Futures & Foresight Science

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Comparative Analysis of Long-Term Governance Problems: Risks of Climate Change and Artificial Intelligence

Atte Ojanen[1,2]

[1]Demos Helsinki, Helsinki, Finland | [2]Department of Philosophy, Contemporary History and Political Science, University of Turku, Turku, Finland

Correspondence: Atte Ojanen (atte.ojanen@demoshelsinki.fi)

Received: 10 October 2024 | **Revised:** 10 October 2024 | **Accepted:** 30 October 2024

Funding: This paper was supported by the KT4D and TANDEM projects, which have received funding from the European Union's Horizon Europe program under grant agreement No 101094302 and No 101069653 respectively.

ABSTRACT

Comparative approaches are rarely utilized in futures studies despite the distinctive nature of different policy problems. Issues like climate change, infrastructure investments, and governance of emerging technology are frequently grouped under the umbrella of the "long-term problems" without adequate consideration for their distinct spatial and temporal attributes. To address this research gap, this paper presents a framework to systematically compare long-term policy problems, such as the risks of climate change and artificial intelligence (AI). I conduct a comparative analysis of the risks of climate change and AI—both widely regarded as pivotal questions of our time—focusing on how they differ across eight attributes that affect their governance: scientific certainty, spatiality, temporality, linearity, path dependence, accountability, capacity to address and the costs involved. The findings suggest that climate change involves a more evident intergenerational conflict between generations than risks of AI and might therefore be a more challenging long-term governance problem. Yet, both problems risk triggering irreversible lock-in effects, specifically in extreme scenarios such as crossing climate tipping points or misaligned advanced AI systems. Mitigating these uncertain lock-in effects requires precautionary governance measures, highlighting the potential of comparative approaches at the intersection of foresight and policy analysis.

1 | Introduction

There is a growing recognition of the long-term policy problems that modern governments face. Addressing issues such as climate change, infrastructure investments, pension system, pandemic preparedness, and governance of artificial intelligence (AI) frequently require policymakers to pay upfront costs in anticipation of potential benefits in the future (Jacobs 2016). Arguably, this makes democratic governments reluctant to invest in long-term social goods. As such, these policy issues are prone to short-termist decisions and under-investment, leading many to conclude that democracies suffer from "presentist bias" or "myopia" (Boston 2017). However, the above-mentioned issues are frequently framed as "long-term problems," without consideration for their distinct spatial, temporal, and ethical characteristics within the fields of futures studies and policy analysis. There is a need to examine the different timeframes of policy problems without reducing them to a binary of short or long-term problems (MacKenzie, Setala, and Kyllönen 2023, 270). This includes consideration of whether the problem is cumulative or intermittent in nature, how visible and linear it is and how are the costs and benefits of the policy investment distributed both within and between generations.

Such policy problem-oriented approach to long-term governance is instructive because uniform interventions (e.g., better foresight mechanisms) are unlikely to function effectively across different issues. Accurately structuring the policy

---

Futures & Foresight Science, 2024; e203
https://doi.org/10.1002/ffo2.203

problem and its key characteristics is a prerequisite for anticipating and mitigating related future risks. Yet, there remains a gap between futures studies and policy analysis (van Dorsser et al. 2018, 2020). For example, horizon scanning (Cuhls 2020) over the possible future trajectories of the issue could be effectively integrated into the problem definition phase of policy analysis (Patton, Sawicki, and Clark 2015). While the field of Technology Assessment (TA) has made strides in connecting foresight to policy analysis, its implementation has been inadequate, evident in the defunct institutions dedicated to it. Moreover, TA's initial emphasis on technocratic forecasting and reactive cost–benefit analysis arguably neglected the broader, longer-term normative viewpoint of futures studies (see van Lente, Swierstra, and Joly 2017). As such, alternatives like anticipatory governance have gained prominence, referring to the institutional capacity and preparedness to manage possible societal disruptions, while this is still possible (Guston 2014, 219). Anticipatory governance practices seek to systematically utilize foresight to reduce risk and uncertainty in decision-making, especially in relation to emerging technologies. In the context of long-term governance, the anticipation is directed at problems that are intergenerational by their nature.

This article presents a comparative framework for analyzing attributes of different long-term policy problems and demonstrates it by systematically comparing the risks of climate change and artificial intelligence. Comparative attention will be given to intergenerational justice (Meyer 2021) because of the long-lasting consequences of both issues. The main research question is, first, how do attributes such as scientific uncertainty, temporality, and linearity differ between climate and AI risks, and second, what implications this has for their long-term governance. Both problems are widely recognized as two of the most pressing global catastrophic risks with comparable short-term and long-term risks (Ord 2020; Beard et al. 2021; Vold and Harris 2021; Bucknall and Dori-Hacohen 2022). This is also exemplified by the recent statements by notable AI figures such as "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." (CAIS Center for AI Safety 2023). While such claims warrant skepticism, climate change and AI do share key common features from a long-term perspective, such as potentially cascading future risks. Regardless of their differences, both are pressing problems that need to be compared to make sensible investments into their mitigation and governance.[1] This requires prioritization of limited resources both between the problems and their governance solutions, as noted by the emerging field of global priorities research (Greaves et al. 2020).

I will first outline how the risks of climate change and artificial intelligence relate to intergenerational justice and long-term governance.[2] Next, a comparative framework for contrasting these two policy problems will be presented. Risks of climate change and AI will be analyzed across eight attributes that affect their governability: scientific certainty, spatiality, temporality, linearity, path dependence, accountability, capacity to address, and the costs involved. Afterward, the results of the comparative analysis will be discussed, with specific focus on precautionary governance of lock-in risks, that may violate future generations' self-determination. The article seeks to make a novel contribution to the intersection of foresight and policy analysis by demonstrating how comparisons between policy problems can reveal important differences that inform their long-term governance.

2 | Background

Climate and technology policies of today significantly affect the future, prompting us to consider our duties towards future generations. This has generally been approached as a distributive question, such as guaranteeing a minimum living standard for future generations (sufficientarianism), equalizing the welfare, resources, or capabilities across generations (egalitarianism), or improving the conditions of the worst-off generations (prioritarianism) in theories of intergenerational justice (González-Ricoy 2019, 3). Recently, scholars have proposed approaching intergenerational equity as a question of nondomination between generations (see Beckman 2016).[3] For instance, impacts of climate change not only threaten future generations' well-being materially but also irreversibly limit their autonomy and options, thereby arbitrarily dominating them (Nolt 2011). This becomes especially critical in extreme scenarios, such as crossing climate tipping points that irreversibly lock-in certain outcomes.

Path dependence and lock-in are pivotal concepts that describe the persistence and rigidity of societal trajectories. Path dependence refers to processes where even relatively insignificant technological and economic developments of the past affect future decisions disproportionately, based on mechanisms such as increasing returns and positive feedback effects (Pierson 2000; Cairns 2014). One supposed example of this is the persistence of the QWERTY keyboard layout. Lock-in represents a more stringent and strong form of path dependence, where a particular technology or policy becomes entrenched, making alternative, more sustainable options irreversibly inaccessible in the future (Goldstein et al. 2023). One of the most well-known forms of this is carbon lock-in, in which the initial conditions, economic returns to scale, and social dynamics prevent the deployment of low-carbon solutions (Seto et al. 2016). The lock-in of existing carbon-intensive technologies and infrastructure constrains efforts to reduce carbon emissions. The permanence of lock-in effects is contested issue, with Shackley and Thompson (2012) calling for more nuanced understanding of lock-in, differentiating between shallow and deep lock-in. Nevertheless, minimizing the risks of lock-in, defined as virtually irreversible institutional, infrastructural, or technological path dependence, is a central concern in long-term governance.

These issues are particularly stark with climate change due to its global and intergenerational nature (Gardiner 2011). Global warming caused by greenhouse gas (GHG) emissions is not restricted by national boundaries, but the accumulation of emissions affects the earth's atmosphere as a whole. These carbon emissions remain in the atmosphere for centuries, contributing to warming and increased risks of droughts, depletion of ecosystems, food and water scarcity, diseases, and social instability. Climate change is essentially a back-loaded phenomenon where emissions have immediate and concrete benefits for current people in the form of inexpensive energy and products, but most of the serious costs are deferred to future generations. This is what some have termed the "tyranny of the contemporary" or "intergenerational buck-passing" (Gardiner 2011). This dynamic can ultimately trigger tipping points where severe climate issues of today are transformed into irreversible ones.

A climate tipping point is a critical threshold, which if crossed, triggers a large and often irreversible shift in the climate system. Ecosystems, ice sheets, ocean, and atmospheric currents can all exhibit self-perpetuating tipping behavior with severe consequences for society (Lenton et al. 2019, 2023). Out of the nine global and seven regional tipping elements identified, ice sheet collapse of Greenland and West Antarctic, die-off of tropical coral reefs, and abrupt thaw of boreal permafrost are already estimated to be likely to pass a tipping point if the global warming exceeds 1.5°C (Armstrong McKay et al. 2022). Recent studies suggest Atlantic Meridional Overturning Circulation (AMOC) is nearing tipping point this century as well (van Westen, Kliphuis, and Dijkstra 2024). Crossing these tipping points could already lock-in major harms for future generations such as massive sea level rise over thousands of years and decimation of key ecosystems. Once considered unlikely at low levels of warming, tipping points now pose risks already at 1°C, escalating at 2°C, and becoming severe between 2.5°C and 4°C of warming (OECD 2022). A major risk is that crossing one climate threshold may trigger a "domino effect" of cascading tipping points, leading to catastrophic environmental outcomes. However, the timing and probability of climate tipping points remain highly uncertain, with many expected to unfold over centuries or millennia.[4]

Similarly, responsible governance of artificial intelligence and its risks has rapidly emerged as one of the pressing challenges to long-term governance. AI refers to advanced machine-based systems that operate with a degree of autonomy to produce predictions, recommendations, or decisions (see OECD 2019). The current AI paradigm relies on machine learning, specifically reinforcement learning from large training datasets and substantial amounts of compute, as exemplified by large language models (LLMs) such as GPT-4. AI systems already pose significant societal challenges in terms of bias, disinformation, and privacy (Eubanks 2018), while also raising uncertain long-term safety and security implications as they advance. The current 6-9 month doubling trends in data, compute, and algorithmic efficiency point to significantly more powerful AI systems in the future (Epoch 2023). Training, deployment, and societal diffusion of AI models all give rise to risks— misalignment, misuse, and systemic risks in society—across different timescales. Advanced general-purpose AI models could have emergent capabilities like situational awareness that pose risks to public safety either via misuse or accident (Bengio, 2024). This includes deliberately using AI models to target key infrastructure or to produce dangerous pathogens, as well as loss of control over more agentic models (Kilian, Ventura, and Bailey 2023). Such frontier AI models are a challenging regulatory target because their capabilities might arise unexpectedly, they are prone to misuse when deployed and they can proliferate rapidly, making accountability difficult (Anderljung et al. 2023).

In terms of technology and AI, lock-in has been used to refer to situations where certain technologies become so widely entrenched in societal infrastructure that it is extremely challenging and costly to change paths (Shapiro and Varian 1999). Especially general-purpose technologies, such as the steam engine, electricity, computer—and now potentially AI—have had such transformative long-term effects on society. In the short term, AI systems may create lock-in effects by concentrating power in the hands of a few dominant entities, perpetuating existing injustices, and influencing human preferences. In the longer term, radically transformative AI such as artificial general intelligence (AGI) exceeding human capabilities could potentially lock-in certain values (see Kilian, Ventura, and Bailey 2023). These lock-in risks can vary dramatically depending on how likely the development of human-level AI is seen. A 2023 survey with 2778 AI researchers concluded that there is 50% probability of high-level machine intelligence that outperforms humans in every possible task by 2047, which is 13 years earlier than the previous year's survey (Grace et al. 2024). Significantly shorter AGI timelines have also been espoused recently, such as the median estimate of 2031 in Metaculus (2024) forecasting platform. However, the practical consequences of advanced AI systems are poorly understood and reliant on contested estimates about "takeoff speeds."[5] Moreover, even less transformative AI, like lethal autonomous weapons, could lead to major lock-in effects and irreversibly change narrow societal domains (Gruetzemacher and Whittlestone 2022).

Despite the uncertainties, the risks of climate change and AI are perhaps the two most well-recognized global catastrophic risks, alongside threats such as nuclear weapons and pandemics (Ord 2020). Both problems share similar origins as unintended negative externalities of otherwise economically beneficial innovations. Climate change is a result of burning fossil fuels for energy, which has been the backbone of global economic growth since industrialization. AI promises to automate large sectors of the economy and science, leading to potentially unprecedented productivity, R&D, and innovation gains. Overall, climate change and AI share relatively comparable short-term and long-term risks that could endanger society, as demonstrated in Table 1.

In fact, both AI and climate change can be seen as the latest of historical technological progress that increases societal risks

TABLE 1 | Illustration of the most salient and severe climate and AI risks across different timeframes.

	Short-term (0–5 years)	Medium-term (5–20 years)	Long-term (20+ years)
Climate change	Extreme weather events such as heatwaves and floods	Loss of biodiversity and regional ecosystem services	Collapse of key ocean currents or permafrost
Artificial Intelligence	Biased and discriminatory automated decision-making	Concentration of power and AI-automated surveillance	Misaligned and unsafe advanced AI systems

before appropriate investments into safety. Aschenbrenner (2020), building on Jones (2016), argues that technological progress tends to on average make the world more dangerous *ceteris paribus*. As technologies mature and wealth increases, societies invest more into the safety of technologies through research, regulation, and standards, as seen with the car industry and nuclear energy. Similarly, Sandbrink et al. (2022) have argued in favor of differential technology development, which leverages risk-reducing interactions between technologies by affecting their relative timing. This principle calls on regulators to delay risk-increasing technologies such as advanced AI systems while preferentially advancing risk-reducing defensive, safety, or substitute technologies. Climate change is arguably ahead of AI risks on this technological maturity scale—as the negative externalities of fossil fuels have been recognized, more has been invested into safety and governance of these technologies, that is, renewable energy, carbon capture, and international agreements on limiting carbon emissions. Having established an understanding of the risks of climate change and AI as long-term governance challenges, we now turn our attention to examining the policy attributes that facilitate their comparative analysis.

3 | Methodological Framework

I employ a comparative approach between policy problems to discern differences in attributes or characteristics that are relevant for their long-term governance. To do so, I compose a comparative framework with eight attributes to systematically contrast different long-term policy problems such as the risks of climate change and artificial intelligence. The approach is inspired by efforts to strengthen the connection between policy analysis and foresight (see van Dorsser et al. 2018, 2020). To facilitate anticipatory governance of long-term policy problems, foresight should be systematically integrated into the policy cycle, beginning at the problem definition stage to identify differences between problems (Guston 2014). This comparative problem centric approach represents a novel and interdisciplinary method for combining horizon scanning and problem definition as the first steps of foresight and policy analysis (Patton, Sawicki, and Clark 2015) respectively. The problem definition is inherently connected to horizon scanning, referring to systematic examination of potential future problems, trends, and weak signals (Cuhls 2020). The comparative approach aids categorization of policy problems based on their level of uncertainty to enhance their anticipatory governance (Muiderman et al. 2020). Building on Walker (2011), van Dorsser et al. (2018, 2020) identify four levels of uncertainty surrounding policy problems linked to the projected, probable, plausible, and possible futures, each requiring different forms of forecasting or foresight to address. The comparative approach can enhance this process by providing a systematic framework to map, structure, and contrast the future uncertainty associated with policy problems. Moreover, it also holds value for adaptation into scenario analysis and technology foresight (Pietrobelli and Puppato 2016) to aid identification of emerging yet uncertain risks and their potential governance responses.

I base my comparative framework on existing categorizations of policy problems (Peters 2005; Jacobs 2016; Boston 2017; Hansson 2023) by compiling the most relevant attributes for long-term issues. Although such categorizations have been overlooked in foresight, they have been discussed in policy analysis and design literature. Peters (2005, 355) influentially differentiates between solubility, complexity, scale, divisibility, monetarization, scope of activity, and interdependencies as attributes of policy problems. The first three attributes refer to the policy problem itself, while the latter ones refer to the policy instruments required to address it. Contrasting climate change with pandemics, Hansson (2023) makes a distinction between causality, epistemology, temporal patterns as well as spatial and interpersonal distribution of a policy problem. Causality includes issues such as whether the policy problem is a natural phenomenon triggered by human activities and whether its solutions require government action or lifestyle changes by individuals. Epistemology relates to the degree of scientific consensus and uncertainty, disinformation, denialism, and power of private interest groups on the problem. Temporal patterns include the problems' long-term effects, urgency, and whether it is an intermittent or accumulating problem. Lastly, problem's spatial and interpersonal distribution refers to the number of people affected, distribution of costs, governance level, and the possibilities for international cooperation (Hansson 2023, 90). Arguably the last two on temporal and spatial patterns are especially relevant for governance of long-term problems.

Similarly, Boston (2017, 101-103) identifies four factors that make a policy problem more likely to exhibit presentist bias. First, whether the policy problem has significant intertemporal dimension by placing costs on future generations. Second, is it a wicked problem, thereby being intractable, causally complex, disputed, and uncertain? Third, whether a front-loaded policy investment is required to mitigate the problem. Fourth, if the policy problem involves intertemporal trade-offs between incomparable values like financial versus immaterial goods. These factors affect both the demand and supply side of policy responses to mitigate the problem. Furthermore, it is important to assess the nature of the required policy investment and to what degree this imposes upfront costs in exchange for future benefits, such as regulatory framework to improve safety of a technology (Jacobs 2016). Policy investments are easier to justify the more credible and concrete the harms and those (people, sectors, or industries) negatively affected by the problem are, for example, through clear warning signals (Boston 2017, 111). Prospects for long-term policy investment are also strengthened if its expected benefits are relatively certain, tangible, and direct, whereas complex policies with high up-front costs and delayed, immaterial benefits face greater resistance. Lastly, policy flexibility is also beneficial, namely if the policymakers can alter the visibility, magnitude, distribution, or timing of the costs to different interest groups (Boston 2017, 120–122).

As a synthesis of the literature outlined above, eight framework attributes that enable holistic comparison of long-term policy problems such as risks of climate change and AI were identified. The selection and compilation of attributes is based on their relevance to long-term problems with temporal complexity to make for a concise and parsimonious framework. For example, solubility raised by Peters (2005) is not included, since the continuous nature of long-term problems makes them less susceptible to one-time interventions. Boston (2017, 104) initially identified 13 different attributes of policy problems: problem definition, spatial considerations, temporal considerations, distributional considerations, the kind of impacts, predictability, visibility and tangibility, linearity, path dependence, urgency, complexity, and causal certainty, technical capacity to solve, and fiscal (and other) costs. In contrast, this framework combines aspects of multiple attributes, such as the problem definition, predictability, and causal certainty, under a single heading of scientific certainty in the interest of parsimony. One notable omission from the attributes is the severity or urgency of the problem, but this analysis treats them as a combination of the other attributes. The first five attributes of the framework seek to capture the intrinsic characteristics of policy problems whereas the last three concern the required policy investments. The attributes are as follows.

- Scientific certainty: how strong is the scientific consensus on the causes and effects of the problem?
- Spatiality: what is the geographical scope of the problem and its impacts?
- Temporality: over what timeframe is the problem expected to unfold?
- Linearity: does the problem progress steadily in cumulative manner?
- Path dependence: is there a risk that the problem becomes irreversibly locked-in?
- Accountability: who is most responsible for the problem?
- Capacity to address: does the technical and political capability to amend the problem exist?
- Costs involved: how expensive is the problem to address?

4 | Analysis

This section demonstrates the use of the comparative framework by contrasting the risks of climate change and AI based on the eight attributes above, covering both the nature of the policy problems themselves and their required policy investments.

4.1 | Scientific Certainty

There is over 99% scientific consensus that human activities are causing global warming (Lynas, Houlton, and Perry 2021) and its severe impacts on human and nonhuman life, such as extreme weather events, crop shortages, and diseases. There are reliable projections about future climate conditions under different emission pathways, and the historical track record of these climate models has been accurate (Hausfather et al. 2020). However, some uncertainties remain about the causality and severity of climate impacts, especially in the more extreme cases such as positive feedback loops and tipping points (IPCC 2023). Conversely, the risks of AI systems constitute a contested and ambiguous problem where the scientific consensus is still emerging. For instance, between 38% and 51% of AI researchers estimated at least a 10% chance of advanced AI causing catastrophic outcomes in a recent survey (Grace et al. 2024). AI risks, especially in the long term are relatively uncertain, unpredictable, and related to complex socio-technical systems (see Bengio 2024). As such, AI governance is more focused on anticipation and prevention of potential, yet uncertain future risks. Similar tipping points to climate change, such as recursively self-improving AI systems have been postulated but remain controversial. However, there is a relative consensus on the ethical principles that should guide AI development (see OECD 2019).

4.2 | Spatiality

Climate change is a fundamentally global problem affecting the whole planet. However, some regions and countries will warm more and bear extreme climate risks unevenly. More importantly, populations and socioeconomic groups who have contributed least to climate change tend to be most vulnerable and exposed to it, such as poor communities in small island states (Caney 2021). The risks of climate change are therefore unevenly distributed both intra- and internationally, requiring global coordination to distribute the burden, compensate damages, and fund clean technologies. Risks of AI are similarly a global problem as AI systems and algorithms are not bound by national boundaries. However, the risks of AI systems can be more localized depending on the specific system and where it is deployed (e.g., a local predictive policing system). Currently, AI governance can still be described as a mesoproblem, in which measures taken in a single country have their largest effects within that country, thereby incentivizing national policy action (see Hansson 2023). While one could argue climate risks also materialize differently in local contexts, climate change remains a global phenomenon, whereas most AI risks can be avoided by not deploying the system. Similar to climate change, the risks of AI systems such as discrimination, surveillance, and job displacement are likely to be unequally distributed and exacerbate existing social inequalities (Eubanks 2018). More advanced general-purpose AI systems are likely to have pronounced global effects.

4.3 | Temporality

Climate change is a long-term problem that will affect future generations for centuries or millennia to come. While the global mean temperatures are already up by 1.2°C since preindustrial times with devastating effects, the impact of climate change will worsen progressively (30–100 years[6]) as temperatures rise, especially if GHG emissions are not reduced (IPCC 2023). These disproportionate effects on future generations make climate risks fundamentally a problem of intergenerational justice. In contrast, risks of AI generally involve shorter timeframes (5–25 years) due to rapid technological development. Despite lacking a similar intergenerational conflict, AI systems can reshape societal norms and institutions in long-lasting ways, even if the risks are more episodic and nonlinear. AI already affects various sectors through automation, with uncertain and potentially transformative impacts as AI systems become more autonomous (Gruetzemacher and Whittlestone 2022). Risks from advanced AI systems are arguably proportional to how soon they are developed and how much time there is to ensure social adaptation and safety. For example, similarly advanced AI system might possess a significant chance of severe harm if deployed in 2030, but only minimal in 2060 (see Kilian, Ventura, and Bailey 2023). Nonetheless, shorter timeframes of AI risks make it easier to justify regulatory response due to greater clarity on the concrete harms, those affected, and the effects of regulation.

4.4 | Linearity

The risks of climate change such as ocean acidification are generally invisible, slowly accumulating "creeping problems" (Glantz 1999) that tend to generate few focusing events, evading human attention. This is despite the more concrete, observable, and measurable physical indicators of climate change such as temperatures, sea levels, and carbon dioxide concentrations. Climate change can be characterized as a linear process of rising greenhouse gas emissions and adverse environmental effects (excluding possible tipping points). In comparison, risks of AI systems tend to be intermittent and nonlinear, depending on how they are deployed in society. While higher training compute and larger neural networks—both doubling every 6-9 months (Epoch 2023)—are associated with increased AI model capabilities and risks (Bengio, 2024), this is not directly analogous to the harms of GHG emissions. Due to the context specific and emergent nature of AI risks, there may be more visible and concrete warning signals of harm than with climate change, such as the release of ChatGPT, which spurred governments to respond. Somewhat paradoxically, harms of climate change appear linear but less visible whereas AI risks are nonlinear but rather tangible.

4.5 | Path Dependence

The risks of climate change are likely to become harder and more costly to address the longer GHG emissions continue accumulating in the atmosphere. In fact, certain catastrophic harms of climate change are potentially irreversible if tipping points such abrupt boreal permafrost thaw are crossed (Lenton et al. 2019; 2023), which might already unfold with 1.5°C warming. Climate science suggests there is only a limited window of opportunity to avoid crossing these physical tipping points (Armstrong McKay et al. 2022). Moreover, there are carbon lock-in effects and inertia in socioeconomic systems that make it difficult to reduce emissions once specific infrastructures, technologies, or policies are in place (Seto et al. 2016). Current AI systems can already contribute to harmful technological path dependency by extrapolating existing social patterns into future predictions and by becoming infrastructurally entrenched. Yet, irreversible lock-in risks become pronounced under radically transformative AI scenarios, such as misaligned advanced AI systems. Some argue there are critical thresholds related to AI capabilities that could lead to transformative impacts and undermine safety if crossed (Kilian, Ventura, and Bailey 2023; Vold and Harris 2021). Scientific uncertainty regarding these extreme scenarios is considerable on both domains, but lock-in risks are likely to be more abrupt with AI, as most climate tipping points take decades to centuries to materialize (Wang et al. 2023). In both cases, these path dependencies can also stem from the policy and governance responses adopted, potentially affecting future generations irreversibly.

4.6 | Accountability

Fossil fuels are deeply ingrained in the societal infrastructure and lifestyles of large numbers of people, resulting in a collective action problem. Tackling climate change requires collective action from society, even if responsibilities may vary depending on who has most contributed to climate change, their ability to pay, and how they have benefitted historically (Caney 2021). Regardless of the specific principle chosen, large polluting nations, companies and citizens of the Global North are the ones most accountable for addressing the risks of climate change. AI risks are rather specific regulatory issue with the main responsibility falling on those developing and deploying AI systems, such as industrial AI labs, including Microsoft's OpenAI or Google DeepMind. Regulation of AI development therefore requires less up-front costs to enact in comparison to climate change, which is tied to controversial and personal issues such as energy consumption, diet, and transportation. This is subject to change as AI systems become more integrated into societal infrastructure and open-source models proliferate (Bucknall and Dori-Hacohen 2022). Although AI companies face near-term regulatory costs, the benefits of regulation such as avoidance of "arms race" between companies might also incentivize them to advocate for regulation in some instances.

4.7 | Capacity to Address

The technical and political capacity to mitigate climate change largely exists, even if not fully deployed or scaled. Deployment of renewable energy like solar and wind power has risen rapidly as prices have fallen (IEA 2023). Further technological innovations such as carbon capture or solar radiation management are required under most projections that limit the warming to 1.5°C–2°C (see IPCC 2023). All of this requires international coordination on climate policy, which largely exists through the United Nations system (e.g., UNFCCC, COP, Paris Agreement, IPCC). Like climate change, the risks of AI systems are not easily solved by one-time interventions and require ongoing governance efforts. Yet, the technical and political capacity for tackling AI risks is only emerging and severely lacking behind the rapid progress in AI capabilities, requiring foresight and anticipation. While techniques for improving the trustworthiness and safety of AI systems (e.g., adversarial debiasing, mechanistic interpretability, model evaluations) are emerging, there are barriers to their deployment and standardization, such as arms race dynamics and geopolitical competition (Bengio, 2024). In terms of political capacity, mechanisms of international AI governance are still nascent in comparison to climate governance, but are materializing briskly (e.g., EU's AI Act). While the rapid development of AI models poses a constant "catch-up" challenge to regulators, it also enables more policy flexibility and opportunities for innovative regulatory design, like windfall clauses (O'Keefe et al. 2020), to alter the distribution of near-term costs of AI regulation.

4.8 | Costs Involved

Governance of climate change and AI risks both require tangible up-front costs and deliver mostly delayed and indirect benefits, especially in the case of climate mitigation. The short-term fiscal costs of climate mitigation and adaptation are significant even if a rapid transition to green energy by 2050 is likely to be ultimately beneficial economically (Way et al. 2022).[7] The sustainability transition necessitates changes in the employment of citizens (e.g., coal mine and manufacturing workers), potentially lessening support of such policies. Overall, climate mitigation is other-regarding action that does not particularly benefit the large emitters. However, emissions trading and the current investment boom into clean energy and green jobs (IEA 2023), partly alleviate these worries. On the other hand, the development of AI represents direct economic, business, and scientific benefits by potentially adding trillions to GDP.[8] While regulation may slow down AI adoption, it could ultimately boost economic growth and innovations by fostering more aligned and safer AI systems. Overall, the costs of AI regulation appear feasible to alter and less prohibitive than climate action.

Table 2 summarizes how the risks of climate change and AI compare against each other as long-term policy problems.

Generally, the more complex, uncertain, global, future-oriented, costly, and invisible the policy problem is, the harder it is to govern (Boston 2017, 129). Based on the above analysis, the risks of AI fare better on ‘long-term governability’ in terms of four attributes: temporality, accountability, linearity, and costs involved. As such, anticipatory AI governance benefits from shorter timeframes, less diffused accountability, more visible harms, and positive economic benefits associated with AI. Conversely, risks of climate change are more governable only in terms of two attributes: greater scientific certainty and capacity to address the problem. In terms of spatiality and path dependence, it is harder to decipher if one of the problems is more conducive to long-term governance, even if there is more clarity over climate tipping points than the respective lock-in risks of AI.

Overall, it could be argued that climate governance is therefore more prone to harmful short-termism than AI governance. Yet, both problems require policy investments that generate delayed and intangible benefits such as improved health and safety. These benefits are rather indirect, uncertain, and invisible compared to the near-term costs, hindering long-term governance. Despite these challenges, regulatory responses to risks of climate change and AI both seem to enjoy public support (see AIPI, Artificial Intelligence Policy Institute 2023). This comparison serves as a proof of concept. Further methodological research could develop more objective, attribute-based criteria and evaluate the relative importance of different attributes, but that is beyond this article's scope. As a key difference in the comparison, the inverse temporality of climate change and AI risks is illustrated in Graph 1.

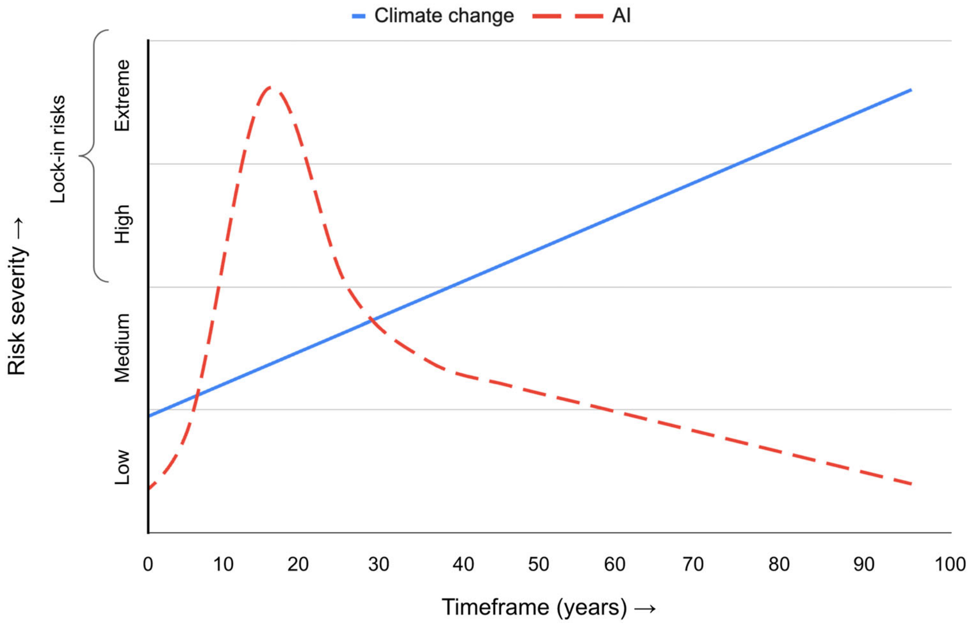
5 | Discussion

The main finding from the comparative analysis is that the risks of climate change and AI differ importantly as long-term policy problems, especially as AI does not involve a similar intergenerational conflict as climate change. However, the comparison did highlight path dependence, specifically irreversible lock-in effects emerging as key attributes with both problems, even if on different timeframes. Fossil fuels and AI are both innovations that deliver immediate benefits but are almost entirely dependent on future technological progress to solve their respective risks and externalities. Using nonrenewable resources for technological development without sufficiently considering their risks effectively constrains future generations' decisions, creating strong path dependencies that force people

TABLE 2 | Comparison of the risks of climate change and AI as long-term problems.

	Climate change	Risks of AI	Similarity
Scientific certainty	High; certainty and predictability based on physical evidence (+)	Low; uncertain, unpredictable, and disputed nature of the problem	Different
Spatiality	Global problem; although unequally distributed impacts	Global-national problem; although unequally distributed impacts	Similar
Temporality	Long timeframe; 30–100 years for severe risks	Short timeframe; 5–25 years for severe risks (+)	Different
Linearity	Cumulative problem; linear but less visible risks	Intermittent problem; visible but nonlinear risks (+)	Different
Path dependence	Physical lock-in effects as the problem exacerbates	Technological lock-in effects as the problem exacerbates	Similar
Accountability	Collective issue; focused on social infrastructure and lifestyles	Sectoral issue; focused on specific AI developers and deployers (+)	Different
Capacity to address	Medium–high; technical and political capacity largely exist (+)	Low–medium; technical and political capacity do not exist, although are emerging	Different
Costs involved	Medium–high economic costs to address; delayed benefits	Low–medium economic costs to address; potential benefits of AI (+)	Different

Note: The plus sign (+) signifies conduciveness to long-term governance.



GRAPH 1 | Illustration of the temporal and linear differences between risks of climate change and AI. The severe risks of AI might actualize nonlinearly within shorter timeframes (5–25 years) than climate change, which worsens cumulatively (30–100 years) with lock-in risks.[10].

to focus their resources on addressing these harms (Hendlin 2014, 18). While there is no scientific consensus on the likelihood of extreme lock-in risks or tipping points, both AI and climate change are susceptible to harmful forms of path dependency, such as carbon lock-in. This is increasingly true as AI becomes part of entrenched energy-intensive infrastructure, with largest models costing over $1B to train with significant environmental footprint, potentially locking in harmful trajectories (Robbins and van Wynsberghe 2022).

The attributes of policy problems are not independent of each other but often interrelated, with significant long-term governance implications. For instance, the capacity to address a problem is closely related to the costs involved, as demonstrated by their similar assessments in this comparison. Moreover, lack of capacity coupled with high costs is likely to intensify path dependence in turn. More importantly, while strong path dependence and lack of scientific certainty are problematic on their own, together they make governance of long-term problems particularly toxic or "wicked," as in the case of AI. Uncertainty complicates decision-making, as it obscures which path to take, while lock-in effects make certain decisions irreversible. The inability to reverse decisions renders governance approaches which would otherwise be favored under uncertainty, such as experimentation[9], infeasible. This combination necessitates a precautionary approach to manage potential risks, even before clear causal evidence has been established scientifically. Precautionary principle is generally used to justify action to prevent scientifically uncertain, but potentially serious and irreversible harms (Bourguignon 2015), as is the case with lock-in risks of climate change and AI.

There are multiple interpretations of the precautionary principle(s) and its stringency, but the comparative analysis of climate change and AI suggests that the uncertainty over their potentially irreversible lock-in risks obligates action to mitigate them (see Rechnitzer 2020; Aven 2023). This implies that AI and climate policies should be subject to stringent precautionary measures and evaluation mechanisms before implementation, given the scientific uncertainty regarding the probability and magnitude of harmful lock-in risks. Contrary to inaction, this means proactively identifying and addressing potential lock-ins and tipping points through policy interventions. Although precautionary policies might hinder innovation, they are still justified from long-term governance perspective due to the paramount importance of preventing irreversible lock-in effects that harm autonomy of future generations. This necessitates some overcaution and false positives, where precautions prove to be unnecessary later (Rechnitzer 2020). However, this should not amount to a complete reversal of the burden of proof as demonstrating safety in relation to the long-term risks posed by climate change and AI can be rather infeasible.

Precautionary policy and technology measures can also contribute to the lock-in risks. For instance, Cairns (2014) warns that geoengineering technologies, especially stratospheric aerosol injections, risk becoming irreversibly locked-in themselves due to the "termination effect" and rapid warming that would occur if they were discontinued prematurely. McKinnon (2019, 449) also argues that failing to anticipate and mitigate harmful lock-in risks of research into emerging technologies would be to wrong future generations by recklessly endangering them. This requires governance provisions that enable significant slowing or shutdown of technological research programs if they pose serious lock-in risks (e.g., by risk assessment, ethical requirements, defunding, moratoriums), such as research into more capable AI systems. Yet, the application of the precautionary principle is contextual and does not by itself presuppose any specific measures, such as R&D bans (Rechnitzer 2020). The exact precautionary measures against lock-in risks of climate change and AI, their restrictiveness and triggers are beyond the scope of this article.

However, based on the comparison, the standard precautionary measures such as cost–benefit analysis appear insufficient in the context of AI and climate risks given their long-term nature, and tendency for policymakers to discount future benefits. Moreover, AI risks might justify more stringent precautionary measures given the pervasive scientific uncertainty and unknown unknowns surrounding AI capabilities (see Anderljung et al. 2023; Bengio 2024). The uncertainty concerning AI arguably derives from ambiguity and ignorance of potential risks, whereas with climate change the risks and probabilities are known better, and uncertainty is related to their complexity (see Bourguignon 2015). Hence, straightforward prevention principle is more applicable to climate risks that do not amount to tipping points. In addition, lock-in risks of AI may require more active steering of technological development through regulation in contrast to the more static and well-defined physical mitigation targets in climate policy. Nevertheless, building anticipatory capacity (Guston 2014) for emerging risks in AI and climate governance appears essential because of the disconnect between technological progress and policymaking – regulation will be invariably too late if only enacted after concrete harms and lock-in effects materialize.

I hope to have demonstrated how the policy problem-focused approach can inform long-term governance of different issues. The comparative approach outlined in the article can potentially aid the problem definition stage of policy analysis by integrating it with foresight and horizon scanning (van Dorsser et al. 2018; Cuhls 2020). This approach aids in categorizing policy problems based on their level of uncertainty and aligns with Muiderman et al.'s (2020) first two approaches to anticipatory governance: assessing probable futures to inform strategic risk reduction policy and exploring plausible futures to build preparedness for uncertain outcomes. Based on the categorization by Walker (2011) and van Dorsser et al. (2018, 2020), climate change and AI risks generally fall under level 2 and 3 uncertainty—probable and plausible futures—requiring foresight and risk-based policies to mitigate uncertainty. Yet, the greater uncertainty and “unknown unknowns” of AI development and lock-in risks also warrant exploration of possible futures to improve precautionary measures. However, this also points to a limitation in the current categorizations, which tend to equate uncertainty with the time horizon (see van Dorsser et al. 2018, 83). In contrast, this comparison found that the severe risks of AI systems are more uncertain despite their shorter timeframes, whereas the risks of climate change are more predictable, but also occur over longer time horizon. Exploring these dynamics further holds potential as a future research direction. The comparative framework presented could also be applied to a broader selection of issues and signals to assess its robustness and to identify potential refinements as part of horizon scanning.

6 | Conclusion

In this article, I have compared the risk of climate change and artificial intelligence as long-term policy problems to examine their distinctive attributes with relevance to their governance. The comparative analysis contrasted the two policy problems in terms of eight dimensions: scientific certainty, temporality, linearity, spatiality, path dependence, accountability, capacity to address, and costs involved. The comparison shows that the risks of AI are likely more conducive to long-term governance due to the less pronounced intergenerational conflict and more favorable environment for policy investments. However, the risks of harmful lock-in effects that threaten to curtail options of future generations were identified as particularly relevant in both contexts, especially in relation to more uncertain long-term risks such as climate tipping points and advanced AI systems. Mitigating these lock-in effects requires precautionary governance measures. This article has illustrated the value of comparative methods at the intersection of foresight and policy analysis, especially in relation to problem definition and horizon scanning of emerging risks. Going forward, this hopefully informs more concrete research on long-term governance across different policy problems and technologies.

Acknowledgments

This paper was supported by the KT4D and TANDEM projects, which have received funding from the European Union's Horizon Europe program under grant agreements No 101094302 and No 101069653, respectively. I want to thank Maija Setälä, Simon Caney, Oskari Sivula, Henri Huttunen, and anonymous reviewers for their constructive comments on the article.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Endnotes

[1] To limit its scope, this article specifically focuses on the governance of risks of AI and climate change, rather than on them as general phenomena. Hence, the potential benefits of these issues are not covered even if important.

[2] By long-term, I generally refer to intergenerational problems that can affect nonexistent future generations and demonstrate intertemporal dilemmas (Jacobs 2016).

[3] Although some see non-domination as a specific type of egalitarian views, see González-Ricoy (2019, 4).

[4] The possibility of a runaway warming resulting in multiple degrees of additional warming over the next centuries still remains unlikely (Wang et al. 2023). The nine global tipping points could individually impact global warming by -0.5°C to 0.6°C, averaging at 0.1°C (Armstrong McKay et al. 2022).

[5] Takeoff scenarios refer to how slow (years or decades) or fast (weeks or months) advanced AI system might improve from human-level to superintelligent capabilities, measured in, for example, economic productiveness.

[6] These timeframes are illustrative and refer to when the most severe risks might actualize, not their duration as both problems can affect the long-run future. According to Wang et al. (2023), most climate tipping points do not possess the potential for abrupt change within the next 50 years.

[7] According to the IPCC (2018, 22) limiting global warming to 1.5°C requires an annual energy investment of around $2.4 trillion, between 2016 and 2035, which is 2.5% of the world GDP.

[8] PwC Global (2017) has predicted AI could boost global GDP by 14% ($15.7 trillion) by 2030, while Goldman Sachs (2023) estimates a 7% ($7 trillion) increase over the next decade due to generative AI.

[9] Although AI arguably allows for regulatory sandboxes and experimenting with systems in controlled environments before broader deployment.

[10] Note that the graph focuses on AI as defined by OECD (2019). It does not cover all emerging technologies, which might generate new risks.

References

AIPI, Artificial Intelligence Policy Institute. (2023, September 19). Overwhelming Majority of Voters Believe Tech Companies Should Be Liable for Harm Caused by AI Models. https://theaipi.org/poll-shows-voters-oppose-open-sourcing-ai-models-support-regulatory-representation-on-boards-and-say-ai-risks-outweigh-benefits-2/.

Anderljung, M., J. Barnhart, J. Leung, et al. (2023). Frontier AI Regulation: Managing Emerging Risks to Public Safety.

Armstrong McKay, D. I., A. Staal, J. F. Abrams, et al. 2022. "Exceeding 1.5 C Global Warming Could Trigger Multiple Climate Tipping Points." *Science* 377, no. 6611: eabn7950.

Aschenbrenner, L. 2020. *Existential Risk and Growth*. Global Priorities Institute.

Aven, T. 2023. "A Risk and Safety Science Perspective on the Precautionary Principle." *Safety Science* 165: 106211.

Beard, S. J., L. Holt, A. Tzachor, et al. 2021. "Assessing Climate Change's Contribution to Global Catastrophic Risk." *Futures* 127: 102673.

Beckman, L. 2016. "Power and Future People's Freedom: Intergenerational Domination, Climate Change, and Constitutionalism." *Journal of Political Power* 9, no. 2: 289–307.

Bengio, Y. 2024. *International Scientific Report on the Safety of Advanced AI: Interim Report*. UK: Department for Science, Innovation and Technology.

Boston, J. 2017. *Governing for the Future: Designing Democratic Institutions for a Better Tomorrow*. Bingley: Emerald Group Publishing Limited.

Bourguignon, D. 2015. "The Precautionary Principle: Definitions, Applications and Governance." *European Parliamentary Research Service* PE 573: 1.

Bucknall, B. S., and S. Dori-Hacohen. 2022. Current and Near-Term AI as a Potential Existential Risk Factor. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*, 119–129. New York: Association for Computing Machinery.

Cairns, R. C. 2014. "Climate Geoengineering: Issues of Path-Dependence and Socio-Technical Lock-In." *WIREs Climate Change* 5, no. 5: 649–661.

CAIS Center for AI Safety. 2023. Statement on AI Risk. https://www.safe.ai/statement-on-ai-risk.

Caney, S. 2021. Climate Justice. The Stanford Encyclopedia of Philosophy (Winter 2021 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/win2021/entries/justice-climate/.

Cuhls, K. E. 2020. "Horizon Scanning in Foresight–Why Horizon Scanning Is Only a Part of the Game." *Futures & Foresight Science* 2, no. 1: e23.

van Dorsser, C., P. Taneja, W. Walker, and V. Marchau. 2020. "An Integrated Framework for Anticipating the Future and Dealing With Uncertainty in Policymaking." *Futures* 124: 102594.

van Dorsser, C., W. E. Walker, P. Taneja, and V. A. W. J. Marchau. 2018. "Improving the Link Between the Futures Field and Policymaking." *Futures* 104: 75–84.

Epoch 2023. *Key Trends and Figures in Machine Learning*. https://epochai.org/trends.

Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Gardiner, S. 2011. *A Perfect Moral Storm: The Ethical Tragedy of Climate Change*. New York: Oxford University Press.

Glantz, M., ed. 1999. *Creeping environmental problems and sustainable development in the Aral Sea basin*. Cambridge: Cambridge University Press.

Goldman Sachs. 2023, April. *Generative AI could raise global GDP by 7%*. https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html.

Goldstein, J. E., B. Neimark, B. Garvey, and J. Phelps. 2023. "Unlocking "Lock-In" and Path Dependency: A Review Across Disciplines and Socio-Environmental Contexts." *World Development* 161: 106116.

González-Ricoy, I. 2019. "Intergenerational Justice and Institutions for the Long Term." In *The Oxford Handbook of Time and Politics*, edited by K. Goetz. New York: Oxford University Press.

Grace, K., H. Stewart, J. F. Sandkühler, S. Thomas, B. Weinstein-Raun, and J. Brauner. 2024. Thousands of AI authors on the future of AI. arXiv preprint arXiv:2401.02843.

Greaves, H., W. MacAskill, R. O'Keeffe-O'Donovan, et al. 2020. *A Research Agenda for the Global Priorities Institute*. University of Oxford.

Gruetzemacher, R., and J. Whittlestone. 2022. "The Transformative Potential of Artificial Intelligence." *Futures* 135: 102884.

Guston, D. H. 2014. "Understanding 'Anticipatory Governance." *Social Studies of science* 44, no. 2: 218–242.

Hansson, S. O. (2023). The Covid-19 Pandemic and Global Climate Change: Comparing Two Long-term Dangers. In Democracy and the Future: Future-Regarding Governance in Democratic Systems, edited by M. MacKenzie, M. Setälä, and S. Kyllönen. Edinburgh: Edinburgh University Press.

Hausfather, Z., H. F. Drake, T. Abbott, and G. A. Schmidt. 2020. "Evaluating the Performance of Past Climate Model Projections." *Geophysical Research Letters* 47, no. 1: e2019GL085378.

Hendlin, Y. H. 2014. "The Threshold Problem in Intergenerational Justice." *Ethics & the Environment* 19, no. 2: 1–38.

IEA. 2023. *World Energy Investment 2023*. Paris: IEA.

IPCC. 2018. "Summary for Policymakers." In *IPCC Special Report on Global Warming of 1.5°C*, 3–24. Cambridge, UK and New York, NY, USA: Cambridge University Press.

IPCC. 2023. "Summary for Policymakers." In *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, edited by H. Lee, and J. Romero, 1–34. Geneva, Switzerland: IPCC.

Jacobs, A. M. 2016. "Policy Making for the Long Term in Advanced Democracies." *Annual Review of Political Science* 19: 433–454.

Jones, C. I. 2016. "Life and Growth." *Journal of Political Economy* 124, no. 2: 539–578.

Kilian, K. A., C. J. Ventura, and M. M. Bailey. 2023. "Examining the Differential Risk From High-Level Artificial Intelligence and the Question of Control." *Futures* 151: 103182.

van Lente, H., T. Swierstra, and P. B. Joly. 2017. "Responsible Innovation as a Critique of Technology Assessment." *Journal of Responsible Innovation* 4, no. 2: 254–261.

Lenton, T. M., D. I. Armstrong McKay, S. Loriani, et al. 2023. *The Global Tipping Points Report 2023*. Exeter, UK: University of Exeter.

Lenton, T. M., J. Rockström, O. Gaffney, et al. 2019. "Climate Tipping Points—Too Risky to Bet Against." *Nature* 575, no. 7784: 592–595.

Lynas, M., B. Z. Houlton, and S. Perry. 2021. "Greater Than 99% Consensus on Human Caused Climate Change in the Peer-Reviewed Scientific Literature." *Environmental Research Letters* 16, no. 11: 114005.

MacKenzie, M., Setälä, M., and Kyllönen, S., ed. 2023. *Democracy and the Future: Future-Regarding Governance in Democratic Systems*. Edinburgh: Edinburgh University Press.

McKinnon, C. 2019. "Sleepwalking into Lock-In? Avoiding Wrongs to Future People in the Governance of Solar Radiation Management Research." *Environmental Politics* 28, no. 3: 441–459.

Metaculus. 2024, April. When Will the First General AI System Be Devised, Tested, and Publicly Announced? https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/.

Meyer, L. 2021. "Intergenerational Justice." In *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), edited by E. N. Zalta. Stanford: Stantford University.

Muiderman, K., A. Gupta, J. Vervoort, and F. Biermann. 2020. "Four Approaches to Anticipatory Climate Governance: Different Conceptions of the Future and Implications for the Present." *WIREs Climate Change* 11, no. 6: e673.

Nolt, J. 2011. "Greenhouse Gas Emission and the Domination of Posterity." In *The Ethics of Global Climate Change*, edited by D. Arnold, 60–76. Cambridge: Cambridge University Press.

OECD. 2019. *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. Paris: OECD Legal Instruments.

OECD. 2022. *Climate Tipping Points: Insights for Effective Policy Action*. Paris: OECD Publishing. https://doi.org/10.1787/abc5a69e-en.

O'Keefe, C., P. Cihon, B. Garfinkel, C. Flynn, J. Leung, and A. Dafoe. 2020. The Windfall Clause: Distributing the Benefits of AI for the Common Good. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*, 327–331. New York: Association for Computing Machinery.

Ord, T. 2020. *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury Publishing.

Patton, C., D. Sawicki, and J. Clark. 2015. *Basic Methods of Policy Analysis and Planning*. New York: Routledge.

Peters, B. G. 2005. "The Problem of Policy Problems." *Journal of Comparative Policy Analysis* 7, no. No. 4: 349–370.

Pierson, P. 2000. "Increasing Returns, Path Dependence, and the Study of Politics." *American Political Science Review* 94, no. 2: 251–267.

Pietrobelli, C., and F. Puppato. 2016. "Technology Foresight and Industrial Strategy." *Technological Forecasting and Social Change* 110: 117–125.

PwC Global. 2017. *Sizing the Price—PwC's Global Artificial Intelligence Study—Exploiting the AI Revolution*. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.

Rechnitzer, T. 2020. "Precautionary Principles." *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/pre-caut/.

Robbins, S., and A. van Wynsberghe. 2022. "Our New Artificial Intelligence Infrastructure: Becoming Locked Into an Unsustainable Future." *Sustainability* 14, no. 8: 4829.

Sandbrink, J., H. Hobbs, J. Swett, A. Dafoe, and A. Sandberg. 2022. "Differential Technology Development: A Responsible Innovation Principle for Navigating Technology Risks." *Available at SSRN*. https://doi.org/10.2139/ssrn.4213670.

Seto, K. C., S. J. Davis, R. B. Mitchell, E. C. Stokes, G. Unruh, and D. Ürge-Vorsatz. 2016. "Carbon Lock-In: Types, Causes, and Policy Implications." *Annual Review of Environment and Resources* 41: 425–452.

Shackley, S., and M. Thompson. 2012. "Lost in the Mix: Will the Technologies of Carbon Dioxide Capture and Storage Provide Us With a Breathing Space as We Strive to Make the Transition from Fossil Fuels to Renewables?" *Climatic Change* 110, no. 1: 101–121.

Shapiro, C., and H. R. Varian. 1999. *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business Press.

Vold, K., and D. R. Harris. 2021. "How Does Artificial Intelligence Pose an Existential Risk?" In *Oxford Handbook of Digital Ethics*, edited by Carissa Véliz. Oxford: Oxford University Press.

Walker, W. E. 2011. "Policy Analysis, 1962-2012: From Predict and Act to Monitor and Adapt." *Farewell Lecture at TU Delft*. https://repository.tudelft.nl/record/uuid:2d09b2a7-de07-402d-b73d-471a438e7f79.

Wang, S., A. Foster, E. A. Lenz, et al. 2023. "Mechanisms and Impacts of Earth System Tipping Elements." *Reviews of Geophysics* 61: e2021RG000757.

Way, R., M. C. Ives, P. Mealy, and J. D. Farmer. 2022. "Empirically Grounded Technology Forecasts and the Energy Transition." *Joule* 6, no. 9: 2057–2082.

van Westen, R. M., M. Kliphuis, and H. A. Dijkstra. 2024. "Physics-Based Early Warning Signal Shows That AMOC Is on Tipping Course." *Science Advances* 10, no. 6: eadk1189.