, S. (2017) 'US Government Lifts Ban on Risky Pathogen Research', *Nature*, 553: 11–11.
- Rip, A., and Kemp, R. (1998) 'Technological Change', in *Human Choice and Climate Change*. Vol. II, pp. 327–99. Columbus: Battelle Press.
- Sandbrink, J. B. et al. (2021) 'Safety and Security Concerns regarding Transmissible Vaccines', *Nature Ecology and Evolution*, 5: 405–6.
- Sandbrink, J. B., and Koblenz, G. D. (2022) 'Biosecurity Risks Associated with Vaccine Platform Technologies', *Vaccine*, 40: 2514–23.
- Sandbrink, J. B., and Shattock, R. J. (2020) 'RNA Vaccines: A Suitable Platform for Tackling Emerging Pandemics?', *Frontiers in Immunology*, 11: 608460.
- Schaller, R. R. (1997) 'Moore's Law: Past, Present and Future', *IEEE Spectroscopy*, 34: 52–9.
- Schelling, T. C., and Halperin, M. H. (1961) *Strategy and Arms Control*. New York: Twentieth Century Fund.
- Schoch-Spana, M. et al. (2017) 'Global Catastrophic Biological Risks: Toward a Working Definition', *Health Security*, 15: 323–8.
- Schot, J., and Geels, F. W. (2008) 'Strategic Niche Management and Sustainable Innovation Journeys: Theory, Findings, Research Agenda, and Policy', *Technology Analysis & Strategic Management*, 20: 537–54.
- Schot, J., and Rip, A. (1997) 'The past and Future of Constructive Technology Assessment', *Technological Forecasting and Social Change*, 54: 251–68.
- Seger, E. et al. (2020) 'Tackling Threats to Informed Decisionmaking in Democratic Societies', *The Alan Turing Institute*, 2020.
- Shevlane, T. et al. (2023) 'Model Evaluation for Extreme Risks', Google Deepmind.
- Shevlane, T. (2024) 'Structured Access: An Emerging Paradigm for Safe AI Deployment', in J. B. Bullock, Y.-C. Chen and J. Himmelreich et al. (eds) *The Oxford Handbook of AI Governance*, pp. 604–18. Oxford: OUP.
- Stilgoe, J., Owen, R., and Macnaghten, P. (2013) 'Developing a Framework for Responsible Innovation', *Research Policy*, 42: 1568–80.
- Sykora, P., and Caplan, A. (2017) 'The Council of Europe Should Not Reaffirm the Ban on Germline Genome Editing in Humans', *EMBO Report*, 18: 1871–2.
- Taylor, P. L. (2005) 'The Gap between Law and Ethics in Human Embryonic Stem Cell Research: Overcoming the Effect of U.S. Federal Policy on Research Advances and Public Benefit', *Science and Engineering Ethics*, 11: 589–616.
- Tetlock, P. E., and Gardner, D. (2016) *Superforecasting: The Art and Science of Prediction*. New York: Random House.
- The Power of Treaties (2013) 'The Power of Treaties', *Nature*, 501: 5.
- Tunyasuvunakool, K. et al. (2021) 'Highly Accurate Protein Structure Prediction for the Human Proteome', *Nature*, 596: 590–6.
- UK Department for Science, Innovation and Technology (2024) 'Frontier AI Safety Commitments, AI Seoul Summit 2024 [WWW Document]', GOV.UK <https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024>, accessed 14 Jun. 2024.
- UNODA (2024) 'Lethal Autonomous Weapon Systems (LAWS) [WWW Document]', <https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/>, accessed 14 Jun. 2024.
- US Cybersecurity & Infrastructure Security Agency (2024) 'Roadmap for AI [WWW Document]', <https://www.cisa.gov/resources-tools/resources/roadmap-ai>, accessed 14 Jun. 2024.
- van Dijk, A. et al. (2013) 'Skin Cancer Risks Avoided by the Montreal Protocol-Worldwide Modeling Integrating Coupled Climate-Chemistry Models with a Risk Model for UV', *Photochemistry and Photobiology*, 89: 234–46.
- Vaynman, J. (2021) 'Better Monitoring and Better Spying: The Implications of Emerging Technology for Arms Control', Texas National Security Review. <https://tnsr.org/2021/09/better-monitoring-and-better-spying-the-implications-of-emerging-technology-for-arms-control/>, accessed 14 Jun. 2024.
- Vaynman, J., and Volpe, T. A. (2023) 'Dual Use Deception: How Technology Shapes Cooperation in International Relations', *International Organization*, 77: 599–632.
- Vergne, J.-P., and Durand, R. (2010) 'The Missing Link between the Theory and Empirics of Path Dependence: Conceptual Clarification, Testability Issue, and Methodological Implications', *Journal of Management Studies*, 47: 736–59.
- Volpe, T. A. (2019) 'Dual-use Distinguishability: How 3D-printing Shapes the Security Dilemma for Nuclear Programs', *Journal of Strategic Studies*, 42: 814–40.
- Waldrop, M. M. (2016) 'The Chips are down for Moore's Law', *Nature*, 530: 144–7.
- Walshe, T. and Simpson, A. (2020) 'An Empirical Study of Bug Bounty Programs', Presented at the 2020 IEEE 2nd International Workshop on Intelligent Bug Fixing (IBF), London, ON, Canada, pp. 35–44.
- Wandall, B. (2004) 'Values in Science and Risk Assessment', *Toxicol Letter*, 152: 265–72.
- Wetterstrand, K. (2022) 'DNA Sequencing Costs: Data [WWW Document]', National Human Genome Research Institute, <https://www.genome.gov/sequencingcostsdata>, accessed 28 Mar. 2022.
- Wheelis, A. (2006) *The Way We Are*, 1st edn. New York: W.W. Norton.
- Williams, H. (2019) 'Asymmetric Arms Control and Strategic Stability: Scenarios for Limiting Hypersonic Glide Vehicles', *Journal of Strategic Studies*, 42: 789–813.
- Wisotzki, S., and Kühn, U. (2021) 'Crisis in Arms Control: An Introduction', *Zeitschrift für Friedens- und Konfliktforschung*, 10: 183–94.
- World Health Organization. (2021) *Emerging Technologies and Dual-use Concerns: A Horizon Scan for Global Public Health*. Geneva: WHO.
- Zhou, Y. et al. (2020) 'Forecasting Emerging Technologies Using Data Augmentation and Deep Learning', *Scientometrics*, 123: 1–29.
- Zwart, H., Landeweerd, L., and van Rooij, A. (2014) 'Adapt or Perish? Assessing the Recent Shift in the European Research Funding Arena from 'ELSA' to 'RRI', *Life Sciences, Society and Policy*, 10: 11.