

Forthcoming Chapter in:
Reflections of AI for Humanity, coord. by B. Braunschweig and M. Ghallab, Springer (2021)

A Framework for Global Cooperation on Artificial Intelligence and its Governance

Pekka Ala-Pietilä¹ and Nathalie A. Smuha²

¹ Chair of the EC High-Level Expert Group on AI & AI Finland

² FWO Fellow at the KU Leuven Faculty of Law & Leuven.AI Institute, Leuven, Belgium,
nathalie.smuha@kuleuven.be

Abstract. In this chapter we put forward a framework for global cooperation on Artificial Intelligence (AI) and its governance, with the aim to ensure that humanity can enjoy its benefits while preventing and minimizing its risks. It is structured along three sections, focusing respectively on the *why*, *what* and *how* of global cooperation. First, we set out *why* AI requires governance and why its governance necessitates global cooperation. Particular focus is given to the need for a level playing field that secures citizen protection across the globe, enables socially beneficial innovation, and stimulates healthy competition to disseminate AI's benefits. Second, we list *what* the substantive areas are on which global cooperation on AI should be prioritized, and argue for a holistic approach along two dimensions. The first dimension is horizontal, and aims to identify minimum requirements that cover the entire socio-technical environment of AI in a transversal manner. In addition to AI-systems, we urge to consider the socio-technical environments of data and digital infrastructure, which are inextricably interwoven therewith. The second dimension is vertical, and aims at cooperation around domain-specific areas that require a more tailored approach to maximize AI's benefits for humanity, to prevent and minimize its risks and to address ad-hoc issues. Third, we assess *how* global cooperation should be organized. We stress the need to balance speed, holism and contextualism, and provide a number of guiding principles that can inform the process of global cooperation initiatives on AI and its governance.

Keywords: Artificial Intelligence, Global Cooperation, Competition, Regulation, Governance, Data, Digital Infrastructure.

Table of Contents

1	Introduction	3
2	The need for global cooperation on AI.....	4
2.1	Why AI necessitates governance.....	5
2.2	Why AI's governance requires a global approach	6
2.3	Challenges to overcome	8
3	Areas for global cooperation on AI and its governance	10
3.1	The horizontal dimension of AI governance	11
(a)	<i>Building a trustworthy environment for AI-systems</i>	<i>12</i>
(b)	<i>Building a trustworthy environment for data</i>	<i>14</i>
(c)	<i>Building a trustworthy environment for digital infrastructure</i>	<i>17</i>
3.2	The vertical dimension to AI governance	19
(a)	<i>Cooperation on maximizing AI's benefits for humanity.....</i>	<i>20</i>
(b)	<i>Cooperation on preventing and minimizing AI's risks for humanity</i>	<i>21</i>
(c)	<i>Cooperation on ad hoc matters</i>	<i>22</i>
4	Organizing global cooperation on AI and its governance	22
4.1	Balancing the need for swift action, a holistic approach and attention to context-specificity	23
4.2	Clarifying the rules of engagement	23
4.3	Building on existing cooperation structures	24
4.4	Developing a network of networks	25
4.5	Maintaining openness to differentiated cooperation	26
4.6	Securing an inclusive and transparent way of working, mindful of power imbalances	27
4.7	Establishing a feedback loop and preparing for the future.....	29
5	Conclusions	29
	Acknowledgement	31
	References.....	31

1 Introduction

Artificial Intelligence (AI), an umbrella term for a range of ‘intelligent’ technological applications, is not a new phenomenon [45]. Yet the combination of an increased availability of high computing power and large amounts of data, as well as advances in AI’s research domain, significantly boosted the technology’s possibilities [15, 34]. As a result, it is being used in an ever-wider range of applications and in ever-more domains – both in the public and private sphere. Given that these opportunities are accompanied by significant challenges and risks, attention to AI’s governance peaked throughout the world. Today, it is widely agreed that – if we are to ensure that humanity can enjoy AI’s benefits while preventing and minimizing its risks – global cooperation is a necessity [18]. Yet the motives behind the need for cooperation, the concrete areas of cooperation and the manner in which cooperation should take place often still remain debated. With the aim of diving into this debate, this chapter puts forward a framework for global cooperation on AI and its governance.

A large number of actors are already engaged in global cooperation in one way or another. Accordingly, global cooperation on AI does not take place in a vacuum. In addition, new cooperation initiatives around AI are regularly emerging – the Global Partnership on AI being an example [48]. Recognizing this multitude of efforts, we take a step back to consider the bigger picture that these initiatives are part of by shedding light on the overarching orientation they may consider, each from their own angle. Our framework is guided by three fundamental questions: *Why* is there a need for global cooperation on AI and its governance? (Section 2); *What* are the substantive areas on which global cooperation should take place? (Section 3) and; *How* should the process of such cooperation be organized? (Section 4). Together, these elements provide the contours of a comprehensive framework that can guide cooperation across the globe.

The need for global cooperation on AI can be argued from various angles. Bearing in mind the dual aim of maximizing AI’s benefits for humanity while preventing and minimizing its risks, in **Section 2** we set out the importance of steering the relevant actors from a finite to an infinite mindset. The current ‘race to AI’ risks spurring short-term and protectionist approaches that result in a zero-sum game whereby, in the long run, everyone’s gains are severely compromised. Considering the different motives that drive private and public actors towards competition, cooperation and cooptation, we argue for a level playing field around commonly agreed values for the development and deployment of AI. Such level playing field can not only provide a cross-border layer of protection against AI’s risks, but also opens up the conditions for beneficial competition to foster innovation and materialize AI’s benefits for all.

In **Section 3** we list the areas of cooperation that should be prioritized to achieve this goal, bearing in mind the resource-intensiveness of cooperation

initiatives. We argue for a holistic approach with a dual dimension. The first dimension is *horizontal*, and focuses on laying down minimum requirements to enable a trustworthy socio-technical environment around AI in a transversal manner. Cautioning against a myopic view, we believe this comprises not only the environment around AI-systems, but also the interwoven socio-technical environments of data and digital infrastructure. The second dimension is *vertical*. In parallel to establishing trustworthy environments, cooperation should also focus on domain-specific areas that require a tailored approach to maximizing AI's benefits, preventing and minimizing its risks, and addressing AI-related issues that arise ad hoc.

Having established *why* and in *which* areas global cooperation on AI should take place, **Section 4** addresses *how* such cooperation should be organized. Preliminary, we raise the importance of striking the right balance between the need for speed given AI's fast-paced roll-out and impact, the need for a holistic approach and the need to consider the context-specificity of AI's concerns. Against that background, we urge cooperation partners to clarify the rules of engagement in advance and in an open manner. Moreover, we propose favoring existing cooperation fora rather than the creation of new ones – all the while creating a network of networks across initiatives to allow for collective know-how to be built up and disseminated. While the need for speed imposes openness to differentiated cooperation, it remains essential for cooperation to be organized in an inclusive way, with a transparent way of working and mindful of power imbalances. Finally, we emphasize the need to establish accurate information streams to inform discussions, to set up feedback loops that improve cooperation outcomes and to secure AI-education for future generations who will need to continue today's cooperation efforts.

There is no silver bullet to ensure the success of global cooperation initiatives on AI. Nevertheless, by reflecting on an overarching framework in which such cooperation can take shape, we hope this chapter provides some useful indications.

2 The need for global cooperation on AI

In this section, we provide a number of reasons to stimulate and accelerate global cooperation on AI and its governance. Notably, the underlying motivations for setting up cooperation initiatives typically also inform their scope and process. We first consider what distinguishes AI as a technology so as to warrant new governance mechanisms (2.1). Second, we assess why these mechanisms require cooperation at the global rather than merely local level (2.2). Finally, we list some of the challenges that may need to be overcome by aspiring global cooperation partners (2.3).

2.1 Why AI necessitates governance

Already in 1986, Melvin Kranzberg stated that technology is neither good nor bad, nor neutral [32]. The applicability of this statement is equally true today. Technology is a tool created by human beings. It is shaped by the values of its makers and of the society they live in. Once technology becomes embedded in society, it can also start shaping society in turn. To ensure that this mutual shaping process respects and fosters appropriate values, governance mechanisms have been established to steer human behavior when developing and using technology. These governance mechanisms consist of many different tools – from non-binding organizational guidelines to sectoral standards to (inter)nationally applicable regulations. They can be applied *ex ante* (prior to being circulated on the market) or *ex post*, and are typically tailored to the technology and/or risks at stake. This also holds true for technologies based on AI, which are already covered by many existing (binding and non-binding) governance regimes, most of which predate AI's wide-scale incorporation in our lives. As a consequence, some of these regimes do not adequately capture the opportunities and risks that AI-systems generate in light of their specific properties.

There seem to exist as many definitions of AI as there are people discussing it, with each definition laying emphasis on different aspects. Moreover, definitions are typically guided by the context in which they are established [6]. Nevertheless, regardless of the definition one upholds, AI-systems share a number of distinct properties that are generally accepted to constitute both its strengths and weaknesses. It is, for instance, precisely AI's ability to reason and learn autonomously, and to subsequently act thereon in an autonomous manner in the physical or digital world, that renders it so useful for human beings – as well as hazardous. The same can be said for AI's evolving nature, its remarkable speed in processing information, and the immense scale on which it can be deployed. Consequently, when we get it right, AI's properties can generate significant benefits that positively impact not only a happy few, but humanity at large. However, when we get it wrong, the very same properties are capable of causing significant individual and societal harm. AI-systems do not operate in a void, but are always part of the environment in which they are developed and used [25]. Therefore, the risks they entail are not limited to the technical realm, but also encompass ethical, legal, social and economic challenges that must be dealt with [9]. Most of these challenges are not new, yet due to the properties specific to AI they can manifest themselves in novel ways [50, 63]. In particular, AI-systems can pose new threats to human rights, the democratic process and to the rule of law [4, 43, 44]. Just as the decision to design, develop and use AI is intrinsically human, so is the decision to set the conditions under which this should occur, and the responsibility for those decisions [49, 50].

As a consequence, an increasing number of actors (often as part of existing regional or global cooperation initiatives) have called upon the establishment

of governance mechanisms to secure the prevention and minimization of these risks, while at the same time fostering the maximization of its benefits. The European Commission’s High-Level Expert Group on AI [25] as well as the OECD [46] for instance urged for measures to ensure “Trustworthy AI”. This trustworthiness should emanate from the (demonstrable) fulfillment of requirements around the development, deployment and use of AI-systems by human beings. Hence, whether the term used is “Trustworthy AI”, “Responsible AI” or any other denomination is of less importance than the actual requirements this term embodies.

As defined by the European Commission’s expert group, the term “trustworthy” denotes the need to be (1) lawful, complying with all applicable laws and regulations; (2) ethical, ensuring adherence to ethical principles and values; and (3) robust, both from a technical and social perspective [25]. Throughout this chapter, the term *trustworthy* will be used as encompassing these three components. All three are necessary but not sufficient in themselves; they should work in harmony, overlap in their operation and be updated to ensure this harmony. To this end, existing governance mechanisms should be carefully scrutinized and, where needed, reshaped or complemented with new ones [52]. Fostering the trustworthiness of AI is not only essential to secure legitimate trust in the fact that, when AI-systems are used, this happens in a legal, ethical and robust manner. It is also required to enable AI’s benefits, as a lack of trustworthiness will stand in the way of AI innovation and uptake by citizens and consumers, companies and institutions - and hence of the materialization of its opportunities.

2.2 Why AI’s governance requires a global approach

Over the last few years, numerous countries and regions established their own AI-strategies. These strategies typically focus on incentivizing local stakeholders to join forces and maximize AI’s benefits, and – to more or lesser degrees – to minimize the risks raised thereby. While local initiatives are necessary, they are not sufficient to duly tackle the challenges and opportunities at stake and to secure AI’s trustworthiness [18]. In a globalized world, countries are increasingly interdependent. The policy choices pursued by one country can thus have a significant impact on others – directly and indirectly. This impact is particularly pronounced in the context of AI, a technology transforming entire economies and societies, with clear cross-border effects [28, 40, 64]. Consequently, the need for global cooperation on AI and its governance arises in a growing number of areas.

Global cooperation is an elusive concept and can be interpreted in multiple ways. Given that instances of truly ‘global’ cooperation are limited, one could even question the concept of ‘global cooperation’ as such. When referring to global cooperation throughout this paper, we intend to denote cooperation that

takes place internationally or across-borders, with an as large as possible number of international actors that share certain interests, objectives or values, driving them to cooperate. It is, moreover, important to note that cooperation initiatives take place not only at the level of states, but also at stakeholder level, including for instance private companies, public institutions, research and academia, civil society organizations and individuals. In fact, cooperation increasingly takes place through mixed models, whereby governments and other stakeholders are represented around the same table.

The relationship amongst these actors can take various shapes and typically depends on the specific context. Drawing on insights from industrial organization, roughly three types of relationship can be identified: a relationship of competition, a relationship of cooperation and relationship of cooptation (or cooperative competition) [3]. Each of these relationships has a role to play in enabling AI's trustworthiness, depending on the issue at stake. Cooperation is not a goal in and of itself, and is not necessary for each and every AI-related aspect. However, even in areas where competition is preferred over cooperation – for instance in light of the stimulation it can provide to socially beneficial innovation – a certain level of pre-emptory cooperation may be needed to secure that such competition can take place under fair conditions and to the ultimate benefit of all.

As has been raised by various authors, when it comes to the domain of AI, the current global landscape is marked primarily by a relationship of competition, particularly in light of the so-called 'race to AI' [3, 22, 35]. Virtually all national AI-strategies emphasize the desire to develop, maintain or strengthen a position of 'leadership' in developing and using AI, often with explicit references to the comparative position of rival states. And while this race to AI must not necessarily lead to a race to the bottom that sacrifices aspects like quality, safety and ethical values [51], such a scenario may nevertheless materialize if driven by the belief that AI's opportunities are part of a zero-sum or finite game, with set winners and losers [7]. This belief may lead to a unilateral focus on the state's own interest, triggering short-sighted – and often protectionist – measures, as well as a disregard for the negative externalities of its policies. A zero-sum game not only limits the scope of potential cooperation areas, but also hampers fair competition and undermines the incentives it can provide for beneficial innovation. As a result, the finite mindset may well become a self-fulfilling prophecy in which, in the longer term, everyone loses. Conversely, if a shift in perspective can be secured towards an infinite mindset, the protectionist approach can be cast aside and a reorientation can take place towards long-term growth and sustainable well-being.

Achieving this shift, however, requires a collective effort to establish a level playing field based on a common set of values to be respected when developing and using AI, and safeguarded by appropriate governance mechanisms. Once established, such a level playing field can protect citizens, as well as creating

the conditions for healthy competition that allow AI's benefits to be augmented, scaled and widely disseminated. Through common rules and standards, a space of mutual trust can be fostered, ensuring that AI-products and services can travel across the globe without crossing red lines that may cause individual or societal harm. In addition, global cooperation can leverage know-how and capabilities from multiple actors and orient these toward beneficial applications that are not captured by market incentives. In this way, cooperation initiatives can help secure that AI's opportunities are enjoyed by humanity at large – and by those who would benefit from it most – rather than solely by the traditionally privileged.

Global cooperation on AI should not be equated with harmonizing regulation. Cooperation on AI can also meaningfully take place around non-regulatory areas, such as for instance the incentivization of cross-border research collaborations in AI for social good. Moreover, it should be stressed that not all aspects of the development and use of AI must be governed by regulation, nor must this necessarily occur at the global level. AI raises different challenges for different countries, some of which are better dealt with in a manner tailored to the local situation [51]. A balance must thus be found, whereby meaningful cooperation for the benefit of humanity is fostered, without overlooking the particular circumstances of individual states.

2.3 Challenges to overcome

Global cooperation is both desirable and necessary to reach the aims set out above, and some actors even explicitly included the intention to engage therein in their AI strategies [15, 68]. However, enabling such cooperation is not devoid of challenges. The urge to compete rather than to cooperate – and to prioritize one's own interests to the detriment of longer term and increased benefits for all – is pervasive at all levels that require cooperation. This encompasses not just the geopolitical level, but also the level of stakeholders (companies, public institutions, researchers), the level of international organizations that bring these stakeholders together, as well as within single organizations. Accordingly, several obstacles will need to be overcome.

First, as was already raised above, there is currently no universally agreed definition of AI. Moreover, there is no such thing as a single AI, as various techniques and application domains of the technology exist, each with their own benefits and risks. Furthermore, these techniques and applications continuously evolve. Given that different actors can interpret the scope of AI in different ways – and are sometimes incentivized to do so [51] – there is a risk of misalignment when cooperation initiatives are established. At the same time, rather than focusing on a strict delineation of AI, we propose to consider a holistic approach to the technology, as outlined in more details under Section 3.1, thereby largely overcoming this definitional obstacle.

Second, the stability of cooperation initiatives typically hinges on the underlying motives that drive actors to cooperate rather than compete. Generally speaking, competition can occur at two levels: at the level of values and at the level of markets. With the former, we refer to the political and ideological values underpinning the organization of national or regional societies. With the latter, we refer to the space of trade in which economic (public or private) actors are engaged. Many national and regional AI-strategies have focused on positioning their jurisdiction – including their citizens, companies and organizations – in the best possible position to compete in and benefit from the global AI market. Such competition is typically not limited to the trading of goods and services, but also extends to establishing the best regulations and standards for AI, with the aim to turn national standards into the global norm. Competition also focuses on attracting AI resources, such as AI researchers and developers, AI-developing companies, AI-enabling infrastructure and financial investments – all of which influence one’s position on the global AI market.

As long as the competing actors have a similar underlying value-system, these values will also be reflected in their way of competing. This opens up a basis for cooperation regarding the conditions that AI’s socio-technical environment should meet to ensure that competition ultimately benefits all. However, when actors have differing value-systems, overcoming competition in favor of cooperation is significantly more complex. Substantive areas for cooperation can still be identified in those instances where economic interests, or the desire to attain a specifically delineated objective, may diminish the importance of the value discrepancy. Yet the agreement that can be reached in such scenario risks not only being less far-reaching but also less stable, as any change in those interests – whether through internal or external factors – can terminate the underlying motive for cooperation altogether.

Third, even where value-systems are more closely aligned, global cooperation on AI can be complicated by differing priorities.¹ While Artificial Intelligence is a subject appearing on the agenda of most (geopolitical) actors, it is far from the only one. The fact that AI-based technologies are already transforming our lives on numerous fronts does not render other transformative (human-made or natural) phenomena less important to focus on. As the establishment of cooperation initiatives requires both time and resources, engaging in cooperation on AI means less resources are available for (cooperation on) other goals. Moreover, even when sufficient partners are ready to spend their resources on AI-related cooperation, the partners’ priorities may not necessarily align when it comes to the substantive domain or concern that should be ad-

¹ In this regard, we also refer to the recommendations we propose in the final section of this chapter, focusing in particular on clarifying the rules of engagement between cooperation actors and ensuring a transparent cooperation process.

dressed first. It can in this regard also be noted that, while individual and societal interests impacted by AI often coincide, under some circumstances, these interests can instead collide. Depending on the underlying societal values and their hierarchy, cooperation actors may be driven towards different approaches when faced with colliding interests.

Finally, and closely linked thereto, is the differing social and economic conditions of the cooperation partners. Not all actors – whether countries or stakeholders – have an equal starting position when engaging in cooperation. Those with less financial means or in more vulnerable circumstances not only encounter more difficulties in finding a seat around the cooperation table, but might also struggle to find cooperation initiatives that are sufficiently tailored to their particular needs. It is evident that the challenges and opportunities faced by countries who still lack basic connectivity infrastructure will be different from the challenges and opportunities of highly advanced nations that dominate the global AI landscape. This uneven position will have an unavoidable impact on their respective priorities. Global cooperation on AI can help bridge the digital divide and is necessary to secure that AI is not used in a manner that further deepens it [28]. At the same time, efforts will be needed to ensure that potential differences in value-systems, political priorities and socio-economic conditions can be overcome to bring as many as possible global actors together.

3 Areas for global cooperation on AI and its governance

Having established *why* global cooperation on AI and its governance is needed, we can now focus on *what* areas should be cooperated on. AI is a multifaceted technology and can be used in a myriad of manners and domains, for better and for worse. As the establishment of global cooperation initiatives is resource-intensive, prioritization of cooperation goals is needed. Not all aspects surrounding AI's development and use need to be addressed through global cooperation. As a general rule, we believe that areas of increased risk require increased cooperation, and should be addressed as a matter of priority. Setting commonly agreed rules in this regard is essential to prevent and minimize AI's risks not only at local but also at global level. In addition, we also consider cooperation essential to establish a level playing field that is based on a shared set of values. This will not only secure citizen protection across the globe, but also enable socially beneficial innovation and stimulate healthy competition to disseminate AI's benefits.

To achieve this, we argue for a holistic governance approach, along a dual dimension. The first dimension is **horizontal (3.1)**, and aims at identifying minimum requirements for a level playing field to secure trustworthy AI in a transversal manner. Not only AI-systems, but the entire socio-technical environment

around such systems should be considered (a). Moreover, in addition to the environment of AI-systems, it is essential that the interwoven socio-technical environments around data (b) and digital infrastructure (c) are also taken into account. While many cooperation initiatives around AI are still myopically focusing on AI-systems alone, it is only by considering these three environments collectively - the '*system-data-infrastructure trinity*' - that AI's trustworthiness can truly be advanced. The second dimension is **vertical (3.2)**, and focuses on domain-specific areas where cooperation efforts should be tailored to the context or sector. This encompasses areas in which specific benefits for humanity can be realized and maximized (a), where AI's risks must be prevented and minimized with more immediate urgency (b), and areas where the need for cooperation can arise in a more ad hoc fashion (c). We believe that the horizontal and vertical dimension of AI governance should be addressed in parallel.

3.1 The horizontal dimension of AI governance

What is it that makes us trust financial institutions sufficiently to hand over our savings to them? What is it that inspires our trust in the aviation system, so that we dare step on a plane flying over 10.000 meters above the ground? It is not the trustworthiness of the bank's staff, nor the trustworthy reputation of the airline. Rather, we trust the broader socio-technical environment around these systems. We know that financial institutions are subjected to regulatory requirements, that standardized procedures are in place to ensure the quality of their services, and that – in case something goes wrong – there is a possibility for redress to ensure we get compensation. Similarly, we know that airplanes are built in accordance with certain standards, that they undergo multiple verifications prior to and during their deployment, and that the pilots flying them have the certified competences to man them. Moreover, we know this not only for the banks and airlines that are established in our home country, but we trust these environments across the globe, due to the global cooperation on and harmonization of the relevant standards, processes and regulations.

By analogy, securing the trustworthiness of AI cannot be limited to considering individual AI-systems, but must extend to their broader socio-technical environment. We consider the socio-technical environment of AI to be an overarching concept, encompassing three distinct environments that each need to be rendered trustworthy: the socio-technical environment around AI-systems (a), the socio-technical environment around data (b) and the socio-technical environment around digital infrastructure (c) – or the '*system-data-infrastructure trinity*'. Each of these three environments form a distinct yet interlinked web, which means they ought to be considered holistically. For each, it must be assessed which minimum requirements the systems, processes and actors involved should meet to be rendered trustworthy, as well as the human skills that are necessary to meaningfully engage with these environments.

For some aspects, binding regulation will be the most appropriate governance mechanism to secure the aims sought. In this regard, we argue for a risk-based approach, whereby elements that carry a higher extent of risk for individuals and society should be addressed more stringently. For other aspects, however, different tools should be explored, such as voluntary standards or certification mechanisms. In each case, it must also be established to which extent the mechanism should be imposed *ex ante* or *ex post*, and which entity should be responsible for its enforcement. By arguing in favor of a holistic approach that covers the trustworthiness of the entire socio-technical environment around AI rather than just the system, and by considering the interwoven environments of AI-systems, data and digital infrastructure comprehensively rather than in isolation, we wish to counter the myopia that existing initiatives at times suffer from.

(a) Building a trustworthy environment for AI-systems

As explained above, regardless of their shape, AI-systems do not exist independently but are part of their broader socio-technical environment. They influence this environment and are influenced by it. To secure their trustworthiness, a systemic approach is thus required, focusing on the trustworthiness of all actors and processes that are part of this environment [25]. This includes *inter alia* the social, legal and economic context in which the systems are used, the design and technical specifications of the systems' software, the purpose for which they are deployed and the business model in which they fit.

Concretely, as a first step, the various actors and processes that are part of the systems' environment should be mapped. Specific attention is needed for people who may be negatively affected by AI-systems even if not directly engaging therewith. As a second step, minimum requirements that ascertain and enhance the legality, ethicality and robustness of the socio-technical environment of AI-systems should be identified. These requirements should reflect – by priority – those aspects posing the largest individual and societal risks, and those aspects needed to establish a global level playing field. Finally, the identified requirements should be globally agreed on and, where needed, complemented with new ones.

As many risks are context-specific, a large number of standards or procedures will need to be established at sector- or application-level, so as to tackle the risks specifically arising within that context. This, however, does not take away the need for standards and procedures that are context-agnostic and apply horizontally – especially since the same AI-system can be repurposed for different contexts. Regardless of the sector in which AI-systems are used, those interacting therewith or subjected thereto must be able to trust in the fact that basic safeguards are in place. For instance, they need know that certain guarantees are foreseen to protect their human rights, that the accuracy and robustness

of the systems are duly verified or certified by competent experts where necessary to safeguard those rights, that the necessary documentation and tools are available to render the systems auditable by independent authorities, that measures have been taken to counter the impact that these use of these systems may have at work or on the labor market, and that – in case something goes wrong – accountability and redress mechanisms are available.

We believe that finding agreement on these requirements is significantly more important than the quest for an agreement on AI's definition. They should be met regardless of whether it concerns a basic algorithm, a sophisticated deep-learning system or any other automated decision-making process that may or may not fall under the ever-changing AI definition. Safeguarding these minimum requirements can take numerous (complementary) shapes, from principle-based regulatory provisions to certifiable standards. In our view, both are necessary, with the former ideally informing the latter. As noted above, the risk-based approach that increases the stringency in case of increased risk can be of help. Several initiatives have started preparatory work in this field [11, 15, 27, 46, 48, 56, 74], increasingly also across borders. As AI-systems often consist of components developed or used in different countries, these requirements must be agreed on globally to ensure that they are met regardless of their place of development. Encouragingly, while each of the existing cooperation initiatives on AI highlighted slightly different aspects in light of their mandate, so far, most established similar outcomes. Nevertheless, many of the current requirements are non-binding and not yet sufficiently concretized to secure accountability by those involved. Moreover, few of them have sufficient attention to the necessary skills and competences that may need to be certified.

In addition, besides identifying the minimum content of these requirements, cooperation efforts must also focus on the processes that are necessary to demonstrate and verify the fulfillment of these requirements, for instance through standardized reporting or audit procedures.² We believe that agreement on these procedural matters is equally if not more important than agreement on the requirements' substance. To illustrate: two countries may, for instance, demand a different level of demonstrable accuracy of an AI-system. However, without a commonly accepted procedure on how the system's accuracy should be measured, demonstrated or verified, neither country will be able to assess compliance with its standards. Consequently, to avoid that trust in the socio-technical environment of AI-systems will remain local and volatile, agreement on the demonstrability and verifiability of the requirements' fulfillment is essential.

² It can be noted that a number of initiatives have started work in this regard, including ISO [70], CEN-CENELEC [65], ETSI [67] and NIST [72].

(b) Building a trustworthy environment for data

Whether it's referred to as the new oil or the new electricity, the role of data in our economies and societies is growing in importance, and this trend will undoubtedly persist. Not all AI-systems necessitate big data. Yet many AI-systems nevertheless heavily rely thereon. For those systems, the availability of qualitative, integer, comprehensive and representative data is necessary to materialize the benefits they can achieve [14]. Data plays an essential role throughout those system's entire lifecycle. At the same time, storing and processing data comes with significant responsibility [23, 24], even more so when it concerns personal data. In that case, such data can be considered as a constitutive part of a person's identity [20] and is therefore typically accompanied by a protective right of control thereof [38]. Given data's relationship to AI, securing a trustworthy socio-technical environment for data that enables its availability and protection across borders and secures the trustworthiness of all actors and processes involved, should be part of any global cooperation framework on AI.

This means that the parameters of data's socio-technical environment must be mapped, and that global cooperation on defining their substantive and procedural dimensions should be stimulated. While this mapping exercise is well underway - and actors on regional level (such as the EU) have started setting out some of the constitutive elements of a trustworthy data environment - political convergence on these issues at global level is still limited. Certain countries implemented particularly protectionist stances in this field, under the not always justified guise of national interest and security – thereby fostering a zero-sum game approach and creating obstacles for mutual trust and growth opportunities. Matters are further complicated by the large diversity of issues at stake. As data is contextual, certain requirements will need to be specified at domain-level. Yet the fact that it can be reused and repurposed for numerous applications – including applications that were not foreseen or foreseeable at the time the data collection took place – render horizontally applicable policies an indispensable base layer for a trustworthy data environment. To this end, and without the ambition of being exhaustive, we list here below a number of data-related issues that should be addressed in global cooperation fora.

First, a convergence of approaches should be sought on data's legal status, particularly including ownership and intellectual property rights thereon. Such convergence is not only useful as regards first-hand data, but also for secondary and tertiary data, including 'new' data that originates through the combination of other data points. Data's applicable property regime – which also governs its accessibility – typically differs depending on the category of data (personal, non-personal or mixed; pseudonymized or anonymized; public or confidential) and context (such as business-to-business (B2B), business-to-consumer (B2C), business-to-government (B2G), government-to-citizens (G2C) and other varia-

tions). Whereas in certain contexts – such as the B2B domain – ownership questions are often settled through contractual clauses, these matters are typically less straightforward when consumers or governmental actors are involved. Since different models are conceivable, fostering a mutual understanding around these issues at global level is fundamental, and can subsequently form a basis for further convergence where needed.

Second, and closely linked thereto, is the governance of personal data – a data category for which the term ‘ownership’ is, in fact, often deemed controversial for various reasons [26]. In recent years, awareness around the (mis)use of personal data, especially when emanating from consumers or citizens and used to influence their behavior, has steadily grown. The *MyData* movement and other initiatives greatly stimulated the empowerment of individuals through increased data autonomy [36, 71]. While codified in multiple human rights instruments, the right to personal data protection is not absolute; it must be considered in relation to its function in society and be balanced against other fundamental rights. Tensions can arise between individual and societal perspectives, for which concepts like data trusts and data commons have aimed to provide (partial) solutions. In addition, (extra-EU) countries formulated the right to personal data protection in different ways and with different safeguards, and some still lack any specific data protection law. Even in countries where personal data is protected, mechanisms for protection are not always efficient. The emphasis on self-determination through consent is, for instance, increasingly deemed problematic [8, 59], given the substantial asymmetries of information and power that are often at play and that may render meaningful consent an illusion. For AI-systems and their data to travel within a trustworthy environment, common definitions for data categories, common regimes for (shared) data ownership and protection, and common procedures to determine which regime is applicable, should be established across borders.

Third, individuals, objects and organizations produce millions of data points every day, that could be harnessed for a myriad of beneficial uses. These data points are, however, not always collected and - if they are - this collection sometimes only occurs by a small number of large players who currently dominate the data space. Data can not only provide significant value, but can also be used as a means of amassing power – whether in the private or public sphere [41]. Legal clarity on the property and use of data will not necessarily reduce the power asymmetry between consumers, citizens and smaller companies on the one hand, and incumbent entities on the other. Depending on the regional regulatory and market structure, governments can be one of these dominant entities or they can be entirely dependent thereon. Whichever the case, a concentration of power in the form of data hoarding rarely results in anything other than a zero-sum game with suboptimal outcomes. A lack of access to the necessary datasets to provide a service plays a major role in this regard. Consequently, cooperation to incentivize data sharing under fair conditions – including

through federated learning models – as well as measures to ensure that market concentration does not lead to (private or public) abuses, should be encouraged. Moreover, the creation of open (annotated) data sets, common data pools and data commons can help balance existing power asymmetries. Just as is the case for open source repositories of AI codes and the sharing of (already trained) neural networks, much work is still needed to foster cooperation on access to open data sets. Such access is particularly useful for researchers and not-for-profits who rely thereon to conduct fundamental research or work on socially beneficial projects. In addition, a comprehensive regime for data philanthropy within a trustworthy data environment can also stimulate the development of AI-applications that benefit humanity as a whole, rather than data incumbents [53].

Fourth, cooperation is needed on technical standards for the digitization, storage, processing and encryption of data. Data's interoperability and portability hinges not only on its legal and market environment, but also on the manner and format in which data it is collected and kept. Moreover, the value of datasets is highly dependent on their quality and integrity, for which common measurement procedures should likewise be developed. Besides agreement on the (legal) definition of anonymous and pseudonymous data, harmonized procedures on a given anonymization or pseudonymization method - including the potential demonstrability of compliance therewith - can likewise enhance trust. This also holds true as regards requirements for data encryption and security. Similar to the context of AI-systems, we believe the focus should lay not only on the substantive dimension of these requirements, but also on commonly agreed standards to measure, demonstrate and verify the steps and processes followed for their fulfilment.

Last, attention must be paid to the appropriate collection and use of data. The datafication of society is an increasingly common phenomenon, spurred particularly by the aim to realize the benefits that sound data analysis can provide [39]. At the same time, this has also led some to mistakenly believe that everything is quantifiable and measurable, and that the collection of sufficient data can unambiguously inform policy-making [47]. Data should, however, not be equated with objective facts. It concerns a (partial) representation of the world, which is always shaped by a specific interpretation and needs to be interpreted in turn to be of use [23]. Hence, to secure a trustworthy socio-technical environment for data, it is important to foster the necessary competences and literacy for the appropriate handling of data, including the awareness of its inherent limitations when used to draw conclusions. Especially when data-driven applications are used to inform policy-making, common guidelines on providing basic information – such as how much data a given dataset contains, how it was collected, whether and what kind of sampling was used, what data is missing, and in which manner data points are being used as proxies – would not be a luxury.

(c) Building a trustworthy environment for digital infrastructure

The importance of digital infrastructure as an enabler of various digital technologies has rendered it one of the most desired assets for both private and public actors. AI-systems cannot function without an underlying infrastructure that supports their development and use, and through which their accessibility towards users is enabled. Moreover, the data that AI-systems rely on depend on digital infrastructure for their storage, processing and transfer. As the deployment of remote tracing applications to fight the COVID-19 pandemic has shown, entities controlling the digital infrastructure do not even require AI-applications or access to (personal) data in order to nevertheless extract value therefrom and have an influence on societies [60, 61]. Consequently, securing a trustworthy socio-technical environment for AI must necessarily go hand in hand with securing a trustworthy environment for digital infrastructure.

The term digital infrastructure refers to a range of elements enabling digital services, including the internet backbone; the connectivity of systems through broadband, mobile telecommunications, Wi-Fi networks and communication satellites; cloud computing; data centers; platforms to develop and operate AI and other software systems as well as API's. Global cooperation in this field is rendered particularly difficult in view of its connection to discussions on sovereignty and national security. Geopolitical competition over infrastructure has proven to cause negative spillover effects on state relationships in other domains too, thereby affecting the potential success of cooperation in this field. It is, however, precisely because of digital infrastructure's importance that global cooperation on its environment must be sought. Besides the need to avoid a zero-sum game approach and a loss of benefits in the longer term, cooperation can also help counter the risk that negative consequences will be suffered primarily and most severely by those who are already more vulnerable. Convergence on the substantive outcomes that cooperation in this field should lead to is still difficult to define. Here below, we nevertheless list four aspects that we believe global cooperation fora should address.

First, discussion is needed on the evolution of the global digital infrastructure and its critical points of control. As a function of market and technological developments, a limited number of actors emerged – often referred to as ‘gatekeepers’ – that have a strong influence on these control points, as well as the conditions for access to the underlying infrastructure. The capability to influence the development of, and access to, digital infrastructure comes with significant responsibility, as the shape of digital infrastructure is capable of shaping society – a capability that can be used in ways that enhance the infrastructure's utility for the benefit of individuals, companies and other stakeholders, but also in ways that can harm them. Global cooperation initiatives on AI should therefore duly reflect on the governance mechanisms needed for the evolution of the digital infrastructures on which AI is run. Such governance

mechanisms should extend not only to the infrastructure's technical requirements, but should also focus on ensuring the responsibility of the actors involved, as they form an inherent part of the infrastructure's broader socio-technical environment. In particular, a dialogue is needed on (and with) the gatekeepers for internet access, for storage and processing capacity, for the accessibility of applications, for the digital marketplace and for news curation. Special attention should also be given to the gatekeepers of the infrastructures that shape digital identity (such as social media platforms).

Second, cooperation efforts should focus on bridging the digital infrastructure divide. While some countries are rolling out 5G, others still lack basic connectivity infrastructure [1]. This has a profound impact on the ability of the citizens in those countries to meaningfully participate in society, and on the economic progress that they can achieve. It is futile to speak of the many benefits that AI can generate as long as basic digital capacity is not secured. Global cooperation is hence needed to support countries across the globe with the build-out of digital infrastructure where this is lacking or underdeveloped. Only once such capacity is in place – as part of a trustworthy environment that governs it – can we truly start materializing the benefits of AI for humanity. Global cooperation initiatives such as the Digital Public Good Alliance [66] and the Global Data Access Framework [69] can help foster this [58].

Third, attention is needed for the environmental footprint of AI's digital infrastructure. Data centers, cloud services and connectivity equipment consume a large amount of energy and thereby significantly contribute to emissions, with the ICT sector as a whole being estimated to use around 5 to 9% of the world's total electricity and generating over 2% of all emissions [14]. It is projected that, in the next few years, ICT operations will start representing up to 20% of global electricity demand, with one third stemming from data centers alone [58]. To reduce our negative impact on the environment and tackle climate change so as to preserve our planet for future generations, states must take their collective responsibility – and enforce such responsibility also upon the private actors active or incorporated within their jurisdictions. In its Communication on “Building Trust for Human-Centric AI” [16], the European Commission already raised the need for AI developers and deployers to foster environmental well-being, which constitutes one of the seven requirements for Trustworthy AI put forward by its High-Level Expert Group. This need goes beyond AI-systems and also includes the broader digital infrastructure – and entire value chain – on which AI-systems rely. A trustworthy environment for digital infrastructure must secure attention to its sustainability and ensure that the benefits realized with AI today will not jeopardize the future of next generations. Given the cross-border nature of environmental harm, the environmental requirements that AI's digital infrastructure should meet must be agreed at the global level.

Last, we highlight the need for global cooperation on digital infrastructure’s security, especially when such infrastructure is critical. The importance of digital infrastructure for states’ daily operation – from the functioning of hospitals or transport services to services affecting national security – was already raised above. Evidently, the more we rely on digital infrastructure to operationalize our essential services, the more vulnerable we are when this infrastructure proves to be defective or comes under attack. Since an increasing number of private and public actors procure (part of) such infrastructure abroad, it is essential that the trustworthiness of this infrastructure can be ensured across borders. Beyond technical robustness and safety requirements for AI-systems, similar requirements should therefore be discussed for their underlying digital infrastructure.

3.2 The vertical dimension to AI governance

Building a trustworthy socio-technical environment around AI – covering the *system- data-infrastructure trinity* – can help ensure that it embeds appropriate values and that it is used in a manner that fosters rather than hampers human capabilities. A trustworthy AI environment can also help minimize the possibility that the technology is underused out of mistaken fear, a lack of clear rules or other obstacles that may result in opportunity costs [19]. At the same time, the risk remains for AI to be intentionally misused or to inadvertently cause harm. As the concrete materialization of AI’s opportunities and risks is often linked to the domain and application in which it is used, the horizontal framework that was set out above needs to be complemented by domain-specific cooperation initiatives, allowing for a more tailored approach where needed. Some of these domains can even start constituting (socio-technical) environments in and of themselves, for instance when linked to a specific sector such as healthcare, transport or education. Rendering these environments – and particularly the development and use of AI within such environment – trustworthy, will necessitate domain-specific expertise. Moreover, in some instances, a specific focus on one particular issue or problem – considered against a holistic background and from multiple angles – might be more conducive to finding solutions. Also here, we believe that an increased level of risk typically calls for an increased level of cooperation to assure a coordinated approach.

Domain-specific areas of cooperation can broadly be captured along three axes: areas aimed at maximizing AI’s benefits for humanity (a), areas aimed at preventing and minimizing AI’s risks for humanity (b) and areas dealt with on an ad-hoc basis, typically in light of new or unexpected challenges (c). Rather than listing the numerous vertical cooperation domains for AI, in what follows we briefly address these three overarching axes. It should be born in mind that the type of relationship that different actors are engaged in (competitive, cooperative or co-competitive) as well as the cooperation initiative’s organizational set-

up (which is further described under Section 4) will to a large extent shape – and be shaped by – the substantive domain in question.

(a) Cooperation on maximizing AI's benefits for humanity

AI can enable a multitude of beneficial applications. The fact that the technology is being oversold (often for commercial reasons) can at time lead to unfeasible expectations. Nevertheless, a realistic perspective of its capabilities acknowledges AI's potential to contribute to individual and societal well-being in numerous ways. To maximize AI's benefits for humanity, we see the need for cooperation on two main fronts.

First, cooperation is needed to secure the minimum requirements for a level playing field that provides a base layer of protection, and that enables market competition to take place under fair conditions. By eliminating distortions to competition and ensuring that all actors can contribute, such a level playing field can stimulate and incentivize beneficial AI research and innovation and ensure that the fruits thereof can be accessed by all. Achieving this result requires not only transversally applicable rules, but also attention to the developments taking place in specific sectors and domains.

Second, cooperation is needed to materialize socially beneficial AI-applications that are not fostered by market competition. While claims are being made about AI's benefits, it is not always specified who the beneficiary of those benefits are. This question is rightfully receiving renewed attention, and efforts are increasingly focusing on the benefits' wider dissemination. Moreover, as already noted above, situations can arise whereby one group of individuals benefits from an application while another group is harmed, and where individual benefits are traded off against societal harm – or vice versa. These situations are not always straightforward, and the perspective with which one looks at them is often colored by the underlying value-system of the onlooker. Nevertheless, global cooperation can help ensure that the beneficiaries of AI's benefits also encompass those people who are most in need.

Initiatives aimed at developing so-called 'AI for social good' applications have been mushrooming [2, 12, 62], and are often aimed at advancing one or more of the UN Sustainable Development Goals and its 169 targets [57]. Examples are AI-applications that help reduce humans' negative impact on the environment, render the provision of healthcare more accessible, optimize the allocation of scarce resources in developing countries or foster educational opportunities. Cooperation initiatives can bring together the necessary systems, data and infrastructure to build and scale such applications in various domains. In this context, it is however important to be attentive to local and cultural specificities, for instance by having due regard to the preservation of local languages and customs. Moreover, it is not because AI applications are not built for profit, that their careless use cannot cause harm. Accordingly, it is essential that these

applications – however beneficial their intended aim may be – meet all relevant requirements and are built with the help of domain experts that can steer them towards their most relevant uses. As AI is but a tool towards an end, cooperation efforts to maximize AI’s benefits for humanity in specific domains should first clarify the goals to be achieved, and only then identify to which extent AI can help therewith.

(b) Cooperation on preventing and minimizing AI’s risks for humanity

The use of AI can lead to several types of harm. As with any technology, this harm can stem from a malicious use of AI and be caused intentionally, but it can also arise out of negligence. It can affect individuals and groups, private and public organizations, as well as entire economies, societies and humanity at large. Moreover, in certain situations, the only way to detect and counter AI-driven harm is by using AI-applications as a shield and countermeasure. Given AI’s digital nature, the adverse effects of its negligent use and misuse are typically not limited to the country in which it is developed, but can easily wreak havoc in other countries too – or can even be built to do precisely that. No country can control these risks by itself. Therefore, the prevention and minimization of AI’s risks for humanity is par excellence an area for which global cooperation is not a choice, but an absolute necessity. For these reasons, competition between global actors may more easily be overcome in this field, as competitors are naturally forced to seek mutual aid and protection from those risks, which opens the door to cooperative and cooperative relationships.

The various manners in which the use of AI can harm humanity can be linked to a specific domain or application, and hence often require tailored approaches. These include, for instance, the use of AI in cybersecurity attacks that target states’ critical infrastructure [54, 55], the use of AI for disinformation purposes which threatens democracies [4] and, more broadly, AI’s dual-use risks [5]. The use of AI in the military, especially when having the potential to cause lethal effects, requires heightened attention [42]. Across the globe, calls are also increasing made by civil society organizations and private actors alike to regulate the conditions under which AI-enabled (remote) biometric identification should be deployed. In addition, consideration should be given to critical risks for humanity in the long term.³ Several international fora have ongoing discussions on some of the above risks. However, we believe that the speed and efficiency with which (state) actors are currently engaged in these discussions is not proportionate to the risks involved.

³ A number of research centers scattered around the world have the specific aim (and funding) to develop applications of Artificial General Intelligence and Artificial Consciousness. Regardless of the likelihood that this aim proves realistic or successful, the magnitude of risk that would be raised thereby calls for a common global stance and vigilance.

(c) Cooperation on ad hoc matters

Besides domains that can be anticipated, periodically the need will arise to organize global cooperation on AI-related issues in a more ad hoc manner. These can concern AI's positive or negative consequences – or both. The developments surrounding the COVID-19 pandemic for instance made it clear that, in times of crisis, AI-systems can be used for numerous socially beneficial applications [37], that the exchange of qualitative data can contribute to saving lives [31] and that the control points of the infrastructures on which AI-systems run (including AI-driven tracing applications) are more crucial than ever [61]. At the same time, AI-systems have been used to fuel the spread of mis- and disinformation about the virus, the harm of which goes far beyond the medical realm. It is the collective task of all actors involved to remain attentive for new matters that may require the setting aside of competition in favor of ad hoc cooperation at the global level. While the development and use of AI still brings with it many unknown unknowns, the establishment of sustainable networks of global cooperation can help foster their anticipation.

4 Organizing global cooperation on AI and its governance

The annals of global cooperation precede the second world war, yet it is in the aftermath thereof that such cooperation truly started taking off [17]. This leaves us with a long and rich experience of how global cooperation has been organized – on issues as diverse as nuclear weapons, aviation safety, space or the law of the sea – which can be drawn from in the context of AI. Many existing cooperation initiatives already touch upon matters that are (directly or indirectly) related to AI's concerns, from their own perspective. These initiatives have beaten tracks, distinct ways of working, ingrained procedures and established actors. As global cooperation on AI is thus embedded in a pre-existing setting, a close link exists between *how* such cooperation is organized (the process and shape) and *who* is involved (the cooperation organizer and partners). Inexorably, the cooperation process will be influenced by the actors around the table as well as by the type of table. Bearing this interwovenness in mind and acknowledging the numerous actions that are already taking place the field of AI, in this section we put forward seven elements that can help guide the organization of both existing and new global cooperation initiatives.

In particular, we address the need to: balance speed, holism and context-specificity (4.1), clarify the rules of engagement of cooperation partners (4.2), favor cooperation through existing fora rather than creating new ones (4.3), develop a network of networks to build up collective wisdom (4.4), being open to a model of differentiated cooperation (4.5), secure an inclusive and transparent way of working, mindful of power imbalances (4.6) and establish feedback loops while preparing for cooperation by future generations (4.7).

4.1 Balancing the need for swift action, a holistic approach and attention to context-specificity

As the scale and pace with which AI is being developed and used is growing, so is the need for swift cooperation. Consequently, a red thread that should shape the organization of global cooperation on AI is the need for speed and efficiency. It has become increasingly clear that the transformative influence of AI goes well beyond the simple sum of the influence of individual AI-systems. While we are still in a steep learning curve as regards the technology itself and its impact on individuals and society, AI-systems are already being rolled-out in virtually all domains of our lives. Speed is therefore of the essence to secure that this roll-out occurs in a manner that fosters AI's benefits in the long term and in a manner aligned with our values, rather than in a manner hampering them. The importance of speed, however, needs to be balanced off against two other elements, namely the necessity of a holistic approach on the one hand, and sufficient attention for domain-specific concerns on the other.

The holistic approach we argued for above requires attention not only to the socio-technical environment of AI-systems, data and infrastructure, but also to the large variety of angles through which AI's challenges and opportunities can be considered, necessitating the diversity and inclusivity of cooperation partners. At the same time, AI's challenges and opportunities are often context-dependent, requiring appropriate attention for domain-specific concerns that require more tailored measures. These three elements – speed, holism and context-specificity – are, evidently, not always harmonious. The more parties at the table and the more pieces of the puzzle to consider, the less likely it is that the desired speed of action can be attained. The shape of the cooperation initiative, as well as the type and number of actors taking part, will have an inevitable influence on the speed of the process. A careful balancing exercise is hence needed, requiring often difficult trade-off decisions from the global cooperation partners. For areas where the risks are particularly extensive and irreversible, speed may temporarily need to be prioritized over holism so as to secure a fast base layer of protection.

4.2 Clarifying the rules of engagement

Setting up a global cooperation initiative – whether through a multi-stakeholder partnership or intergovernmental process – is typically a time- and resource-intensive process. To avoid a waste of these resources and enhance efficiency, the rules of engagement between cooperation partners need to be clarified and rendered explicit prior to the cooperation's initiation. If the initiative is to be fruitful, all of its participants need to be seriously committed to the defined mandate and goals, and must communicate their expectations in advance – both internally and externally – to allow for common engagement rules to be agreed

on. The rules of engagement will likely depend on the substantive area of cooperation, and on whether the actors start the process from a stage of competition, cooperation or cooperation.

As noted under Section 2, for global cooperation initiatives to succeed, competition need not always be entirely eliminated. When competition takes place on top of a level playing field based on common values, it can foster beneficial AI-applications and stimulate a race to the top – in terms of both regulation and innovation. However, when competition is taking place at the level of value-systems, it typically affects the partners' relationship far beyond the field of AI, thereby rendering cooperation more difficult. Potential cooperation partners should therefore reflect – and come clean – on the type of competition at stake and the envisaged relationship within a cooperation initiative. The underlying value-systems governing the political, social and economic views of the respective partners will necessarily contribute to the manner in which the cooperation rules of engagement will be shaped, thereby determining not only the initiative's outcomes but also its durability.

4.3 Building on existing cooperation structures

Over the last century, numerous international organizations were set up with the specific aim of fostering cooperation amongst states and other stakeholders, each with distinct compositions, mandates and collaboration processes. Many of these organizations are in the position – or already took decisive steps – to include the issues raised by AI on their agenda. Some of these organizations have a working scope that also encompasses issues around data and digital infrastructure, while others are solely focused on the immediate environment of AI-systems (or even narrower, on maximizing its benefits through 'AI for social good' projects). Therefore, prior to establishing a new cooperation structure, it is advisable to map existing cooperation initiatives on AI, so as to assess the new structure's added value. Given the need for speed, we believe that – as a default – the use of existing cooperation initiatives should be favored over the creation of new ones. This is particularly the case when those organizations have a broad mandate and can thereby foster a more holistic approach. Of course, in order to enable fruitful cooperation on AI and its governance, these existing initiatives may need to include new and more diverse cooperation partners, and will need to ensure the involvement of domain experts – yet many of them have set procedures in place to do so.

Only where, after a thorough mapping exercise, existing cooperation initiatives appear to fall short, can the establishment of new cooperation fora be justified. This is for instance the case when no existing organization has a mandate to address the relevant issue, and where obtaining such mandate would be an unduly lengthy process. A new initiative could also be justified when no exist-

ing organization can engage in swift and fruitful cooperation on a specific matter due to burdensome, inefficient or non-inclusive processes. Each time a new forum is created – or a new mandate is defined within an existing cooperation forum – the delicate balance between speed, holism and context-specificity as outlined above must be considered.

4.4 Developing a network of networks

In some instances, the working scope and mandate of existing cooperation initiatives can (partially) overlap. Overlapping mandates have not only led to organizational competition, but also to geopolitical competition within these organizations. Indeed, in light of the first-mover advantage for (state) actors in terms of regulatory impact, it has been argued elsewhere that the race to AI also brought forth a race to AI regulation [51]. This regulatory race at organization-level is at least in part fueled by the differing composition of the organizations in question, though competition can also take place between organizations with similar memberships.

While these overlaps carry a risk of duplication of work, inconsistency and waste of scarce resources, they also secure that the same subject matter is looked at from different perspectives, which may limit the risk of gaps. Thus, the Council of Europe will for instance have a particular focus on the human rights risks raised by AI, while the OECD's perspective will be colored by the lens of economic progress and world trade. Both organizations may converge on the need for legal, ethical and robustness requirements for AI, yet they will consider this need from their specific angle. A multiplicity of initiatives can also avoid the risks of a herd mentality or of a concentration of regulatory power at the global level, which could start from the best of intentions but evolve into an undesirable regulatory monopoly. In this sense, the various organizations in the global cooperation field can maintain a balance of power amongst them and – in the ideal case – work towards the most optimal outcomes by competing with each other in a race to the top.

Where cooperation mandates overlap, it is however important to secure the dissemination and cross-pollination of information between the relevant initiatives. This will enable a mutual learning process and thereby maximize the chances of a regulatory race to the top – effectively establishing a relationship of co-competition between these organizations rather than competition. The dissemination of information can take many shapes, but should ideally be as speedy and as efficient as possible. In practice, existing initiatives within international organizations have for instance allowed other organizations to partake as an observer. Thus, the Council of Europe's Ad Hoc Committee on AI (CAHAI) counts the OECD, the European Commission, UNESCO and the UN Panel on Digital Cooperation – each having partially overlapping mandates – as observers in its meetings. The status of observer has also been granted to

states that do not have a Council of Europe membership, yet have an interest in its work, and to other types of stakeholders. In turn, the European Commission's High-Level Expert Group on AI likewise comprised observers representing (non-member) states, stakeholders and international organizations - including the Council of Europe and the OECD.

The practice of inviting observers from regulatory competitors generates a network effect that allows for collective knowledge building – both on the subject matter and on cooperation processes – thereby benefiting all actors involved. In addition, some organizations have set up more concrete mechanisms for information sharing. For instance, in 2019, the OECD launched an AI Policy Observatory [73] in cooperation with the European Commission, containing a repository of state- and stakeholder policy initiatives around AI. Such information gathering and dissemination helps compensate the unavoidable overhead costs that a multiplication of work generates, and contributes to the mutual advancement of cooperation efforts. By maintaining a dialogue across cooperation initiatives, a network of networks can be created and informal coordination can be secured.

4.5 Maintaining openness to differentiated cooperation

Given the borderless nature of AI's positive and negative impact, cooperation initiatives should ideally gather the widest group of (geopolitical) actors, aiming towards truly 'global' participation. However, and particularly in light of the potential value-based discrepancies highlighted above, this may not always be possible or desirable – nor is it always strictly necessary to achieve relevant outcomes. The European Union is a primary example of a cooperative framework in which differentiated integration has proven to be essential to advance multinational agreement on certain matters. Essentially, the EU has inbuilt mechanisms that allow for countries to opt-in and opt-out of a cooperation regime for certain substantive matters, and for a smaller group of countries to pursue closer cooperation in some domains when other countries are hesitant to join [13, 33]. In this manner, enhanced cooperation need not be delayed but can be initiated by like-minded countries – with an open invitation for other countries to join the cooperative regime at a later stage.⁴ When modelling this feature to the global stage, openness to differentiated cooperation in the context of AI can prove an asset for countries willing to invest time and resources in finding agreement on certain issues amongst themselves.

At the same time, it should be born in mind that the enhanced cooperation by some countries is liable to create externalities upon those remaining outside

⁴ It should be noted, however, that the EU Treaties only allow for the use of this mechanism when no agreement can be found with all EU member states. At the same time, EU countries have also pursued 'differentiated cooperation' outside the scope of the EU Treaties.

the cooperation framework. Those externalities can be both positive or negative. When positive, external countries are only likely to join the initiative at a later stage if the benefits they can gain from joining minus the costs thereof outweigh the positive externalities. When negative, external countries are only likely to join if the costs of joining are lower than the negative externalities they face by remaining outside. Alternatively, to the extent the costs of doing so are lower than the costs of joining, they may also choose to boycott the enhanced cooperation initiative through political, economic or other means. In this regard, it is also important to consider the unequal geopolitical and economic circumstances in which countries at the global stage find themselves. Sensitivity is needed for asymmetries of bargaining power – including when it comes to accessing the cooperation table.⁵

Being mindful of the above, we believe two approaches ought to be combined. On the one hand, cooperation should be sought with an as large as possible group of partners, especially to seek agreement on the requirements for trustworthy socio-technical environments around AI's *system-data-infrastructure* trinity. As a general rule, cooperation initiatives should be open and collaborative, and welcome new partners that share the initiative's vision, values and goals. The size of such a group and the diversity of actors may make it difficult to achieve consensus that is far-reaching in the short term, yet the immediate focus should lay on identifying minimum requirements allowing a basic level-playing field that can be further built upon. Once started, the trust-building through this cooperation process can grow, which – ideally – gradually increases the convergence of views and extends the cooperation scope. On the other hand, and simultaneously so, like-minded partners willing to seek closer cooperation and reach further consensus should advance through a differentiated cooperation approach. When doing so, they should however secure consistency with the outcomes of wider cooperation initiatives. Even when starting off at smaller scale rather than at the global level, such cooperation can have a positive impact and act as a catalyst for broader agreement in the longer term.

4.6 Securing an inclusive and transparent way of working, mindful of power imbalances

For cooperation to be both swift and efficient, the teams representing the cooperation partners and partaking in the negotiations should have a clear mandate

⁵ Such power asymmetries are often accompanied by an asymmetry of the (negative) consequences endured by the development and use of AI at global level. Attention should therefore be paid not only to the opportunities and risks generated by AI, but also to the potential shifts in power that AI may induce – whether between and amongst countries, or between and amongst public actors, private actors and individuals [24, 30].

to conduct negotiations, as well as the required knowledge of the area of cooperation. Given AI's multifaceted impact, a multidisciplinary approach is required, allowing for the relevant issues to be considered from all relevant angles. This necessitates inclusive, diverse and gender-balanced negotiation teams, and specific efforts to ensure participation from multiple stakeholders, ideally comprising a balanced representation from the public sector, the private sector, research and academia, and civil society organizations. In addition, cooperation initiatives need to be equipped with sufficient resources, and secure clarity and transparency on monetary contributions in advance. Mechanisms must also be foreseen to ensure that those actors willing to participate but lacking the necessary financial means – particularly when it concerns developing countries or civil society organizations – are not hindered thereby. Not only those with the ability to accelerate or contribute AI-capabilities, but also those meant to benefit therefrom – and that may be adversely affected thereby – should have a seat around the table.

In this regard, specific consideration should be given to the manner in which past injustices might shape today's power asymmetries. Often still, the discourse within global cooperation initiatives on AI and its associated values is driven by more economically developed countries [29], reflecting a power imbalance to which past geopolitical developments directly or indirectly contributed. There is a risk that these power structures are maintained within cooperation initiatives, even when aimed at outcomes that are meant to benefit humanity at large [41]. Structural inequities can not only be perpetuated through opaque algorithmic decision-making that may render asymmetrical power relations more obscure [1], but also through cooperation frameworks that insufficiently acknowledge the disparate manner in which the use of AI-systems affects global populations. This can concern the extraction of data from more vulnerable populations [10], the economic exploitation of ghost workers in weak labor markets [21] or the beta-testing of AI-applications in countries with less safeguards [41]. Countering these practices necessitates the participation of those countries and populations that have historically been underrepresented at the discussion table and are still exposed to the consequences thereof.

Finally, it was already mentioned that transparency on the parties' goals, intentions and expectations is a necessary precondition for durable cooperation. However, transparency is equally important during the cooperation process. Thus, the way of working, procedures for agreement, budgetary implications and conditions for welcoming new cooperation partners should be clearly set out. Procedural transparency is not just important within the initiative, but also externally. The legitimacy – and hence the success – of cooperation outcomes also hinges on their acceptance by those whom cooperation partners represent, be it citizens or other stakeholders. Legitimacy stems not only from a valid political mandate or board decision, but also from transparency about the cooperation process and accountability for the decisions taken in its context.

4.7 Establishing a feedback loop and preparing for the future

To achieve meaningful output, cooperation initiatives must secure adequate sources of input. Given AI's rapid scientific advances, a bridge must be created between cooperation partners and the research community, allowing for a constant information stream that ensures accurate, updated and scientifically sound input to form the basis of cooperation discussions. This research should not just focus on providing information that advances the cooperation initiative, but should also map the concrete impact of the initiative's outcomes. Ideally, a feedback loop is created, whereby the effects of the cooperation outcomes – for instance, a set of requirements for AI-systems – are monitored, and the extent to which they achieve their aim – for instance, a global level-playing field in terms of AI safety – assessed. Subsequently, these findings can serve as new input for the cooperation process and facilitate the evidence-based improvement of cooperation outcomes. This is particularly important in light of the uncertainty surrounding the most optimal governance models for AI at this stage of the technology's uptake.

In addition, the need for cooperation on AI must also be considered in the longer term. Although the urgency to cooperate forms a red thread throughout this chapter, this does not exclude the importance of looking ahead. Global cooperation on AI is unlikely to be a short-term undertaking, and is set to have continued importance for years to come. It is therefore necessary that – both at state and stakeholder level – investments are made in future cooperation resources and capabilities, particularly in the form of educating the cooperation teams of tomorrow. Citizens – whom the cooperation partners not only represent but of which they are also composed – need to be educated about AI's capabilities, limitations and impacts. An increased level of awareness and education on AI not only empowers individuals at home, at work and in the public sphere, but can also help ensure that, in the long run, they are sufficiently equipped to continue an informed and collaborative cooperation process on AI over the next decades. We are convinced that the establishment of a durable global cooperation framework on AI and its governance is a marathon, not a sprint.

5 Conclusions

Global cooperation knows many shapes. In the sections above, we did not seek to describe the numerous variations of cooperation initiatives that one can find in the context of AI today. Instead, we took a bird's eye perspective and aimed to set out an overarching framework in which such cooperation can take place, by setting out its core dimensions. Our framework was guided by three funda-

mental questions: *why* is there a need for global cooperation on AI and its governance; *what* areas should cooperation initiatives focus on; and *how* should cooperation be organized.

Many consider the global landscape of AI to be characterized by competition, despite the fact that – to attain the dual aim of maximizing AI’s benefits while preventing and minimizing its risks – global cooperation on AI and its governance is indispensable. We, however, believe that this aim is globally shared, and see hopeful indications that the need for cooperation thereon is increasingly acknowledged and addressed. To secure its realization, a balancing exercise between three elements must be mastered: speed, holism and contextualism.

First, the need for quick action cannot be overstressed. The technology’s fast development and deployment require an equally fast response, especially in areas where the likelihood and extent of risk may be significant. This implies that cooperation through existing cooperation initiatives should be favored over creating new ones, given the significant time and resources the latter typically requires; that differentiated cooperation structures should be considered so as to incentivize countries with aligned values and objectives to progress even if agreement cannot yet be reached ‘globally’; and that our learning curve about AI’s potential and limitations must become at least as steep as the curve of its rollout.

Second, a holistic approach is essential. This entails that requirements should be adopted to counter risks and enable a global level playing field at horizontal level, aiming to secure not only the trustworthiness of the environment of AI-systems, but also of data and digital infrastructure. These three socio-technical environments are interwoven, and it is only by considering them collectively that the ideal of Trustworthy AI can be approached. In addition, establishing a level playing field can foster healthy competition, thereby stimulating beneficial AI innovation and the wide dissemination thereof. At the same time, a holistic approach also requires a diversity of perspectives around the cooperation table, to secure that AI’s challenges and opportunities are considered from multiple angles and that humanity at large – with particular attention to those who are most vulnerable – is represented.

Third, the importance of context-specificity needs to be acknowledged. Not all aspects pertaining to AI require binding regulation or harmonization, nor do all aspects require global cooperation. Some countries may require a tailored approach to deal with AI’s impact based on their specific situation. In addition, a tailored approach is needed for certain domains in which AI can particularly help materialize benefits for humanity, areas where AI’s risks are particularly pronounced, or domains where the need for cooperation arises on an ad hoc basis. The manner in which cooperation can take shape in those domains will

largely depend on the underlying relationship of the willing cooperation partners and – in case of a competitive relationship – whether such competition takes place at the level of markets or at the level of value-systems.

AI is not a force of nature, but a tool designed by human beings. Consequently, whether, when, how and for what purposes AI is used, is an entirely human responsibility. All stakeholders across the globe – regardless of their nationality, ideology or value-system – carry this responsibility, and they carry it collectively. This raises citizens' legitimate expectation that, for all those areas where cooperation is needed to maximize AI's benefits for humanity and prevent and minimize its risks, actors across the world will roll up their sleeves. With this in mind, in this chapter we aimed to offer some guidance for the actors involved. Undoubtedly, we only managed to scratch the surface of the many complexities at stake. Nevertheless, we hope to have provided a framework that invites further reflection on how global cooperation on AI and its governance can be approached.

Acknowledgement

The authors are grateful for the helpful remarks received from Leo Kärkkäinen and Barry O'Sullivan on an earlier draft of this paper. They also acknowledge the useful comments suggested by Bertrand Braunschweig.

References

1. Benjamin, R.: *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, Medford, MA (2019).
2. Berendt, B.: AI for the Common Good?! Pitfalls, challenges, and ethics pen-testing. *Paladyn J. Behav. Robot.* 10, 1, 44–65 (2019). <https://doi.org/10.1515/pjbr-2019-0004>.
3. Brandenburger, A.M., Nalebuff, B.J.: *Co-Opetition*. Currency Doubleday, New York (1996).
4. Brkan, M.: Artificial Intelligence and Democracy: Delphi - Interdiscip. Rev. Emerg. Technol. 2, 2, 66–71 (2019). <https://doi.org/10.21552/delphi/2019/2/4>.
5. Brundage, M., et al.: *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. (2018).
6. Buiten, M.C.: Towards Intelligent Regulation of Artificial Intelligence. *Eur. J. Risk Regul.* 10, 1, 41–59 (2019). <https://doi.org/10.1017/err.2019.8>.
7. Carse, J.P.: *Finite and Infinite Games*. Ballantine Books, New York (1986).
8. Cate, F.H., Mayer-Schonberger, V.: Notice and consent in a world of Big Data. *Int. Data Priv. Law.* 3, 2, 67–73 (2013). <https://doi.org/10.1093/idpl/ipt005>.
9. Cath, C. et al.: Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. *Sci. Eng. Ethics.* (2017). <https://doi.org/10.1007/s11948-017-9901-7>.

10. Couldry, N., Mejias, U.A.: *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press, Stanford (2019).
11. Council of Europe: *Terms of reference for the Ad hoc Committee on Artificial Intelligence (CAHAI)*, (2019).
12. COWLS, J. et al.: *Designing AI for Social Good: Seven Essential Factors*. SSRN Electron. J. (2019). <https://doi.org/10.2139/ssrn.3388669>.
13. Dyson, K., Sepos, A.: *Which Europe?: The Politics of Differentiated Integration*. Palgrave Macmillan, Basingstoke (2010).
14. European Commission: *A European strategy for data*. European Commission, Brussels (2020).
15. European Commission: *Artificial Intelligence for Europe*. (2018).
16. European Commission: *Building Trust in Human-Centric Artificial Intelligence*. (2019).
17. Farer, T., Sisk, T.D.: *Enhancing International Cooperation: Between History and Necessity*. *Glob. Gov.* 16, 1, 1–12 (2010).
18. Feijóo, C. et al.: *Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy*. *Telecommun. Policy.* 101988 (2020). <https://doi.org/10.1016/j.telpol.2020.101988>.
19. Floridi, L. et al.: *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. *Minds Mach.* 28, 4, 689–707 (2018). <https://doi.org/10.1007/s11023-018-9482-5>.
20. Floridi, L.: *On Human Dignity as a Foundation for the Right to Privacy*. *Philos. Technol.* 29, 4, 307–312 (2016). <https://doi.org/10.1007/s13347-016-0220-8>.
21. Gray, M.L., Suri, S.: *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston (2019).
22. Harari, Y.N.: *Who Will Win the Race for AI?*, <https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/>, last accessed 2020/07/15.
23. Hasselbalch, G.: *A Framework for a Data Interest Analysis of Artificial Intelligence*. *Mediamocracy*. (2020).
24. Hasselbalch, G.: *Making sense of data ethics. The powers behind the data ethics debate in European policymaking*. *Internet Policy Rev.* 8, 2, (2019). <https://doi.org/10.14763/2019.2.1401>.
25. High-Level Expert Group on AI: *Ethics Guidelines for Trustworthy AI*, (2019).
26. Hummel, P. et al.: *Own Data? Ethical Reflections on Data Ownership*. *Philos. Technol.* (2020). <https://doi.org/10.1007/s13347-020-00404-9>.
27. IEEE: *IEEE Ethics In Action in Autonomous and Intelligent Systems*, <http://ethicsinaction.ieee.org/>, last accessed 2020/07/16.
28. ITU: *Assessing the Economic Impact of Artificial Intelligence*. ITU, Geneva (2018).

29. Jobin, A. et al.: The global landscape of AI ethics guidelines. *Nat Mach Intell.* 1, 389–399 (2019).
30. Kalluri, P.: Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature.* 583, 7815, 169–169 (2020). <https://doi.org/10.1038/d41586-020-02003-2>.
31. Kleijssen, J.: Technologies numériques, intelligence artificielle et lutte contre la COVID-19 : la contribution du Conseil de l'Europe. *Obs. Brux.* 121, (2020).
32. Kranzberg, M.: Technology and History: "Kranzberg's Laws." *Bull. Sci. Technol. Soc.* 15, 1, 5–13 (1995). <https://doi.org/10.1177/027046769501500104>.
33. Kroll, D.A., Leuffen, D.: Enhanced cooperation in practice. An analysis of differentiated integration in EU secondary law. *J. Eur. Public Policy.* 22, 3, 353–373 (2015). <https://doi.org/10.1080/13501763.2014.956781>.
34. LeCun, Y. et al.: Deep learning. *Nature.* 521, 7553, 436–444 (2015). <https://doi.org/10.1038/nature14539>.
35. Lee, K.-F.: *AI Superpowers: China, Silicon Valley, and the New World Order.* Houghton Mifflin Harcourt, Boston (2018).
36. Lehtiniemi, T., Haapoja, J.: Data agency at stake: MyData activism and alternative frames of equal participation. *New Media Soc.* 22, 1, 87–104 (2020). <https://doi.org/10.1177/1461444819861955>.
37. Luengo-Oroz, M. et al.: Artificial intelligence cooperation to support the global response to COVID-19. *Nat. Mach. Intell.* 2, 6, 295–297 (2020). <https://doi.org/10.1038/s42256-020-0184-3>.
38. Lynskey, O.: *The foundations of EU data protection law.* Oxford University Press, Oxford (2015).
39. Mayer-Schönberger, V., Cukier, K.: *Big Data: A Revolution that Will Transform how We Live, Work, and Think.* Houghton Mifflin Harcourt (2013).
40. McKinsey: *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy.* McKinsey (2018).
41. Mohamed, S. et al.: Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philos. Technol.* (2020). <https://doi.org/10.1007/s13347-020-00405-8>.
42. Morgan, F.E. et al.: *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World.* 224.
43. Muller, C.: *The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law.* Council of Europe, Strasbourg (2020).
44. Nemitz, P.: Constitutional democracy and technology in the age of artificial intelligence. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 376, 2133, 20180089 (2018). <https://doi.org/10.1098/rsta.2018.0089>.
45. Nilsson, N.J.: *The Quest for Artificial Intelligence: A History of Ideas and Achievements.* Cambridge University Press, Cambridge (2009). <https://doi.org/10.1017/CBO9780511819346>.
46. OECD: *Recommendation of the Council on Artificial Intelligence.* (2019).

47. O’Neil, C.: *Weapons of Math Destruction*. Penguin Books Ltd (2017).
48. Partnership on AI: Joint Statement from founding members of the Global Partnership on Artificial Intelligence, <https://www.gov.uk/government/publications/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence>, last accessed 2020/07/16.
49. Russell, S.: *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking (2019).
50. Smuha, N.A.: Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea. *Philos. Technol.* (2020). <https://doi.org/10.1007/s13347-020-00403-w>.
51. Smuha, N.A.: From a “Race to AI” to a “Race to AI Regulation” - Regulatory Competition for Artificial Intelligence. *SSRN Electron. J.* (2019). <https://doi.org/10.2139/ssrn.3501410>.
52. Smuha, N.A.: The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Comput. Law Rev. Int.* 20, 4, 97–106 (2019).
53. Taddeo, M.: Data philanthropy and the design of the infraethics for information societies. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 374, 2083, 20160113 (2016). <https://doi.org/10.1098/rsta.2016.0113>.
54. Taddeo, M.: Just Information Warfare. *Topoi.* 35, 1, 213–224 (2016). <https://doi.org/10.1007/s11245-014-9245-8>.
55. Taddeo, M.: Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity. *Minds Mach.* 29, 2, 187–191 (2019). <https://doi.org/10.1007/s11023-019-09504-8>.
56. UNESCO Ad Hoc Expert Group (AHEG) for the Preparation of a Draft text of a Recommendation the Ethics of Artificial Intelligence: First version of a draft text of a recommendation on the Ethics of Artificial Intelligence. (2020).
57. United Nations ed: *The 2030 Agenda for Sustainable Development*. Presented at the (2015).
58. United Nations Secretary-General: Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation, <https://undocs.org/A/74/821>, (2020).
59. Van Alsenoy, B. et al.: Privacy notices versus informational self-determination: Minding the gap. *Int. Rev. Law Comput. Technol.* 28, 2, 185–203 (2014). <https://doi.org/10.1080/13600869.2013.812594>.
60. Veale, M.: A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence. *Eur. J. Risk Regul.* 1–10 (2020). <https://doi.org/10.1017/err.2019.65>.
61. Veale, M.: Privacy is not the problem with the Apple-Google contact-tracing app, <http://www.theguardian.com/commentisfree/2020/jul/01/apple-google-contact-tracing-app-tech-giant-digital-rights>, last accessed 2020/07/01.

62. Vinuesa, R. et al.: The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* 11, 1, 233 (2020). <https://doi.org/10.1038/s41467-019-14108-y>.
63. Yeung, K. et al.: AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing. *SSRN Electron. J.* (2019). <https://doi.org/10.2139/ssrn.3435011>.
64. Yeung, K.: Responsibility and AI - A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework, (2019).
65. CEN-CENELEC Focus Group on Artificial Intelligence, <https://www.cen-cenelec.eu/news/articles/Pages/AR-2019-001.aspx>, last accessed 2020/09/13.
66. Digital Public Goods Alliance, <https://digitalpublicgoods.net/>, last accessed 2020/07/15.
67. ETSI Specification group on Securing Artificial Intelligence, <https://www.etsi.org/committee/1640-sai>, last accessed 2020/09/13.
68. Executive Order on Maintaining American Leadership in Artificial Intelligence, <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>, last accessed 2020/07/16.
69. Global Data Access Framework, <https://www.unglobalpulse.org/policy/global-data-access-framework/>, last accessed 2020/07/15.
70. ISO/IEC JTC 1/SC 42 - Artificial intelligence, <https://www.iso.org/committee/6794475.html>, last accessed 2020/09/13.
71. MyData, <https://mydata.org/>, last accessed 2020/07/16.
72. NIST, <https://www.nist.gov/topics/artificial-intelligence>, last accessed 2020/09/13.
73. The OECD Artificial Intelligence Policy Observatory - OECD.AI, <https://oecd.ai/>, last accessed 2020/07/16.
74. The Partnership on AI, <https://www.partnershiponai.org/>, last accessed 2020/07/16.