Ian HUGHES, David E. WINICKOFF
DOI: 10.3217/978-3-85125-855-4-10

Sustainability Transitions and Deep Institutional Innovation—Rethinking RRI

I. Hughes[1] and D. E. Winickoff[2]

[1]SFI MaREI Centre for Energy, Climate and Marine, Environmental Research Institute, University College Cork, Cork, Ireland
[2]Science and Technology Policy Division, Directorate for Science, Technology and Innovation, OECD, 2, Rue André-Pascal, 75775 Paris Cedex 16, France

Abstract. The scale and pace of technological change, alongside the socially disruptive consequences of new technologies, have created a growing perception that the future of work, democracy and other aspects of social order will require new forms of technology governance (Winickoff and Pfotenhauer, 2018). Responsible research and innovation (RRI) has emerged as an important approach and set of practices, aimed at integrating ethical and social issues more directly into innovation and into the governance of science, innovation and technology. RRI frameworks aim to foster science, technology and innovation through a process of anticipation, reflexivity, inclusion and responsiveness. RRI marks an advance in approaches to science and technology governance as it aims to affect upstream development and help direct the very trajectory of technology towards the solution of critical societal challenges. Nevertheless, existing RRI practices have focused largely on scientific and technological practices outside of their economic, social and political context. As such, they have not adequately come to terms with the complexity and multi-institutional nature of the challenges that technology development presents. More effective solutions must begin with the right models for deep institutional change. This paper outlines a new so-called DIIS model (Hughes et al. 2021) of technological change, which presents a re-conceptualisation of the role of technological change in societal transitions. The paper ultimately argues that while existing approaches to technology policy are moving in the right direction, they should seek to address technology more explicitly within its economic, political and social systems.

Keywords: Sustainability transitions; institutional change; technology governance; deep institutional innovation; responsible research and innovation

1 Introduction

While critical for addressing some of society’s most pressing crises, innovation can also have negative consequences for individuals and societies, as witnessed in previous waves of industrial revolution or in current debates around digitization, data privacy, and artificial intelligence. In fact, the ambiguous societal implications of technologies bring them to the forefront of popular media and political debate. Indeed, the governance of AI, blockchain, autonomous vehicles and genome editing have emerged as issues of high concern.

Responsible Research and Innovation (RRI) has emerged as a set of principles and practices aimed at governing technology for the public good (see e.g. Owen et al., 2012; Burget et al., 2017). RRI has arisen from concerns about detrimental impacts of existing and emerging technologies. At the same time, the development of RRI is happening at a time of not only accelerating technological change, but of profound environmental, political, and social change. Climate change has been acknowledged as a planetary phenomenon that threatens the climatic and ecological balance of the planet, with severe consequences for humanity. And climate change is but one of a series of crises, which also include environmental degradation, species extinction, crises of economic and political inequality, democratic malaise, the rise of authoritarian populism, rising geopolitical tensions, the persistence of extreme poverty, and pervasive levels of violence. All of these changes feature technology as constituent elements.

This article examines RRI principles and practices in the context of this particular historic moment of these cascading crises and puts the analysis in the broader context of deep transitions and theories of systems innovation. The outline of the article is as follows. Using artificial intelligence (AI) as a leading case, Section 2 illustrates the array of governance challenges presented by emerging technologies which RRI aims to address. These concerns range from the need to control disruptive transformative technologies, defend against existential threats, steering RDI to address societal challenges, and a democratic imperative that people should participate in decisions that profoundly impact them.

In Section 3, we briefly review existing RRI principles and practices, including Stilgoe et al.'s (2013) framework for RRI and their four dimensions of responsible innovation - anticipation, reflexivity, inclusion and responsiveness - that provide a framework for raising, discussing and responding to the diverse challenges raised by many existing and emerging technologies. We highlight some of the challenges that existing RRI faces in effectively placing ethical boundaries on technology development, including the need for systemic application of RRI instruments both to a diverse range of innovation actors and across the entire R&I cycle.

In Section 4, we step back to look at the larger context within which RRI is being developed and applied. We briefly outline two existing theories of sociotechnical change that both seek to understand technological change and its wider societal drivers, influences and impacts. These theories are the Multi-Level Perspective (MLP) of Geels (2005), and the Deep Transition theory of Schot and Kanger (2018). Both of these theories examine the dynamics of transition from one socio-technical system to another (the transition from a carbon-based energy system for example to a renewable energy system). While the MLP focuses mainly on the transition of single sociotechnical systems, deep transition theory attempts to explain the dynamics of the emergence of entire socio-technical paradigms, particularly the existing paradigm of industrial modernity. Finally in this Section, we introduce our new framing for societal transformation, the model of Deep Institutional Innovation for Sustainability and Human Development (DIIS) which aims to situate technology as a social institution embedded within a range of other social institutions including economics, politics, gender, religion and education, with which the technology system co-evolves.

In Section 5, we return to critique existing principles and practices of RRI in the light of the DIIS model of whole of society transformation. We conclude that while RRI tends to address the problem of science and technology governance through the micro-practices and contexts of the innovation system – and therefore provides a necessary intervention -- the deep embeddedness of innovation within political, economic and cultural systems is left largely unaddressed. In short, as a general matter, RRI may be working without the correct model of a complex and interconnected innovation system. Recent models of technological and societal change that diagnose the deeper drivers of the current moment of crisis may offer a necessary foundation for the development of a new innovation policy.

2 Challenges for Governing Emerging Technology: The Case of AI

In order to illustrate the high stakes of emerging technologies for basic human values, and therefore governance, Artificial Intelligence is an excellent example. AI is a potentially powerful general-purpose technology (GPT) with the ability to cause broad transformation of the economy and society (OECD, 2019, Trajtenberg, 2019). It is anticipated that progress could be rapid, with potential major advances in medicine and health (Goodman et al., 2020), energy (Ahmad et al., 2021), transportation (Abduljabbar et al., 2019), education (Owoc et al., 2021), innovation (Cockburn et al., 2018), and sustainable development (Vinuesa et al., 2020). The risks, however, are substantial and diverse.

2.1 Labour Displacement

One prominent issue of concern is the future of work: AI carries the potential to transform production systems and to replace human labour in many sectors of the economy. The displacement of labour by AI is predicted to be significant in the future and may result in levels of unemployment that could undermine social cohesion and stability. Recent estimates by the OECD are that 14% of all jobs across 32 OECD countries have a high risk of automation (Nedelkoska and Quintini, 2018), while a further 32% of jobs may experience significant changes to how they are carried out. There is an emerging consensus that artificial intelligence is likely to represent a new employment paradigm (OECD, 2018) with far reaching consequences for individuals and societies.

2.2 Inequality

There are concerns also about the impact of AI on inequality, between firms, in terms of exacerbating existing inequalities, and between generations. Current AI research and development activity is concentrated within a small number of companies, raising fears that rapid breakthroughs in AI could result in unprecedented increases in capital accrual by those firms. According to O'Keefe et al. (2020), the growth in economic wealth from advanced AI could be unprecedented in magnitude and speed, potentially disrupting the structure of the global economy and resulting in the rapid creation of an oligopolistic global market structure.

AI could also replicate or even exacerbate the existing inequalities that have already emerged in terms of the digital divide between demographics within society who do and do not have access to digital technologies (Van Dijk, 2017). A rapid transition to AI could also potentially result in increased inequality between generations. According to Sachs and Kotlikoff (2012), since AI is designed, owned and run by skilled workers, who are typically mid-career, AI is likely to impact detrimentally on young unskilled labour, while benefiting older skilled labour, thereby depressing the wages and savings of the next and future generations.

2.3 Range of Safety Concerns

AI raises an array of safety concerns. 'Narrowly competent' AI systems are already being applied across a wide range of domains from transport to energy to banking and finance and beyond. In applications like self-driving cars, automated trading systems, air traffic control, or control of the power grid, AI system failures could lead to severe disruptions or mass casualties.

In addition, the potential for AI to disrupt political systems is already impacting, for example, through the emergence of deep fakes that make it difficult to distinguish between truth and misinformation. The deployment of mass surveillance systems based on AI also highlights the misapplication of AI to reinforce authoritarian forms of government (Wright, 2018). Yet another danger is the emerging arms race in lethal autonomous weapons (Haner and Garcia, 2019).

2.4 Existential Threats—'SuperIntelligence'

The possibility of AI posing an existential threat to humanity constitutes another thread in AI literature. Bostrom (2014, 52) uses the term 'superintelligence' to refer to "intellects that greatly outperform the best current human minds across many very general cognitive domains". At present, AI systems are specialised 'narrowly competent' systems that perform specific, restricted tasks that approximate to, or in some cases exceed, human capacities in those restricted areas. The future evolution of AI systems, however, may see advances towards 'general intelligence' (AGI), that replicates human intelligence is its generality. The key concern regarding AGI research is the development of autonomous artificially intelligent agents which are much more intelligent than humans, and which pursue goals that conflict with our own, perhaps even leading to our own demise. While the feasibility of AGI remains controversial (Fjelland, 2020), one survey of AI experts reflects the belief that there is a significant (>25%) chance that superhuman capabilities in strategic domains could be developed within the next thirty years (Grace et al., 2018).

2.5 Systemic Impacts

This short review of the concerns surrounding the development of AI show that the range of issues of concern are diverse and far reaching. When the focus is broadened to other new and emerging technologies such as biotechnology, neurotechnology, nanotechnology and online digital technologies, the scope, breadth and critical nature of the governance challenges can be seen to be daunting. The development of these new and emerging technologies essentially constitutes what Callon calls 'society in the making' (Callon, 1987). As such, the impacts of R&I are potentially deeply 'systemic' and reach far beyond those stakeholders who are directly involved, such a researchers and funding agencies. In terms of governance, new and emerging technologies therefore raise a number of fundamental difficulties.

First, the ubiquity of their potential impacts raises the question as to who should have a say in R&I, and what processes might possibly be put in place to direct and control such changes.

Second, as Hajer (2003) points out, disruptive technologies typically fall into an 'institutional void', where there are few agreed structures or rules that govern them. Their novelty and ubiquity mean that new applications and their diffusion take place globally in ways that conventional policy find difficult to control (Hajer, 2003). Callon et al. (2011) use the metaphor of science and technology 'overflowing' the boundaries of existing scientific regulatory institutional frameworks and describe this context as one of relative 'lawlessness'.

Third, current forms of regulatory governance offer little scope for broad ethical reflection on the purposes of science or innovation and can do little to identify in advance many of the most profound impacts that may emerge through innovation (Stilgoe et al., 2013).

And fourth, and finally, current forms of technology governance also provide little scope for reflection on the deeply interconnected systems of politics, economics, gender, education and religion in which technology systems are embedded and which deeply influence the path dependence of technological development.

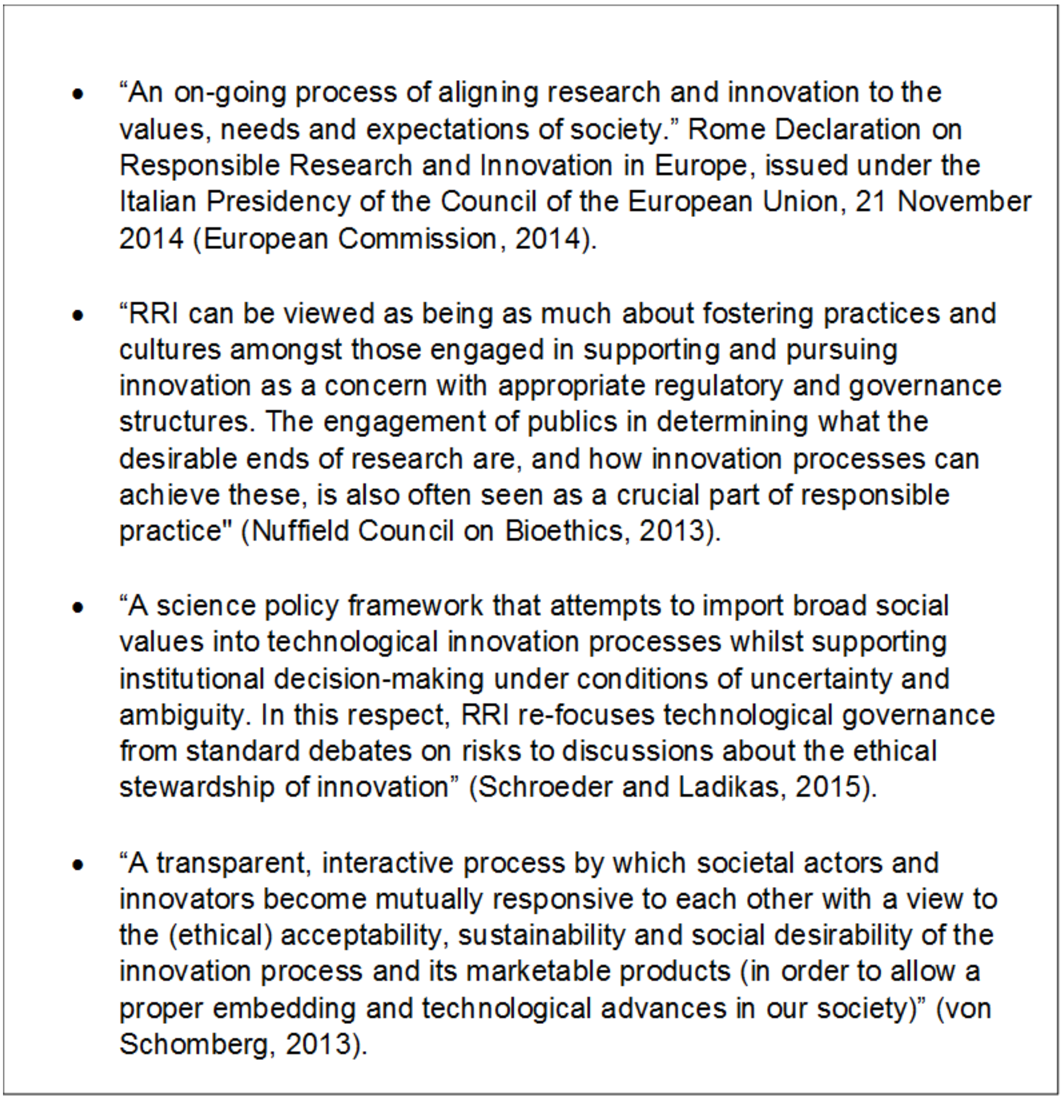
3 Overview of RRI as a Process to Address the Diverse Challenges of Technology Governance

To address some of the concerns around the governance of STI, the European Commission has sought to institutionalise notions of “responsibility”, “responsible development” and “responsible innovation” for science and technology under the banner of “Responsible Research and Innovation” (Tancoigne et al., 2016). Other international organisations including UNESCO have adopted Responsible Research and Innovation as a tool for steering entire innovation systems in ethical directions and as a means of addressing global challenges such as climate change. As a general matter, the RRI framework aims to widen the scope of formal processes of ethics review for research and innovation into a more open approach that addresses wider societal implications of science, services and products. Given this array of demands on RRI, this section turns to briefly outline the range of principles and practices that currently constitute RRI with a view to judging whether existing RRI can meet the critical challenges demanded of it.

There is not one canonical definition or approach to RRI (see Table 1 below). As a general matter, RRI can be seen as the attempt to design and implement an inclusive process for ongoing collective deliberation and decision making to address the multiple and diverse goals that research and innovation are presenting, and to constrain technology development within agreed ethical boundaries.

In one influential exploration of the field, Stilgoe et al. (2013) propose four dimensions of responsible innovation - anticipation, reflexivity, inclusion and responsiveness - to provide a framework for raising, discussing and responding to the diverse challenges raised by many new and emerging technologies.

Table 1: Understandings of “Responsible Research and Innovation”



- “An on-going process of aligning research and innovation to the values, needs and expectations of society.” Rome Declaration on Responsible Research and Innovation in Europe, issued under the Italian Presidency of the Council of the European Union, 21 November 2014 (European Commission, 2014).
- “RRI can be viewed as being as much about fostering practices and cultures amongst those engaged in supporting and pursuing innovation as a concern with appropriate regulatory and governance structures. The engagement of publics in determining what the desirable ends of research are, and how innovation processes can achieve these, is also often seen as a crucial part of responsible practice" (Nuffield Council on Bioethics, 2013).
- “A science policy framework that attempts to import broad social values into technological innovation processes whilst supporting institutional decision-making under conditions of uncertainty and ambiguity. In this respect, RRI re-focuses technological governance from standard debates on risks to discussions about the ethical stewardship of innovation” (Schroeder and Ladikas, 2015).
- “A transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and social desirability of the innovation process and its marketable products (in order to allow a proper embedding and technological advances in our society)” (von Schomberg, 2013).

Anticipation involves systematic thinking on possible consequences of research and innovation. Techniques that have been developed to embed anticipation in R&I processes include upstream public engagement (Wilsdon and Willis, 2004), Constructive Technology Assessment (Rip et al., 1995), and ‘Real-Time Technology Assessment’ (Guston and Sarewitz, 2002). These techniques involve anticipatory discussions of possible and desirable futures.

Reflexivity. According to Schuurbiers (2011), RRI requires deep reflexivity, in which the value systems and theories that shape science, innovation and their governance are themselves scrutinised. Such self-scrutiny and critique need to be public and conducted not only by researchers and innovators, but also at a systemic level involving all stakeholders who are part of the R&I system, including citizens.

Inclusiveness. The participation of diverse stakeholders, including citizens, is a central tenet of RRI. Techniques aimed at including end users in the innovation process include user-driven innovation (Hippel, 2005), open innovation (Chesbrough, 2003), open source innovation (Raymond, 1999), participatory innovation (Buur and Matthews, 2008) and networked innovation (Powell et al., 1996). Small-group processes of public dialogue, described as 'mini-publics' by Goodin and Dryzek (2006), have also been developed including consensus conferences, citizens' juries, deliberative mapping, deliberative polling and focus groups. Such mini-publics aim to include public deliberation as an upstream input in the innovation process.

Responsiveness. Responsiveness means that technology governance responds effectively to the knowledge that results from improved anticipation, reflexivity and inclusion. The mechanisms that might enable responsiveness include, for example, the application of the precautionary principle, a moratorium, or a code of conduct. The implementation of such mechanisms requires hard choices, including challenging dominant norms and values, and overcoming powerful interests that advocate particular technological solutions. RRI requires systemic application of these principles and holistic coordination of multiple individual mechanisms along the entire innovation process. As Stilgoe et al. emphasise, institutional commitment to a framework that integrates all four dimensions (anticipation, reflexivity, inclusion and responsiveness) is vital for effective technology governance within agreed ethical boundaries.

4 Models of Technological/Societal Change

While individual mechanisms such as mini-publics, research integrity, risk management and other RRI instruments may target parts of the governance of innovation, they do not represent a coherent and effective RRI governance system unless they are aligned with and work systemically with other RRI mechanisms. Indeed, as others have pointed out, this is a significant challenge (Stilgoe et al., 2013). The possibility of such coordination of multiple instruments for effective RRI governance is in some doubt. In the following Section we therefore step back to outline two existing models of socio-technical transition and introduce our new model of Deep Institutional Innovation for Sustainability and Human Development (DIIS) which seeks to position technology development within a holistic context.

4.1 Evolution of R&I Policy in the Face of Grand Challenges

For decades, innovation policy makers have been developing innovation models and policy instruments to target investments in science and technology to maximise the impacts of those investments. Until recently, the dominant model of technological change was the so-called linear model of innovation, whereby governments play an active role in financing scientific research on the premise that new scientific discoveries will be taken up by firms to produce new technologies, new industries, economic growth and jobs (Godin, 2006). Over time, this linear model was supplemented by the national innovation system (NIS) approach to innovation in which creating linkages between the various actors in the system, along with building their innovative capacities, are critical (Nelson, 1993).

In the last decade or so, owing to the scale of contemporary grand challenges, the linear model and the NIS model have increasingly been supplemented by a third innovation model, namely the model of System Innovation (SI). While the linear model and NIS both aim to strengthen and enhance the productivity of an existing innovation system, SI recognises that many of our current socio-technical systems are no longer sustainable, and that the optimization of existing systems is no longer sufficient. Instead, the SI approach aims to bring about fundamental change in the socio-technical systems that provide us with energy, food, and transport, among others.

The fundamental reorientation of R&I policy currently taking place has also been accompanied by the emergence of new models for understanding socio-technical transitions, most prominent among them the Multi-Level Perspective (MLP) (Geels, 2005, Grin, et al., 2010.), and the model of Deep Transition (Schot and Kanger, 2018).

4.2 The Multilevel Perspective (MLP) of Socio-technical Transition

The Multilevel Perspective (MLP) focuses on understanding large scale and long-term shifts that take decades to unfold, from one socio-technical system to another. Socio-technical systems are defined as configurations of actors, technologies and institutions that fulfil critical societal functions, such as the energy system, the food system, the transport system etc., that form the material backbone of modern societies.

The basic components making up the multi-level framework are niches, socio-technical regimes and socio-technical landscape, and socio-technical change is typically seen as resulting from interactions at each of these three levels:

1. The micro-level involves innovative experiments, for example by firms and communities developing and adopting new technologies and lifestyle practices. Compared to dominant regimes, the actors in niches are few, their interrelations are limited, and the new technologies and practices are still developing.
2. The meso - level is the existing technological paradigm, for example the existing technologies and practices that make up the current fossil-based energy system.
3. The macro-level landscape comprises high level mega-trends, rules and values, such as long-term changes in technology, the social acceptability of technologies, and the political landscape that supports or opposes change. As Schot and Kanger (2018) point out, this varied set of factors can be combined in a single 'landscape' category because they form an external context that

niche and regime actors cannot influence in the short run, but that do influence activities at the niche and paradigm levels.

According to the MLP model of system change, for new innovations to break through and replace an existing paradigm, multiple policies are needed to overcome current technological infrastructures and practices. A key insight of the MLP perspective is that the transition from one socio-technical system to another results from the interaction of events on all three levels—niche, regime and landscape—and occurs through a specific combination and sequence of endogenous and exogenous sources of change.

4.3 Deep Transition

Schot and Kanger (2018) build on the MLP perspective to provide additional insights into an understanding of long-term socio-technological change. In their model of Deep Transition, Schot and Kanger point out that individual socio-technical systems (for food, energy, transport, production etc) are not free standing, but are instead interconnected, particularly in terms of the meta-rules that are common across them. Their model focuses on understanding the parallel evolution of multiple (as opposed to single) socio-technical systems, complexes of socio-technical systems, and the resulting broader and long-term transformations of industrial societies as a whole.

The Deep Transition model plays particular attention to the role of rule-systems (called regimes and meta-regimes) in driving the directionality of the entire process. Schot and Kanger's hypothesis is that throughout the centuries' long process of industrialization, sociotechnical systems have generated their own macro-level selection environment that impacts on the evolution of individual socio-technical systems. They call this macro-level selection environment, which has been evolving since the Industrial Revolution to become the dominant contemporary socio-technical selection environment, 'Industrial Modernity'.

A Deep Transition is formally defined as a series of connected and sustained fundamental transformations of a wide range of socio-technical systems in a similar direction. Kanger and Schot (2019) cite a number of examples of this directionality, e.g., the move towards increased labour productivity, mechanization, reliance on fossil fuels, resource-intensity, energy-intensity, and reliance on global value chains. Among the beliefs and guiding meta- rules of Industrial Modernity that have emerged during this time, and which continue to shape technological development, Kanger and Schot identify those set out in Table 2.

In addition, Schot and Kanger put forward the following propositions on the macro-dynamics of Deep Transitions:

- The first Deep Transition has comprised successive waves of technological development (Perez, 2002, Freeman and Louca, 2001) that have led to a long-term path dependency and a powerful selection environment, called Industrial Modernity' within which current technological developments are occurring.

Table 2: Meta-rules and values in the landscape of Industrial Modernity

Separation of Nature and Society	Modern industrial society is separate from, and above, nature
Dominance over nature	• nature as a resource to be exploited • control over nature as a desirable goal
Techno-optimism	Belief that societal problems can be solved through STI
Techno-neutrality	STI as inherently value free
Externality of environmental consequences	• Belief in limitless supply of resources • Assumption that waste is not a fundamental problem
Primacy of material progress over other forms of progress	Material progress would lead to emancipation, empowerment and self-realization
Market orientation	Belief in firms as primary drivers of innovation
Productivity	Belief that any human task should be substituted with technologies to increase productivity

- The landscape of Industrial Modernity, with its value and beliefs shown in Table 1, its reliance on fossil fuels, and relentless pursuit of productivity, mechanization, competition etc. have contributed to and intensified the twin challenges of environmental degradation and social inequality.
- The grand challenges that we now face, which have been created in large part by the processes of the First Deep Transition, cannot be definitively fixed within the framework created by this very transition.
- The Second Deep Transition towards sustainability and greater equality will only occur when there is a disruption in the meta-rules and meta-values that constitute the landscape within which technology development is occurring.

The model of Deep Transition adds a new understanding of landscape in the MLP framing and its influence on current technological development. As Schot and Kanger point out, the Deep Transition framing suggests that the expansion or optimization of existing socio-technical systems, or the stimulation of radical niches to promote transitions in single systems, will not be even remotely enough. Only when the broad selection environment of Industrial Modernity itself is transformed can it stimulate the interaction between niches, regimes and meta-regimes in a manner that would alter the directionality of evolution of the broad range of socio-technical systems which constitute the backbone of industrial societies.

In other words, what is needed for transitions to sustainability and greater social and economic equality is a rupture in Industrial Modernity, and the creation of a fundamentally different macro-level selection environment for the future evolution of socio-technical systems: a different type of, or alternative to, Industrial Modernity.

4.4 Deep Institutional Innovation for Sustainability and Human Development (DIIS)

We now present a new model for societal transition that views technology as but one of a set of deeply interconnected social institutions that are currently undergoing profound change. The Deep Institutional Innovation for Sustainability and Human Development (DIIS) model (Hughes et al. 2021) broadens further the conceptualisation of the role of technological change in transitions by placing it firmly within the context of whole of society change. The DIIS model is illustrated schematically in Figure 1.

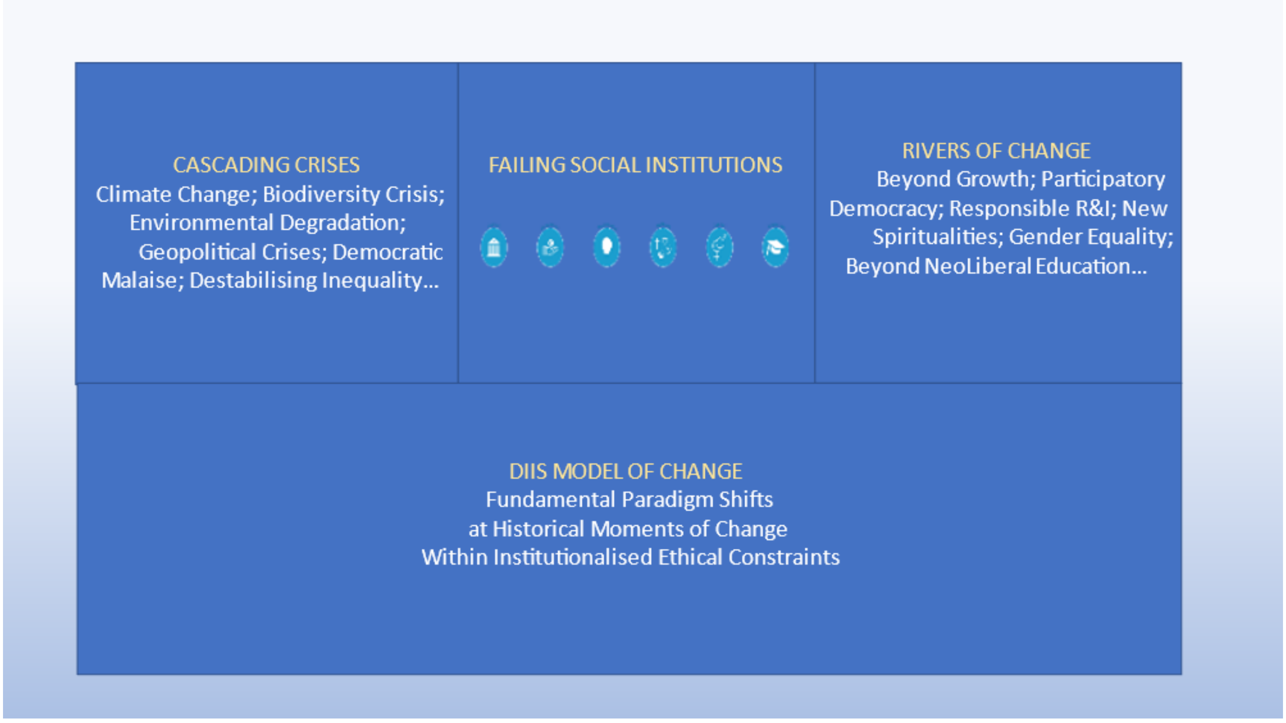

CASCADING CRISES
Climate Change; Biodiversity Crisis; Environmental Degradation; Geopolitical Crises; Democratic Malaise; Destabilising Inequality...
FAILING SOCIAL INSTITUTIONS
RIVERS OF CHANGE
Beyond Growth; Participatory Democracy; Responsible R&I; New Spiritualities; Gender Equality; Beyond NeoLiberal Education...
DIIS MODEL OF CHANGE
Fundamental Paradigm Shifts
at Historical Moments of Change
Within Institutionalised Ethical Constraints

Figure 1: Model of Deep Institutional Innovation for Sustainability and Human Development (DIIS)

The top left part of Figure 1 represents the ‘cascading crises’ that currently afflict humanity. These crises include climate change, extreme biodiversity loss, environmental degradation, destabilising levels of poverty and economic inequality, persistent levels of social and economic inequality, the rise of authoritarian populism, the erosion of democracy, the challenges of digitalisation and emerging technologies, rising international tensions, and the prospect of globally devastating wars.

The top central portion of Figure 1 represents six major social institutions that have traditionally provided stability and direction for societies. While there is no single definitive definition of social institution, the DIIS model adopts the following broad characteristics of social institutions: they play a central and important role in society; they are typically meta-institutions, i.e., systems of ideas, organisations and practices; and being central and important to a society, they are usually long lasting, typically trans-generational. In the DIIS framing, social institutions are taken to span the ideological, material organisational, and 'social practices' aspects of the social systems considered (see Glatz-Schmallegger et al., 2021). In general, social institutions are characterized by continuity, pattern maintenance and social reproduction, rather than by deep structural change, innovation or transformation. The DIIS model identifies six major social institutions: politics, economics, technology, religion, gender, and education.

Since technology spans ideological, material institutional, and social practices aspects, technology qualifies within the DIIS framing as a major social institution. Although technology may change and develop, and in fact, as we will argue, will need to change for the transition to sustainability to take place, the ideology and system aspects of innovation have remained constant for some time. As the Deep Transition framework, as well as work in the field of Science and Technology Studies suggests, stable technological production systems, assumptions and ideologies help constitute our contemporary moment of industrial modernity (e.g., Jasanoff 2004).

That the technology system is deeply intertwined with the other social systems considered in the DIIS model is readily apparent from contemporary discussions within RRI itself. The interconnection between technology, economics and politics is clear, for example, in the central role of technological innovation in economic growth and in the role of governments in setting framework conditions for innovation, including regulation. The intersection of technology and religion is apparent in ethical debates on technologies including gene editing and potential developments in neurotechnology. Gender and technology is an area of focus in RRI, including the consequences of gender imbalances on the outcomes of research and innovation. Education too is deeply interconnected with technology in multiple ways, including the current dominant educational paradigm of skills production for jobs and economic growth.

The DIIS model is focused on understanding the processes of deep structural and functional change within social institutions at historic tipping points, such as the present. The model posits that existing social institutions have both contributed to the cascading series of crises in Figure 1, and are incapable in their present form of resolving the crises they have contributed to creating.

The top left portion of Figure 1 reflects the fact that global movements are already in place that are advocating for, and taking action to create, social institutions more aligned with goals of sustainability and more equitable human development, and more able to address the crises facing humanity. In the realm of politics, movements for participatory democracy, for greater protections against authoritarianism, and for the inclusion of the interests of future generations are growing in strength. In economics, there has been an increasing acknowledgment of the shortcomings of the current dominant neoliberal economic model (OECD, 2020), alongside the development of alternative economic models (Raworth, 2017; Jackson, 2019; Folbre, 2008). In technology, as this paper attests, concerns about the adverse impacts of technologies have prompted the development of normative frameworks and practices such as RRI, as well as movements advocating greater foresight, regulation and accountability of transformative technologies. While religion is often viewed as immutable, the refiguration of religion and spirituality in the context of contemporary cascading crises, including climate change, is evidenced by the contemporary proliferation of religious schisms and emerging spiritualisms. Gender equality, and the social construction of gender, are pervasive issues across all of the major social institutions in society (Smiler, 2019). Finally, education plays a foundational role in terms of enabling (or preventing) deep system change. Education, particularly higher education, can either replicate the status quo in terms of paradigms of knowledge, epistemology, methodologies etc., or can act as enabling institutions, from within which deep system change may emerge. In this regard, movements which question the compartmentalisation of knowledge and research, and the dominant ‘skills-based’ paradigm of education, are gaining momentum in higher education internationally.

The bottom portion of Figure 1 represents the underlying DIIS model for understanding deep societal transformation that results from the simultaneous transformation of multiple contemporary social institutions. This model is outlined in detail in a previous paper (Hughes et al., 2021) and can only be briefly outlined here. The DIIS model posits that deep societal transformations occur at specific moments in history when underlying changes lead to tipping points that necessitate whole-of-society systemic change. It is the premise of the model that we are now at such a historical tipping point. Such moments of change are characterized by the breakdown of social institutions, experienced as periods of liminality, extreme contestation, social unrest and deep institutional innovation. A key focus of the DIIS model is that at such historical moments, the prevalence of particular sources of danger, if they are not restrained within institutionalised ethical constraints, can tip the balance of the transformation to outcomes that are severely detrimental to the public good. Such potential sources of danger include destructive leadership, ideologies of exclusion and blame, and social institutions that prioritise dominance values of hierarchy, inequality, coercion and private gain, over partnership values of equity, cooperation, and public good (Eisler and Fry, 2019).

The four components in Figure 1 together indicate the need for a whole of society paradigm shift, and that such a shift must be constrained within ethical boundaries if it is to result in outcomes that contribute to the common good, particularly increased sustainability and human development.

In summary, the DIIS model can be stated in four primary axioms:

1. achieving planetary sustainability requires deep institutional change across multiple social institutions (technology, politics, economics, gender, religion, education)
2. ongoing deep institutional changes are currently occurring across these social institutions, spurred by, and occurring within the context of, multiple cascading contemporary crises
3. transformation in one institutional arena (e.g., technology) is occurring in deeply coupled interactions with other social institutions (politics, economy, gender, religion, education), and
4. transformations, both within individual social institutions and their summation across society, must be constrained within ethical boundaries if the outcomes are to be in the direction of increased sustainability and greater human development.

5. Discussion and Conclusion—Implications of DIIS for RRI

RRI has emerged as a set of principles and practices aimed at enabling the governance of new and emerging technologies for the public good. The concerns that RRI seeks to address are diverse and critical, ranging from safety concerns over the application of new technologies, their environmental and societal consequences, the potentially existential threats that some technologies may pose, the desire to steer R&I towards the solution of existential threats such as climate change, and the imperative in democratic societies that citizens should have an input into decisions that may profoundly impact their lives. RRI has particular significance at the present moment given the range of crises that humanity is facing, most or all of which involve technology both as their cause and (in part) their potential solution. Current RRI approaches are based on accepted principles of anticipation, reflexivity, inclusion and responsiveness, and comprise multiple instruments which target individual parts of the R&I process. Effective RRI however requires systemic application of these principles and holistic coordination of multiple individual mechanisms along the entire innovation process. Such a coordinated approach has been difficult to achieve because RRI arguably lacks a theory of complex institutional systems. Here the RRI model of change could be enriched by the perspectives on system change outlined above.

The MLP, Deep Transition and DIIS models build on each other, supporting the deep understanding needed to address the cascading crisis we face. The MLP highlights the fact that transition in individual socio-technical systems occurs through the dynamic interplay of changes at the multiple levels of niche, paradigm and landscape. The Deep Transition model shows how the historical development of the multiple sociotechnical systems that underpin modern societies has resulted in a landscape of meta-rules and values that constitute Industrial Modernity. These meta-rules and values in turn acts as a deeply constraining influence on the further development of technology. The DIIS model aims to highlight the fact that technology, in turn, is part of a much wider set of social institutions, comprising politics, economics, gender, religion, and education, which also deeply influence the development pathways of technology. The models suggest therefore that the levers of action for RRI are more varied and pervasive, but also more institutionally embedded and potentially more intransigent. While the MLP and Deep Transitions frameworks both conceptualize the embeddedness of technology and innovation in wider social systems, the DIIS model approaches the technological system as inextricably intertwined with, influencing, and influenced by the other social systems which DIIS is addressing. This parallel conceptualisation and analysis of multiple systems in the DIIS approach significantly broadens both the research questions that present themselves and the transdisciplinary inclusiveness of the analysis required. For the purposes of the current paper, for example, the DIIS approach widens Schot and Kanger's conceptualisation of landscape from 'Industrial Modernity' to a much broader framing of landscape as 'Postmodernity', which can include sociological, psychoanalytical, and feminist perspectives, among others, in the framing. Further development of the DIIS model will aim to acquire a deeper understanding of the systemic intersectionality of the various social institutions included in the DIIS framing (Choo and Ferree, 2010), as well as further research on the simultaneous ideational, structural and process elements of multiple system change.

The perspective brought by the Deep Transition and DIIS models suggests three areas of further research for existing RRI that could potentially deepen the capacity of RRI as norms and practices to help steer technology in ways that support social change for the public good.

First, a potential limitation of RRI is that it views the world from within the STI system. Current RRI practices operate within an incomplete model of change and a limited view of complexity. This arguably limits RRI's capacity to enhance the responsiveness to science and technology, one of RRI's core values as discussed above. True responsiveness requires a systemic diagnosis of the problem and likely solutions. The capacity to respond is limited if the scope is too trained upon itself. Further discussion on the wider whole-of-society contexts within which RRI is being developed and applied, as highlighted by the DIIS model, is warranted.

Second, RRI focuses on behaviour at the niche rather than structural systemic level. As a result, the RRI agenda, as important as it is, may be frustrated in its aspirations to help drive critical R&I transitions. DIIS, alongside the MLP and Deep Transition models, by showing the multi-level and inter-connected nature of the major social systems influencing technology, could provide new and important levers and targets for RRI.

Finally, it is questionable whether RRI—as it focuses squarely on the STI system alone—can aim at the deep systemic changes that are needed for transitions without rethinking innovation itself. Both the Deep Transition and DIIS models assert that a rupture with the landscape of meta-rules and values of industrial modernity is necessary for transitions at this historic moment. The point of intervention of existing models of RRI, which aim to bring about changes to the governance of technology within existing landscape meta-rules and values, might therefore be insufficient to effectively channel technological development for the public good. For these reasons, some commentators argue that RRI may be working with an oversimplified model of innovation itself and may need to rework some of its foundational assumptions (Blok and Lemmens 2015).

The preceding discussion suggests, therefore, that the responsiveness of the science and technology system to social challenges will require a greater appreciation of the complexity of the multiple systems that need to change and the integral role of technology within this broader context. RRI might usefully engage the theories discussed above in order to deepen its capacity to affect social changes that are sorely needed.

Acknowledgments

The authors would like to thank Professor Brian Ó Gallachóir for helpful discussions and insights during the preparation of this paper, as well as the wider DIIS team on whose work this paper builds.

References

Abduljabbar, R., Dia, H., Liyanage, S. and Bagloee, S.A., 2019. Applications of artificial intelligence in transport: An overview. *Sustainability*, 11(1), p.189.

Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y. and Chen, H., 2021. Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. *Journal of Cleaner Production*, p.125834.

Blok, V. and Lemmens, P., 2015. The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In *Responsible innovation* 2 (pp. 19–35). Springer, Cham.

Bostrom, N., 2014. Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press.

Burget, M., Bardone, E. and Pedaste, M., 2017. Definitions and conceptual dimensions of responsible research and innovation: A literature review. *Science and engineering ethics*, 23(1), pp.1–19.

Buur, J. and Matthews, B., 2008. Participatory innovation. *International Journal of Innovation Management*, 12(03), pp.255-–73.

Callon, M., 1987. Society in the making: the study of technology as a tool for sociological analysis. *The social construction of technological systems: New directions in the sociology and history of technology*, pp.83–103.

Callon, M., Lascoumes, P. and Barthe, Y., 2011. *Acting in an uncertain world: An essay on technical democracy*. Inside Technology.

Chesbrough, H.W., 2003. *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press, Boston, MA.

Choo, H.Y. and Ferree, M.M., 2010. Practicing intersectionality in sociological research: A critical analysis of inclusions, interactions, and institutions in the study of inequalities. *Sociological theory*, 28(2), pp.129–149.

Cockburn, I.M., Henderson, R. and Stern, S., 2018. *The impact of artificial intelligence on innovation* (No. w24449). National bureau of economic research.

Eisler, R. and Fry, D.P., 2019. *Nurturing our humanity: How domination and partnership shape our brains, lives, and future*. Oxford University Press.

European Commission, 2014. *Rome Declaration on Responsible Research and Innovation in Europe*, https://ec.europa.eu/digital-single-market/en/news/rome-declaration-responsibleresearch-and-innovation-europe (Accessed on 28 August 2017).

Fjelland, R., 2020. Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7(1), pp.1–9. https://doi.org/10.1057/s41599-020-0494-4.

Folbre, N., 2008. *Valuing children: Rethinking the economics of the family*. Harvard University Press.

Freeman, C., and Louçâ, F., 2001. *As time goes by: from the industrial revolutions to the information revolution*. Oxford University Press.

Geels, F.W., 2005. Processes and patterns in transitions and system innovations: Refining the co-evolutionary multi-level perspective. *Technological forecasting and social change*, 72(6), pp.681–696.

Glatz-Schmallegger, M., Byrne, E.P., Mullally, G., Hughes, I., McGookin, C. and Ó Gallachóir, B.P., 2021. Social innovation and deep institutional innovation for sustainability and human development. *Open Cultural Studies*, 47, pp.176–192.

Godin, B., 2006. The linear model of innovation: The historical construction of an analytical framework. *Science, Technology, & Human Values*, 31(6), pp.639–667.

Goodin, R.E. and Dryzek, J.S., 2006. Deliberative impacts: The macro-political uptake of mini-publics. *Politics & society*, 34(2), pp.219–244.

Goodman, K., Zandi, D., Reis, A. and Vayena, E., 2020. Balancing risks and benefits of artificial intelligence in the health sector. *Bulletin of the World Health Organization*, 98(4), p.230.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B. and Evans, O., 2018. When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, 62, pp.729–754. https://perma.cc/2K2D-LE3A.

Grin, J., Rotmans, J. and Schot, J., 2010. *Transitions to sustainable development: new directions in the study of long term transformative change*. Routledge.

Guston, D.H. and Sarewitz, D.,2002. Real-time technology assessment. *Technology in society*, 24(1–2), pp.93–109.

Hajer, M.,2003. Policy without polity? Policy analysis and the institutional void. *Policy sciences*, 36(2), pp.175–195.

Haner, J. and Garcia, D., 2019. The artificial intelligence arms race: trends and world leaders in autonomous weapons development. *Global Policy*, 10(3), pp.331–337.

Hippel, E. V., 2005. Democratizing innovation: The evolving phenomenon of user innovation. *Journal für Betriebswirtschaft*, 55(1), 63–78.

Hughes, I., Byrne, E., Glatz-Schmallegger, M., Harris, C., Hynes, W., Keohane, K. and Gallachóir, B.Ó., 2021. Deep institutional innovation for sustainability and human development. *World Futures*, pp.1–24.

Jackson, T., 2019. The post-growth challenge: secular stagnation, inequality and the limits to growth. *Ecological economics*, 156, pp.236–246.

Jasanoff, S., 2004. *States of Knowledge: The Co-production of Science and Social Order*. London: Routledge.

Kanger, L., & Schot, J., 2019. Deep transitions: Theorizing the long-term patterns of socio-technical change. *Environmental Innovation and Societal Transitions*, 32, 7–21.

Nedelkoska, L. and Quintini, G., 2018. *Automation, skills use and training*. OECD Social, Employment and Migration Working Papers, No. 202, OECD Publishing, Paris.

Nelson, R.R. ed., 1993. *National innovation systems: a comparative analysis*. Oxford University Press on Demand.

Nuffield Council on Bioethics, 2013. *Novel neurotechnologies: intervening in the brain*, Nuffield Council on Bioethics, https://www.nuffieldbioethics.org/wp-content/uploads/2013/06/Novel_neurotechnologies_report_PDF_web_0.pdf

OECD, 2018, *Putting faces to the jobs at risk of automation*, Policy Brief on the Future of Work, OECD Publishing, Paris https://www.oecd.org/future-of-work/Automation-policy-brief-2018.pdf.

OECD, 2019, *Artificial Intelligence in Society*, OECD Publishing, Paris.

OECD, 2020. *Beyond growth: towards a new economic approach*. OECD Publishing.

O'Keefe, C., Cihon, P., Flynn, C., Garfinkel, B., Leung, J., and Dafoe, A., 2020. *The Windfall Clause: Distributing the Benefits of AI*. Centre for the Governance of AI Research Report. Future of Humanity Institute, University of Oxford. Available at: https://www.fhi.ox.ac.uk/windfallclause/.

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(6), pp.751–760.

Owoc, M.L., Sawicka, A. and Weichbroth, P., 2021. Artificial Intelligence Technologies in Education: Benefits, Challenges and Strategies of Implementation. arXiv preprint arXiv:2102.09365.

Perez, C., 2002. *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Edward Elgar, Cheltenham, UK.

Powell, W.W., Koput, K.W. and Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative science quarterly*, pp.116–145.

Raworth, K., 2017. *Doughnut economics: seven ways to think like a 21st-century economist*. Chelsea Green Publishing.

Raymond, E., 1999. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, O'Reilly and Associates, Inc., Sebastopol, CA.

Rip, A., Misa, T.J. and Schot, J. eds., 1995. *Managing technology in society*. London: Pinter Publishers.

Sachs, J.D. and Kotlikoff, L.J., 2012. *Smart machines and long-term misery* (No. w18629). National Bureau of economic research. https://perma.cc/MS34-B3EK.

Schot, J. and Kanger, L., 2018. Deep transitions: Emergence, acceleration, stabilization and directionality. *Research Policy*, 47(6), pp.1045–1059.

Schroeder, D. and Ladikas, M., 2015. Towards principled Responsible Research and Innovation: employing the difference principle in funding decisions. *Journal of Responsible Innovation,* 2(2), pp.169–183. http://dx.doi.org/10.1080/23299460.2015.1057798.

Schuurbiers, D., 2011. What happens in the lab: Applying midstream modulation to enhance critical reflection in the laboratory. *Science and engineering ethics*, 17(4), pp.769–788.

Stilgoe, J., Owen, R. and Macnaghten, P., 2013. Developing a framework for responsible innovation. *Research policy*, 42(9), pp.1568–1580.

Tancoigne, E., Randles, S. and Joly, P., 2016. Evolution of a concept: a scientometric analysis of RRI. *Navigating Towards Shared Responsibility in Research and Innovation. Approach, Process and Results of the Res-AGorA Project.* Karlsruhe, pp.39–44. http://pure.au.dk/portal/files/98634660/RES_AGorA_ebook.pdf.

Trajtenberg, M., 2019. Artificial Intelligence as the Next GPT. *The Economics of Artificial Intelligence: An Agenda*, 175.

Van Dijk, J.A., 2017. Digital divide: Impact of access. *The international encyclopedia of media effects*, pp.1–11.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M. and Nerini, F.F., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications*, 11(1), pp.1–10.

von Schomberg, R., 2013. A vision of responsible research and innovation. Responsible innovation: *Managing the responsible emergence of science and innovation in society*, pp.51–74. http://onlinelibrary.wiley.com/book/10.1002/9781118551424.

Wilsdon, J. and Willis, R., 2004. *See-through science: Why public engagement needs to move upstream*. Demos, London.

Wright, N., 2018. How artificial intelligence will reshape the global order. *Foreign Affairs*, 10.