

Article

An International Data-Based Systems Agency IDA: Striving for a Peaceful, Sustainable, and Human Rights-Based Future

Peter G. Kirchsclaeger^{1,2,3}¹ Institute of Social Ethics ISE, University of Lucerne, 6002 Lucerne, Switzerland; peter.kirchsclaeger@unilu.ch² Chair for Neuroinformatics and Neural Systems, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland³ ETH AI Center, 8092 Zurich, Switzerland

Abstract: Digital transformation and “artificial intelligence (AI)” —which can more adequately be called “data-based systems (DS)” —comprise ethical opportunities and risks. Therefore, it is necessary to identify precisely ethical opportunities and risks in order to be able to benefit sustainably from the opportunities and to master the risks. The UN General Assembly has recently adopted a resolution aiming for ‘safe, secure and trustworthy artificial intelligence systems’. It is now urgent to implement and build on the UN General Assembly Resolution. Allowing humans and the planet to flourish sustainably in peace and guaranteeing globally that human dignity is respected not only offline but also online, in the digital sphere, and in the domain of DS, requires two policy measures: (1) human rights-based data-based systems (HRBDS): HRBDS means that human rights serve as the basis of digital transformation and DS. (2) International Data-Based Systems Agency (IDA): IDA should be established at the UN as a platform for cooperation in the field of digital transformation and DS, fostering human rights, security, and peaceful uses of DS, as well as a global supervisory institution and regulatory authority in digital transformation and DS. The establishment of IDA is realistic because humanity has already shown in its past that we are able to not always “blindly” pursue the technically possible but also to limit ourselves to what is technically feasible when humanity and the planet are at stake. For instance, humans researched the field of nuclear technology, developed the atomic bomb, and detonated it several times. Nonetheless, the same humans limited research and development in the field of nuclear technology to prevent even worse consequences by establishing the International Atomic Energy Agency (IAEA) at the UN.

Keywords: artificial intelligence (AI); data-based systems (DS); peace; human rights; international data-based systems agency (IDA)



Citation: Kirchsclaeger, P.G. An International Data-Based Systems Agency IDA: Striving for a Peaceful, Sustainable, and Human Rights-Based Future. *Philosophies* **2024**, *9*, 73. <https://doi.org/10.3390/philosophies9030073>

Academic Editor: Jordi Vallverdú

Received: 14 April 2024

Revised: 8 May 2024

Accepted: 9 May 2024

Published: 20 May 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Digital transformation and so-called “Artificial intelligence (AI)” present humanity and the planet with enormous ethical opportunities, which mean something ethically positive, ethically right, or ethically good, and ethical risks, which mean something ethically negative, ethically wrong, or ethically bad. The UN General Assembly has recently adopted a resolution aiming for ‘safe, secure and trustworthy artificial intelligence systems’ [1]. It is now urgent to implement and build on the UN General Assembly resolution. This article aims to identify the ethical upsides and downsides to develop a concrete solution to make humanity and the planet benefit from ethical opportunities and to avoid or master ethical risks.

This article argues against indifference as it does not accept the idea that humans cannot do anything in the field of so-called “AI” but just consume so-called “AI” and produce so-called “AI” by generating and supplying their own data, as well as accepting the violations of their human rights and the destruction of peace and of the planet. In other words, contrary to an attitude to give up some human rights for some technology-based benefits and comfort, it is appropriate to remind the reader how humans have for centuries

fought for human rights, for a global order fostering peace, and how they have in the past striven to protect the planet rather than surrendering everything to the special interests of a few multi-national corporations [2]. The article puts stress on an ethical analysis applying a methodology of research in theoretical ethics that demonstrates that it is the primary responsibility of humans to steer the design, production, deployment, or ethically motivated non-deployment of technologies instead of letting the design, production, and deployment of technologies just happen.

It starts with a discussion of the conceptual problems of the term “AI”—leading to a more accurate description as data-based systems (DS) and providing an understanding of what DS are and what they are not—and of the ethical opportunities and ethical risks of DS.

Based on this analysis, the question is addressed: how can the ethical opportunities and ethical risks of DS be governed by an examination of already existing global governance initiatives?

Informed by this investigation showing that the regulatory oversight in this domain remains insufficient so far, the pressing need to establish robust governance mechanisms to ensure the responsible and sustainable development and deployment of DS is addressed by the introduction and the discussion of two concrete measures—human rights-based DS and the International Data-Based Systems Agency (IDA). This article attempts to offer concrete suggestions on how humans could live up to their primary responsibility, including a normative handling of the entire life cycle of DS based on human rights as well as a legal outline of a global and institutional approach because DS are a global phenomenon of fundamental and existential significance.

2. Data-Based Systems (DS) Rather Than “Artificial Intelligence”

Confronted with the question of the definition of “artificial intelligence”, one becomes aware of its conceptual blurriness [3,4] which should be overcome from an ethical perspective [5]. Artificial intelligence can be defined as “machines that are able to ‘think’ in a human-like manner and possess higher intellectual abilities and professional skills, including the capability of correcting themselves from their own mistakes” [6,7]. The term “artificial” in “artificial intelligence” highlights that “intelligence (is) displayed or simulated by technological means” [8].

From an ethical standpoint, the above-mentioned starting point is criticized because intelligence does not just consist in the solution of a cognitive task but also in the way it is pursued [9]. In view of the nature of artificial intelligence, doubts arise from an ethical perspective if the term is even adequate, because artificial intelligence strives to imitate human intelligence, but this is limited to a certain area of intelligence (e.g., certain cognitive capacities) [10,11]. Furthermore, it is to be assumed that artificial intelligence can at best become like human intelligence in certain areas of intelligence but can never become the same. Among others, in the domain of emotional and social intelligence, machines are only able to simulate emotions, personal interaction, and relationships and lack authenticity. For instance, a health care robot can be trained to cry when the patient is crying, but no one would argue that the robot feels real emotions and cries due to them. On the contrary, one could train the exact same robot to slap the patient’s face when the patient is crying, and the robot would perform this function in the same perfect way. The lack of authenticity of robots in health care is problematic for respecting the dignity of all humans [12]. As it is relevant to the respect of human dignity, authenticity must be part of the equation in data-based health care and in the use of “care robots” [13].

Beyond that, in the domain of moral capability, one cannot ascribe machines with moral capability because they are presupposed to follow patterns and rules given by humans. Technologies are primarily made for their suitability and may set rules as a self-learning system, for example, to increase their efficiency, but these rules do not contain any ethical qualities. E.g., a self-driving car could set the rules for itself, but it is not aware of the ethical quality. It could give itself the rule to get from A to B as fast as

possible including harming humans and nature, to optimally fulfill the task of reaching B in the shortest time possible, without being able to recognize ethical rules for itself, which would allow the machine to perceive the illegitimacy of its rules and actions. A human driver instead possesses the potential to recognize for himself or herself binding ethical rules, which empower him or her to see that harming humans and nature might be more efficient but illegitimate. Machines lack this autonomy. Autonomy encompasses recognizing and setting ethical norms for oneself and basing one's own actions on them. Humans can set the rules for a self-driving car, whether good or bad. Machines fail on the principle of generalizability. This principle has its roots in Immanuel Kant's universalization principle that ethical rules can only be ethical rules if we want them to be universal law [14]. Based on this, the fulfillment of the principle of generalizability presupposes presenting rational and plausible arguments—"good reasons". "Good reasons" means that it must be conceivable that all humans, given their effective freedom and autonomy as well as their full equality, would agree upon these reasons—within a model of thought and not within a real worldwide referendum—on ethical grounds [15]. While a human can know that he or she does something ethically right or wrong, machines cannot identify the ethical quality.

In addition, the potential that technologies possess in relation to ethical decisions and actions is nowhere close to moral capability because machines lack not only autonomy but also vulnerability, conscience, freedom, and responsibility, which are essential for human morality [16].

Finally, sometimes ethics must go beyond principles, norms, and rules to be sensitive to the rule-transcending uniqueness of the concrete [15]. This accounts for the truth that in a concrete encounter with concrete people in a concrete situation, rules can reach their limit because the concrete, in its uniqueness outranks the rule. "The general, concrete ethical, the positive legal and many other norms that are generally applicable, although indispensable, are not sufficient to guarantee the basic humanity which, in the face of diversity (. . .). It is inevitable that we have to cross norms in certain situations in order to act humanely, but this does not mean that we deny the need for norms in general or refute that they are generally applicable" [17] (pp. 42–43). Through the increasing complexity of everyday reality—e.g., when guiding principles diverge or collide—humans are challenged to find ethical insights into the ethical assessment of a concrete encounter with concrete persons in a concrete situation. These ethical considerations in a more differentiated and better manner would be expecting too much of data-based systems due to their lack of moral capability. Transferring ethics completely to mathematics, programming, or training becomes difficult or even impossible.

Therefore, technologies cannot perform as moral subjects or moral agents, but humans carry the ethical responsibility of machines. Humans must lay down ethical principles and ethical and legal norms; set a framework, goals, and limits for digital transformation; and define the use of machines in addition to examining, analyzing, evaluating, and assessing technology-based innovation from an ethical perspective.

The term "data-based systems" would be more appropriate than "artificial intelligence" because this term describes what actually constitutes "artificial intelligence": generation, collection, and evaluation of data; data-based perception (sensory, linguistic); data-based predictions; data-based decisions. In addition, the term "data-based systems" allows for highlighting the main strengths and weaknesses of the present technological achievements in this field. The mastery of an enormous quantity of data represents the key asset of data-based systems.

Pointing to its core characteristic, namely being based on data and relying exclusively on data in all its processes, its own development, and its actions—more precisely, its reactions to data—lifts the veil of the inappropriate attribution of the myth of "intelligence", covering substantial ethical problems and challenges of data-based systems. This allows more accuracy, adequacy, and precision in the critical reflection on data-based systems. For instance, the untraceability, unpredictability, and inexplicability of the algorithmic processes resulting in data-based evaluation, data-based predictions and data-based de-

cisions (“black-box-problem”) [18–21], its wide vulnerability to systemic errors, its deep exposure for confusing causality with correlation (e.g., high consumption of ice-creams by children in a summer-month and high number of children car-accidents due to more mobility during vacation in the same summer-month correlate but there is not any causal relationship between the two statistics, meaning ice-cream-consumption does not cause car-accidents) [22], and its high probability of biased and discriminatory data leading to biased and discriminatory data-based evaluations, predictions, and decisions embrace its major disadvantages [8,23]. “Algorithms are opinions embedded in codes. They are not objective” [24]. They are not neutral. They serve specific goals and purposes.

Finally, this terminological sharpening does not exclude the possibility of relying on and learning from the existing research and discourse on so-called “AI” (including, e.g., “knowledge-based systems”) and its technological and normative dimensions.

3. Ethical Opportunities and Risks of DS

“Data-based systems (DS)” comprise ethical opportunities and ethical risks. DS can be powerful, e.g., for fostering human dignity and sustainability, but also for violating human dignity or destroying the planet. Elon Musk warns: “AI is far more dangerous than nukes [nuclear warheads]. Far. So why do we have no regulatory oversight? This is insane” [25]. Stephen Hawking points out: “Unless we learn how to prepare for, and avoid, the potential risks, AI could be the worst event in the history of our civilization. It brings dangers, like powerful autonomous weapons, or new ways for the few to oppress the many. It could bring great disruption to our economy” [26]. Therefore, it is necessary to identify ethical opportunities and ethical risks precisely and at an early stage in order to be able to benefit sustainably from the ethical upsides of DS and to master or avoid the ethical downsides of DS. In the avoidance and mastering of the downsides, technology-based innovation can, in turn, play an essential role.

Humans need to become active so that digital transformation and DS do not simply happen, but that humans shape it. This is necessary so that digital transformation and DS will not be reduced to an instrument serving pure efficiency but can rise to their ethical potential. More importantly, there is a need for normative guidance to review the economic self-interests that run digital transformation and DS so far almost exclusively and to guide calls for international regulations and governance in the digital domain and in the sphere of DS.

4. Existing Global Governance Initiatives

Several declarations, recommendations, principles, and guidelines have contributed to a debate about the international governance of DS—the first generation of governance initiatives: “the sermons”. Different initiatives by states and civil society on national [27–43], regional [44–55]—e.g., “The Charter of Digital Rights” by European Digital Rights (EDRi)-network [56],—and international level [30,38,39,41,57–70]—e.g., the “Bletchley Declaration 2023” [71], the “AI4People Summit Declaration”, the “G7 International Guiding Principles on AI”, the “AI Code of Conduct”, the “Recommendation on the Ethics of Artificial Intelligence” by UNESCO, the OECD Principles on Artificial Intelligence [72], the G20-Principles of AI, “Declaration of Principles” by the World Summit on the Information Society in Geneva 2003 [73]. Other examples—emerging from professional ethics—are the code of the Institute of Electrical and Electronic Engineers IEEE [74], the code of the National Society of Professional Engineers [75], or the codes of the American Society of Mechanical Engineers [76]. The main challenge with these declarations, recommendations, principles, and guidelines is the step from theory to practice. They remain at maximum “soft law” “which is a tool often used to either avoid or anticipate formal legislation” [77] (p. 26).

Beyond that, the European Parliament and Council reached political agreement on the European Union’s Artificial Intelligence Act (‘EU AI Act’) [78]. The EU AI Act aims to represent a comprehensive legal framework for the regulation of AI systems across the

EU, ensuring the safety of and respect of fundamental rights by DS as well as encouraging investment and innovation in the field of DS.

Other legal initiatives—forming together with the EU AI Act the second generation of governance initiatives: “the locals”—are also pursued in China and in the USA on the federal level, and several governments on the state level released new regulations, leading to the categorization of these activities as “American Market-Driven Regulatory Model”, “Chinese State-Driven Regulatory Model”, and “European Rights-Driven Regulatory Model” [79] (pp. 35–145).

Finally, the debate about global governance in DS also knows the third generation of governance initiatives: “the international players”—consisting of the proposals of two main models: the model of the Intergovernmental Panel on Climate Change (IPCC) [80] and the model of the International Civil Aviation Organization (ICAO) [81,82].

The model of the Intergovernmental Panel on Climate Change (IPCC) consists of a panel of experts. The IPCC was established in 1988 by the United Nations, with member countries from around the world. It provides governments with scientific information they can use to develop climate policies. The panel of experts in the domain of AI would provide policymakers and governments with information, scenarios, and models for their decision-making.

The model of the International Civil Aviation Organization (ICAO) [81] consists of a binding global framework and its implementation. The ICAO, as a United Nations agency, has its basis in the Convention on International Civil Aviation. The ICAO is the global forum of states for international civil aviation. It develops policies and standards, provides compliance audits, studies, analyses, and assistance to states and stakeholders, and contributes to the global alignment of air regulations.

While “the sermons” lack implementation during the working week, and “the locals” possess a national or regional focus, while DS represents a global phenomenon, “the international players” locate the governance of DS on an international level and approach it adequately in an institutional manner in order to create a positive impact in reality and on the ground.

At the same time, it needs to be considered that the IPCC model will not reach the desired effect on the ground because it does not possess neither legal authority nor legal enforcement tools. This weakness of the IPCC model becomes clear listening to the UN High Commissioner for Human Rights, Volker Türk: “Victims and experts (...) have raised the alarm bell for quite some time, but policy makers and developers of AI have not acted enough—or fast enough—on those concerns. We need urgent action by governments and by companies. And at the international level, the United Nations can play a central role in convening key stakeholders and advising on progress. There is absolutely no time to waste. The world waited too long on climate change. We cannot afford to repeat that same mistake” [83].

Regarding the ICAO model, doubts arise if one can compare a self-contained industry like aviation to a cross-cutting technology like DS. The latter causes manifold and multifaceted legal and ethical issues [81].

At the same time, humanity and the planet are struggling with the enormous ethical and legal problems digital transformation and the use of DS pose. Among others, a first global threat represents the growing global inequalities and poverty because of a dramatically widening “digital divide” consisting of and leading to violations of human rights.

The negative impact of DS on the climate and the environment will increase—a second global threat comprising violations of human rights.

A third global threat is the constant violation of human rights to privacy and data protection. Whenever possible, data is stolen from humans and sold to the highest bidder. The continuous disrespect for privacy and data protection forms a massive attack on the freedom of all humans.

Thanks to this vast amount of data about humans, DS know humans better than humans know themselves. This opens the door to economic and political manipulation as

well as disinformation—a fourth global threat. Several democratic decisions have already been manipulated so far with the help of DS. Take the 2016 presidential election in the USA: It has been proven that “Facebook” (now: Meta) sold on user data records. The same thing happened with Brexit [84]. That means that dictators and totalitarian regimes can influence elections and votes in democracies. Manipulation and disinformation lead to the destabilization of democratic countries. The rapidly developing technical possibilities for disinformation and manipulation of people through large language models such as “Chat GPT” open new horizons in this regard. It can be assumed that, for example, “Chat GPT” will further intensify the phenomenon of fake news, which is already so devastating for democracies, with “deep fakes”. At the same time, quality journalism as a pillar of democracy will come under even greater economic and political pressure. This is because media channels can be filled with texts from “Chat GPT” at a low cost. Moreover, economic manipulation affects humans as consumers. DS know exactly—to use a metaphor—which piano keys it must hit to make the music play, in other words, to make humans shop the way it wants humans to.

A fifth global threat represents the security risks for the mental health of children and young people, representing violations of human rights due to the impact of social media as well as for physical health and for the lives of all of us because of the existential consequences of DS-based cyber-attacks and of military applications of DS for global peace and security.

Facing these five global threats, the limitations of the so-far three generations of governance initiatives become obvious if one approaches this risky reality with the following test questions for the already existing “sermons”, “locals”, and “international actors”:

1. Do multinational technology companies need to change anything regarding their human rights-violating business practices because of these proposals?
2. Can human rights-violating state actions or business practices be concretely stopped because of these proposals?
3. Can states or multinational technology companies be held accountable for their human violations based on these proposals?

The already existing “sermons”, “locals”, and “international actors” receive negative responses in all three cases. Therefore, it is necessary to think about ways to combine the strengths of “the sermons”, “the locals”, and “the international actors” and to avoid their weaknesses in order to address these global threats in an impactful and global way while still benefiting from the ethical opportunities of DS. Allowing humans and the planet to flourish sustainably and guaranteeing globally that human dignity is respected not only offline but also online and in the digital sphere as well as in the domain of DS, the below described concrete measures are proposed.

5. Human Rights-Based Data-Based Systems HRBDS

In order to pass the above-mentioned test questions, in order to take the current situation of humanity coined by the above-mentioned five global threats more seriously, and in order to address the ethical opportunities and ethical risks of DS more fervently, human rights as an ethical frame of reference could provide as a minimum requirement the necessary normative guidance. Human rights offer the major benefit of being based on a simple concept and focusing on the essentials: Besides the ethical justifiability of human rights and their universality [85], they define the minimum standards guaranteeing that all humans—always, everywhere—can physically survive and lead a life with dignity—a life worth living. They also encourage and foster innovation by protecting people’s freedom to think, express their opinion, and access information, as well as promote pluralism by respecting each person’s right to self-determination.

Compared with a risk-based approach, human rights-based DS have the advantage that one can identify the respect or violation of rights precisely a priori and a posteriori and that the validity of rights impacts immediately, while a risk-based approach leaves a lot of room for interpretation, causes a time shift, and encompasses the danger of dilution

of the legal protection. The last point gains even more weight if the risk assessment (in other words, if there is a risk and how significant the risk is) lies in the hands of the same companies providing the respective DS because of the obvious conflicts of interest and the explicit conflicting objectives.

Based on these considerations, we should strive for human rights-based design, development, production, use of data-based systems, and non-use of data-based systems based on human rights concerns—we need human rights-based data-based systems HRBDS [16,86] includes a precautionary approach, the reinforcement of existing human rights instruments specifically for data-based systems, and the promotion of algorithms supporting and furthering the realization of human rights. Of course, HRBDS understands human rights as universal—all humans are holders of human rights everywhere and always [87], inalienable, and indivisible. The principle of inalienability means that one cannot lose, give away, or sell his or her human rights. The principle of indivisibility defines that all human rights must go hand in hand. This means that the entire catalog of human rights needs to be respected. Therefore, every human right must be implemented optimally and in a way that accords with all other human rights being implemented optimally at the same time.

HRBDS means—in other words—that human rights are respected, protected, implemented, and realized within the entire life cycle of DS and the complete value-chain process of DS.

The value and significance of HRBDS are also emphasized by the fact that the EU, with its AI Act [78], the Council of Europe, with its work on a Convention on the development, design, and application of artificial intelligence [88], and various UN bodies [1,89] employ human rights as a basis for regulating DS.

At the same time, the so far existing and already aspired legislation does not go far enough because it does not yet implement human rights online, in the digital sphere, and in the domain of DS as it does offline; because it applies—if we look at the EU AI Act—a “risk-based” rather than a “rights-based” approach, causing more space for interpretation and arbitrariness—in front of all, if the private sector itself can assess the risks and the level of risk of their products.

In its strive for protecting the powerless from the powerful, the HRBDS goes further by including the respect and realization of human rights in the entire value chain and the entire life cycle of DS (in other words, in the design, the development, the production, the distribution, the use, or the non-use of DS because of human rights concerns).

To illustrate it with a concrete example, HRBDS means that, e.g., the human rights to privacy and data protection in their relevance for the human dignity and freedom of humans must be defended for all humans, excluding the possibility that only a group of humans are respected as holders of human rights and that humans should be able to sell themselves and their data as well as their privacy as products. This is a substantial argument against data ownership as well. Or would or should one come up with the idea to sell her or his love letter to her or his love to the state and to corporations as data? Or would or should one sell the dinner-table conversation of her or his family to the state or the private sector? Or would or should one sell the behavioral habits of her or his children to the state or a company? No. And not even the offer to sell human rights or a specific human right should be legally made to humans because of the principle of the inalienability of human rights [90]. Even if the question arises in the application of the HRBDS if a specific human right should be prioritized before another specific human right, the HRBDS is able to provide ethical guidance based on the principle of indivisibility of human rights [90].

HRBDS can be illustrated as well by the call for an economically successful, legal, and legitimate business model, e.g., for video-conference software. Current business models of video-conference software, e.g., ZOOM [91], surveille the users and violate their human right to privacy and data-protection by collecting and generating their data and selling it to third parties, although our privacy and our data should not be for sale based on the principle of inalienability. In other words, it must be possible to create a profitable business

model with the provision and promotion of videoconferencing software that does not imply human rights violations by not collecting and not generating data from their users or selling it to third parties.

Another visual example could be automated driving. In order not to overload automated driving with maximum demands and a high ethos, in order to concretize the ethical requirements for automated driving and make them tangible, and in order to succeed in weighting them against other important goods such as mobility and comfort, the approach of human rights-based automated driving (HRBAD) would be worth striving for. Human rights, as a minimum standard that guarantees people to survive and live with human dignity, are achievable for automated driving and allow a focus on what is essential and important—what is necessary to survive and live. Human rights possess a precise focus that can promote clear prioritization based on this minimum standard to be met first. In the agenda-setting process of automated driving, human rights can therefore help not only to set the right priorities but also to adequately define the spheres of influence and responsibility.

The concept of HRBAD also makes it possible to set ethical reference points in relation to other goods (mobility, comfort), thus enabling a conceptual classification. For example, an aspect of automated driving that involves a violation of a human right cannot be outweighed by more comfort. On the other hand, a human rights-neutral luxury solution in the area of the comfort of automated driving that is only made available to a small part of the population through appropriate pricing cannot be described as “unjust” in the sense of distributive justice. Rather, luxury goods can be negotiated with reference to transactional justice. It would be different if a human rights-relevant element of automated driving (e.g., safety) were involved. Here, such exclusion via the high price would not be legitimate.

6. International Data-Based Systems Agency IDA

6.1. *The Purpose of IDA*

Beyond HRBDS as a regulatory framework and considering the inherently dual nature of data-based systems (DS) from an ethical perspective and its substantial impact on humanity and the planet, the aim should be to implement the regulatory framework guaranteeing the use of the ethical positive potential of DS for the benefit of all humans and the planet as well as the handling of its ethical negative potential, including the destruction of humankind and the planet. Serving this aim is the establishment of robust governance mechanisms to ensure the development and deployment of HRBDS. Therefore, an International Data-Based Systems Agency (IDA)—analogous to the International Atomic Energy Agency (IAEA) (www.iaea.org)—needs to be established at the UN. It will be a platform for technical cooperation in the field of digital transformation and DS for state and non-state actors (including, of course, the private sector and civil society, as well as organizations and institutions [already] active in this field, among others, ITU, the Broadband Commission, UNESCO, the UN Department of Economic and Social Affairs, UN DESA), fostering human rights, safety, security, and peaceful uses of DS as well as a global supervisory and monitoring institution and regulatory authority in the area of digital transformation and DS, responsible for access to market approval. Integrated in or associated with the UN, it should work for the safe, secure, and peaceful uses of data-based systems, contributing to international peace and security, the respect and realization of human rights, and the United Nations’ Sustainable Development Goals. Its global and inclusive approach will permit it to master the risk of fragmentation in the field.

IDA needs to be built following the model of the International Atomic Energy Agency (IAEA) as an “institution with teeth” because, thanks to its legal powers, functions, enforcement mechanisms, and instruments, the IAEA was able to foster innovation and ethical opportunities while at the same time protecting humanity and the planet from the existential risks in the domain of nuclear technologies, which also embrace the same dual nature as DS, covering both ethical upsides and downsides. Lessons from the IAEA experience underscore the importance of prioritizing human rights, transparency, and accountability

in DS governance frameworks. By combining a human rights-based approach to DS development and deployment with establishing IDA as a central regulatory authority, the international community can proactively address existential AI risks while harnessing the transformative potential of DS to advance human rights and sustainability. Leveraging the lessons learned from nuclear technologies and the establishment of the IAEA, the establishment of IDA presents a viable pathway towards effective global governance of existential AI risks, ensuring the responsible and ethical development of DS for the betterment of humanity and the planet.

6.2. The 30 IDA Principles

Based on human rights and on the United Nations' Sustainable Development Goals, as well as in critical dialogue with already existing soft-law instruments and declarations [27,30,38,41,42,67,92], the following 30 principles should provide ethical guidance to IDA. The IDA principles consist of more principles in order to address the most recent ethical challenges as well as lacunae, which one could argue for in the so-far existing soft-law instruments and declarations. For example, it is not enough to call for transparency because if someone acts illegitimately, the unethical quality of this action is not transformed into something legitimate just due to transparency. Similarly, it is not enough that DS are explainable DS must also be intelligible in order to enable humans to benefit from the ethical opportunities and to master the ethical risks of DS.

The IDA should serve the realization of the following 30 IDA principles:

1st principle: Data-based systems must respect, protect, implement, and serve the realization of human rights.

This principle ensures that the respect, protection, implementation, and realization of human rights are guaranteed not only "offline" but also "online", in the digital sphere, and in the domain of DS. Especially this first principle—together with the following 29 IDA principles—addresses the above-outlined five global threats because they ensure that human rights are respected, protected, implemented, and realized throughout the entire life cycle of DS.

2nd principle: Data-based systems must serve the realization of the United Nations' Sustainable Development Goals.

Beyond their contribution to the realization of human rights, this principle guarantees that DS foster the realization of the UN SDGs.

3rd principle: Data-based systems must be transparent.

This principle ensures the transparency of DS, which means that the inner workings of DS are accessible. Humans should not blindly trust DS because they are neither objective, fair, nor neutral, but their algorithms and the data they are relying on are neither objective, fair, nor neutral, but biased.

4th principle: Data-based systems must be traceable.

Combined with the 3rd principle, this principle guarantees that the steps of the inner-working process of DS are not only accessible, but their traces are also identifiable, allowing a precise and specific account of the single steps within a DS.

Both—the 3rd principle and the 4th principle—build the foundation of the possibility for humans to carry their responsibility for DS. This is necessary because DS do not possess moral capability due to a lack of the principles of vulnerability, conscience, freedom, and autonomy. Thus, they cannot be responsible for their decisions and actions.

5th principle: Data-based systems must be explainable.

This principle ensures the possibility of an explanation of the inner-working process of DS which allows humans to learn about it.

6th principle: Data-based systems must be intelligible.

Combined with the 5th principle, this principle guarantees that humans not only can know about the specific steps of the inner-working process of DS, but that humans are also able to understand and act upon it.

The 5th principle and the 6th principle represent further pillars of the fundament of the possibility for humans to carry their responsibility for DS because humans cannot perform their responsibility for DS if explanations of the inner-working process of DS as well as intellectual access to the inner-working process of DS would not be provided to them.

7th principle: Data-based systems must be auditable.

Combined with the 3rd principle and the 4th principle, this principle—based on transparency and traceability of DS—opens the horizon for auditing DS, allowing for the identification of responsibilities and liabilities. This is the only way to empower humans to restore justice and to enable humans to address possible accidents or mistakes by DS in an adequate manner, which increases, among others, the probability that the same accidents and mistakes will not happen again.

8th principle: Causes and effects, or causality and correlation, must be identifiable in data-based systems.

This principle makes sure that confusions of causes and effects as well as of causality and correlation in DS—both scenarios potentially deleting essential differentiations in cognitive perception and both scenarios potentially resulting in blurriness of responsibility and leading to conceptual and epistemological blindness as well as terrible consequences—can be avoided by identifying them precisely in DS.

9th principle: Data-based systems must be predictable.

Combined with the 8th principle, this principle empowers humans in their decision-making to trust in the assistance, support, and work of DS or not. Only this way can humans live up to their responsibility for DS because they understand what will happen if they use DS as well as what will happen if they do not deploy DS.

This leads to the specific ethical consequence that self-learning DS must be designed and built in a way that full predictability is given. This is not surprising because the usual necessity of humans is that they use machines only if they know what they can expect is going to happen. Or would one, e.g., board an aircraft if one did not know that the aircraft was normally able to fly?

10th principle: Data-based systems must be decidable.

This principle guarantees that DS can take decisions and that the opinion-forming and decision-making processes resulting in decisions can be accessed and followed by humans.

11th principle: Data-based systems must be non-manipulating and respect the autonomy of every human.

This principle explicitly addresses the ethical danger of manipulation of humans by DS as well as disrespect of the autonomy of humans by DS. DS can spread disinformation, “fake news” and “deep fakes” undermining the freedom and autonomy of humans as well as the credibility of human opinion-forming and decision-making processes. Based on the enormous amount of data about humans, DS “know” humans better than humans know themselves. Humans are therefore open victims of manipulation. One can get people with DS to vote the way one wants them to vote. Metaphorically speaking, DS “know” exactly which keys of the piano they must play so that the music resounds.

12th principle: Data-based systems must be able to adapt to humans.

This principle guarantees that DS are human-centered and that DS serve humans, not vice versa. At the same time, it makes sure that DS are inclusive.

13th principle: Data-based systems and their performance (efficiency and effectiveness) must be controlled, monitored, measured, and evaluated on a regular basis, and their assessment must be published each time such that it is accessible to the broader public.

This principle guarantees that the ethical upsides and ethical downsides of DS are adequately addressed in relation to their essential and fundamental nature for humans and that public and transparent participatory and democratic monitoring, oversight, and handling are in place out of respect for the human dignity and freedom of all humans.

14th principle: Data-based systems must include an “emergency button” (metaphorically) for humans and an “ethics-black-box” enabling an ethical analysis.

This principle ensures, with its first part, that humans remain in charge due to their exclusive responsibility for DS as introduced above and that humanity and the planet are prepared for a worst-case scenario. With its second part, it empowers humans to review decisions and actions by DS and evaluate them from an ethical point of view.

15th principle: Data-based systems must be approved by national regulatory authorities—similar to food and drug regulatory agencies protecting public health by ensuring the safety, efficacy, security, and sustainability of data-based systems, regulating the manufacturing, marketing, and distribution of data-based systems, helping to further innovations that make data-based systems more effective, safer, and more affordable, and empowering the public by providing the accurate, independent, and science-based information they need to accept and use data-based systems.

This principle guarantees that only human rights-respecting and sustainable DS can be put on the market—an aspect that is naturally guaranteed in other industries (e.g., in the pharmaceutical industry, of course a medical drug can only be brought to market after a careful approval process that excludes harm to nature and people)—and that human rights and sustainability continue to be respected, implemented, protected, and realized in the digitalized value chains as well as in the entire life cycle of DS.

16th principle: Research and development projects in the area of data-based systems must be approved by national regulatory authorities.

This principle ensures that research and development in such a high-risk area as DS happens within a well-defined legal and ethical framework based on the freedom and independence of research. E.g., in the area of nuclear technologies—knowing similar dual nature and the potential of an existential impact as DS—, research and development are strictly regulated in order to guarantee the flourishing of humanity and the planet, and this regulation is implemented and enforced resolutely by the IAEA.

17th principle: The conduct of research and development must respect these IDA principles.

This principle makes sure that present and future DS honor these IDA principles.

18th principle: Lethal automated weapons and lethal automated weapon systems are forbidden.

This principle guarantees the prohibition of lethal automated weapons and lethal automated weapon systems because automated weapons face fundamental criticism from an ethical standpoint in that they are not able, according to the laws, to distinguish in armed conflicts between combatants and non-combatants and to apply the principle of proportionality, resulting in more wrong actions and crimes. Beyond that, the “strategic robot problem” [93] and its ethical implications underpin the idea that automated weapons are ethically problematic. The “strategic robot problem” lies in the undermining of command and control structures by creating automated weapons to serve as combatants and commanders at the same time. From an ethical viewpoint, the following factors would also speak against automated weapons: the idea that they would lead to more wars [94] because of the reduction of “ability to make credible threats and assurances in a crisis” [95], the growing distance between human actions and their consequences, [96,97] less human involvement for the actors deploying them [98], a lower number of victims would be expected, [95,99] and a lower political price would need to be paid [100]. The latter becomes obvious for reasons including but not limited to the replacement of the fundamental reciprocity of combat, namely possessing as a soldier the power to kill while running the ongoing risk of being killed [101–103]. As a consequence of the latter, legislative oversight would not be respected, and this would work against the system of checks and balances.

19th principle: Data-based systems for human rights-violating surveillance are forbidden.

This principle ensures that these already existing and already deployed human rights-violating DSs are stopped and prohibited.

20th principle: Data-based systems for social scoring of humans by the state or by non-state actors are halted and forbidden.

This principle ensures that these already existing and already deployed human rights-violating DSs are stopped and prohibited.

21st principle: Data-based systems manipulating and undermining democracy are forbidden.

This principle ensures that these already existing and already deployed anti-democratic DS are discontinued and prohibited.

22nd principle: Data-based systems supporting or reinforcing totalitarian systems and dictatorships are forbidden.

This principle ensures that these already existing and already deployed anti-democratic DS are terminated and prohibited.

23rd principle: Data-based systems blazing the trail for “super-data-based systems”, or “singularity” are forbidden.

This principle guarantees that technological progress, which should lead to “super-DS” or “singularity”—both endangering humanity and the planet—is still stopped at a moment when humans are still in control and prohibited.

24th principle: “Super-data-based systems” or “singularity” are forbidden.

This principle ensures the flourishing of all humans and the planet by prohibiting “super-DS” or “singularity” because they represent existential threats for humanity and the eco-system.

25th principle: These principles so far must be included in the parameter setting for the creation, design, programming, development, production, training, and use of data-based systems.

This principle guarantees that the IDA principles are included in the creation, design, programming, development, production, training, and use of data-based systems.

26th principle: Designers, software engineers, manufacturers, producers, operators, providers, and users of data-based systems, as well as infrastructure providers and data analytics companies and their employees, must have adequate knowledge, skills, and competencies, including a basic applied ethics expertise.

This principle ensures that the professionals involved in the creation, design, programming, development, production, training, and use of data-based systems—because of their high relevance to the background of the IDA principles—possess adequate knowledge, skills, and competencies, also in applied ethics.

27th principle: Designers, software engineers, manufacturers, producers, operators, providers, and users of data-based systems, as well as infrastructure providers and data analytics companies and their employees, must be accountable. They must be able to take legal and ethical responsibility.

This principle guarantees the legal and ethical empowerment of the professionals involved in the creation, design, programming, development, production, training, and use of data-based systems because of their crucial role in the realization of the IDA principles.

28th principle: The principle of indivisibility of all IDA principles must be respected.

This principle ensures that all IDA principles complement each other, go hand in hand, and that they are all optimally realized.

29th principle: Any supplement or modification to these principles must be undertaken only by humans.

This principle guarantees the respect of the exclusive responsibility of humans for DS.

30th principle: Any supplement or modification to these principles must undoubtedly serve the realization of human rights for all humans and the United Nations’ Sustainable Development Goals.

This principle ensures that any necessary change of the IDA principles contributes to the realization of human rights of all humans and the United Nations’ Sustainable Development Goals.

6.3. Precise Regulation: Stimulating Economic Growth

More and stricter commitment to the legal framework is necessary, as is regulation that is precise, goal-oriented, and strictly enforced. The IDA would serve this necessity. In this way, regulation may also be advantageous economically. For example, the American regulation of air traffic and the aviation industry allowed an entire industry to flourish economically thanks to its high degree of precision, its clear orientation, and its uncompromising enforcement.¹

Compared to other models for global governance of DS, IDA promises to reach the precision, goal orientation, and strict enforcement not only necessary to guarantee

the flourishing of humanity and the planet from an ethical standpoint but also to foster innovation from an economic point of view.

6.4. IDA Is Realistic

What makes the establishment of an IDA realistic is not only its essential and minimum normative framework, its practice-oriented and participatory governance-structure as well as its strive for legitimacy combined with fostering innovation but also that in the past, humanity has shown that when the well-being of people and the planet is at stake, humanity can focus on what is technically feasible rather than blindly pursuing all that is technically possible.

Humanity did pursue nuclear technology, develop the atomic bomb, and even deploy it more than once. But to prevent yet worse events, humanity then massively restricted the research and development of nuclear technology despite overwhelming opposition by state and non-state actors. That nothing worse has happened is largely due to international guidelines, concrete enforcement mechanisms, and the International Atomic Energy Agency (IAEA) of the UN.

In the case of chlorofluorocarbons (CFCs), humanity also decided under the Montreal Protocol of 1987 [104] to ban substances that damage the ozone layer and to enforce the ban consistently. Here, the resistance was also huge, *inter alia*, due to special interests from the private sector. This regulation and its uncompromising enforcement led to the fact that the hole in the ozone layer is now slowly closing.

Beyond that, DS distinguish themselves from nuclear technology and from CFCs in front of all in three characteristics that increase the realizability of the establishment and the existential impact of IDA for humanity and the planet:

1. In order to function, DS must have power. This means that if a DS is violating human rights, threatening peace, or destroying the planet, it can be stopped by taking it off the power grid or by cutting off the power supply.
2. In order to function, DS must be connected because of its dependence on data flow. This means that if a DS is violating human rights, threatening peace, or destroying the planet, it can be stopped by disconnecting it.
3. While operating, every DS leaves data traces, allowing identification and accountability.

Finally, both examples—nuclear technology and CFCs—were, at their time, not able to benefit from the ethical positive potential of DS to provide an innovative and technology-based solution for the ethical challenges they were addressing, as is the case nowadays with HRBDS. HRBDS possesses the ethical positive potential to support IDA to avoid or master the ethical risks of DS.

6.5. Legal Basis for IDA

As introduced above, the line of argumentation showing the necessity of the establishment of IDA at the UN includes an analogy with the International Atomic Energy Agency (IAEA). These first preliminary thoughts about the legal basis of IDA developed in the following sections are inspired by the legal architecture of the IAEA [105–110].

The legal basis for the establishment of IDA should be a UN resolution elaborating and adopting the text of the Statute of IDA, which constitutes the following elements:

- a. Purpose: The purpose of IDA is—as defined above—to be a platform for technical cooperation in the field of digital transformation and DS, fostering human rights, safety, security, and peaceful uses of DS; and to act as a global supervisory and monitoring institution and regulatory authority, partnering with and supporting on a global level the work of the national regulatory authorities in the area of digital transformation and DS. It should foster the safe, secure, and peaceful use of data-based systems, contributing to international peace and security, the respect and realization of human rights, and the United Nations' Sustainable Development Goals.
- b. 30 IDA principles (please see above)

- c. Legal Status of IDA (please see below)
- d. Membership in the IDA (please see below)
- e. Rights and Responsibilities of IDA (please see below)
- f. Mechanisms, Measures, and Instruments of IDA (please see below)
- g. Governance of IDA (please see below)

6.6. *Legal Status of IDA*

While IDA is not a state under international law, it is an entity with “international legal personality”. States should recognize IDA as an entity that has some rights and privileges normally associated with a sovereign state. One of the primary attributes of an international organization such as IDA is its capacity to conclude international agreements with other “persons” having international legal personality under international law.

Beyond that, the legal status of IDA also embraces the relationship of IDA with the UN, including its regular reports to the UN General Assembly as well as, if necessary, to the UN Security Council and its institutional cooperation within its own decision-making processes.

These points ensure the ability of IDA to act and have a concrete impact, as well as its checks and balances.

6.7. *Membership*

State and non-state actors become members of the IDA by ratifying its statute.

Offering non-state actors membership in IDA is necessary due to the growing political importance of multi stakeholder participation, the increasing economic power (and corresponding responsibility) of multinational technology companies in the field of DS, as well as the global and multidisciplinary nature of DS.

6.8. *Rights and Responsibilities of IDA*

Vis-à-vis state and non-state actors (e.g., corporations), IDA is authorized to establish and administer safeguards to foster the realization of the 30 IDA principles, among others, and primarily,

- To guarantee that data-based systems (DS) are developed, produced, and deployed with respect to human rights;
- To ensure that data-based systems (DS) are promoting peace;
- To secure that data-based systems (DS) are fostering the realization of the UN SDGs;
- To apply safeguards, at the request of the parties, to any bilateral or multi-lateral arrangement;
- To apply safeguards to any of the data-based systems activities of a state or non-state actor at that state actor’s or non-state actor’s request.

IDA enjoys rights and responsibilities, including, among others:

- The right to permit or prohibit as regulatory authority research and development-projects of data-based systems (DS) in order to protect all humans and the planet (in analogy with such legally defined approval processes in the pharmaceutical industry protecting humans and the planet);
- The right to examine, as a global supervisory and monitoring institution, the design, research, and development of DS;
- The right to decide as a regulatory authority about the approval of access to the market of a DS (in analogy with such legally defined approval processes in the pharmaceutical industry protecting humans and the planet);
- The right to examine the production of DS as a global supervisory and monitoring institution;
- The right to examine the deployment of DS, including requesting operating records to assist in ensuring accountability for and control of DS as a global supervisory and monitoring institution;

- The right to require the submission of reports on the design, production, and/or deployment of DS regarding human rights, peace, and sustainability in global supervisory and monitoring institutions;
- The right to send as global supervisory and monitoring institution into the state actor or to the non-state actor members of the “enforcement committees” and/or inspectors, designated by IDA after consultation with the state actor or state actors, respectively, non-state actor or non-state actors concerned, who shall have access at all times to all places and data, and to any person who, by reason of his or her occupation, deals with data-based systems (DS);
- In the event of non-compliance and failure by the state or non-state actor concerned to take as global supervisory and monitoring institution requested corrective steps within a reasonable time, the right to curtail or suspend assistance and call for the disconnection and deactivation of a data-based system (DS) or data-based systems (DS) combined with the enforcement of corresponding fines (in percentages of the annual budget of the state, respectively, the profit before tax of a non-state actor) and with the execution of the suspension of the state or non-state actor from the exercise of the privileges and rights of IDA membership.
- The responsibility to serve as a platform for technical cooperation in the field of digital transformation and DS—of course collaborating and joining forces with already existing events and formats (e.g., the World Summit on the Information Society (WSIS), the International Governance Forum IGF, the AI for Good Global Summit . . .)—fostering human rights, safety, security, and peaceful uses of DS.

These rights and responsibilities of IDA serve the goal that IDA realizes a concrete and sustainable ethically positive impact in the domain entrusted to IDA.

6.9. Instruments, Enforcement Mechanisms, and Enforcement Measures of IDA

Legally binding instruments as part of international law should be developed under the auspices of IDA, by UN member states, fostering the enjoyment of ethical opportunities by all humans as well as mastering or avoiding ethical risks.

The process of concluding a legally binding instrument is as follows:

- either the usual process of UN treaties and can be started by either a state actor, a civil society actor or IDA
- or—in the case of urgency or emergency regarding the respect of human rights, peace, and sustainability—by IDA.

Every legally binding instrument will have an enforcement committee consisting of independent experts examining the implementation by the state and non-state actors based on annual reports by the state and non-state actors. The work of the enforcement committee is also informed by visits as well as inspectors equipped with enforcement-measures, more specifically:

- addressing the concrete opportunities and risks as well as the measure calling upon the state or non-state actor to remedy non-compliance;
- reporting non-compliance to the UN Member states and to the UN Security Council and the UN General Assembly;
- defining counter-measures against the state or non-state actor, including fines (in percentages of the annual budget of a state, respectively, the profit before tax of a non-state actor) and the suspension of the state or non-state actor from the exercise of the privileges and rights of IDA membership.

Non-binding instruments (e.g., IDA recommendations; IDA guidelines, IDA code of conducts, . . .) should be developed under the auspices of IDA for the promotion of human rights, peace, and sustainability even beyond the letter of the law, as well as addressing in the first phase new technology-based ethical opportunities and risks potentially inspiring new legally binding instruments. Every non-binding instrument will be accompanied by an advisory group consisting of independent experts consulting state and non-state

actors, receiving and reviewing the reports of state and non-state actors in line with the conceptualization of IDA as a learning organization, which is of utmost importance in the area of rapidly evolving technology-based innovation.

Already existing regional and national legally binding instruments should be promoted by IDA as good practices for other regions or states, including general comments evaluating them based on the purpose of IDA and the 30 IDA principles.

Already existing regional and national non-binding instruments should be promoted by IDA as good practices for other regions or states, including general comments evaluating them before based on the purpose of IDA and 30 IDA principles.

These instruments, enforcement mechanisms, and enforcement measures of IDA serve the aim of creating, in an inclusive and participatory way, the aspired concrete effects in the field of DS.

6.10. Governance of IDA

IDA governance shall include

- a. Director General (an individual endowed with executive powers), elected by the UN General Assembly
- b. Secretariat, elected by the Director General and confirmed by the Board (implementing body)
- c. Board (strategic body), elected by the UN General Assembly
- d. Tripartite Council consisting of three representatives of states, of the private sector, and of civil society (advisory body to the Director General and the Board), confirmed by the UN General Assembly
- e. UN Council consisting of the UN Secretariat General and all UN Agencies—especially UNESCO formulating the first global ethical standards for so-called “Artificial Intelligence” (collaborative body for the Director General and the Board)
- f. Enforcement committees, elected by the UN General Assembly (executive bodies)
- g. Advisory groups, elected by the UN General Assembly (consulting bodies for state and non-state actors)
- h. Inspectors team, appointed by the Director General and the relevant enforcement committee (monitoring body).

This governance structure of IDA ensures its ability to decide, act, and deliver concrete results and impact on the ground, its good governance, its embeddedness in the UN and the international community, as well as its interconnectedness with other international institutions and organizations in this field, its multi-stakeholder-participation, its inclusive approach integrating and benefiting from all the work and efforts already pursued and continuing to be pursued in this field, its excellent expertise, knowhow, and experience, as well as its ability to implement and enforce according to its mandate.

6.11. HRBDS Supporting IDA in Fulfilling Its Responsibilities

HRBDS as innovative technology-based solutions should support IDA in fulfilling its responsibilities as far as possible (ethical problems that could occur during the deployment of DS, e.g., contextual bias, should be foreseen as much as possible), e.g., by identifying human rights-violating DS such as racist or sexist apps. (In order to illustrate this point, if DS is able to master the complexity of identifying malignant tumor cells on a screen, it must also be possible to use DS to identify available racist or sexist apps.) This means concretely, e.g., that a state or non-state actor needs to submit digitally an application for its research- and development-project in the field of DS before starting it, and the assessment and evaluation before the 30 IDA principles and the legal requirements of this project can be performed by DS informing a final decision by humans.

An international fund should be established by state and non-state actors providing impact-investments in ventures striving to bring HRBDS on the market, supporting IDA in fulfilling its responsibilities.

This article could serve as a conceptual basis for the further realization of HRBDS and IDA as well as for further interdisciplinary research on HRBDS and IDA.

7. Broad Global Support for IDA

Besides a growing international and interdisciplinary network of experts that calls for the establishment of HRBDS and IDA [111], the Elders—an independent group of world leaders founded by Nelson Mandela that includes former UN Secretary General Ban Ki-moon, Ellen Johnson Sirleaf (former President of Liberia and Africa’s first elected female head of state as well as Nobel Peace Laureate), and Ireland’s first female President Mary Robinson—endorsed the concrete recommendations for human rights-based DS and a global agency to monitor them and called upon the UN to take appropriate action. In their statement of 31 May 2023, the Elders made two specific suggestions for action from the book “Digital Transformation and Ethics. Ethical Considerations on the Robotization and Automation of Society and the Economy and the Use of Artificial Intelligence” [16]—“human rights-based data-based systems” and above all the creation of an “International Data-Based Systems Agency IDA” at the UN following the model of the International Atomic Energy Agency IAEA.

Thus, the Elders declared that: “A new global architecture is needed to manage these powerful technologies within robust safety protocols, drawing on the model of the Nuclear Non-Proliferation Treaty and the International Atomic Energy Agency. These guardrails must ensure AI is used in ways consistent with international law and human rights treaties. AI’s benefits must also be shared with poorer countries. No existing international agency has the mandate and expertise to do all this. The Elders now encourage a country or group of countries to request as a matter of priority, via the UN General Assembly, that the International Law Commission draft an international treaty establishing a new international AI safety agency” [112].

The idea of a human rights-based and legally binding regulatory framework, as well as the establishment of an institution enforcing global regulation, enjoys the support of Pope Francis [113].

UN Secretary General António Guterres also supports the creation of an international AI watchdog body like the International Atomic Energy Agency (IAEA): “I would be favorable to the idea that we could have an artificial intelligence agency (. . .) inspired by what the international agency of atomic energy is today” [114,115]. He has called for a new UN body like IDA to tackle threats posed by artificial intelligence in the UN Security Council on 18 July 2023 [116].

UN High Commissioner for Human Rights Volker Türk has demanded “urgent action” and proposed human-rights-based DS and a coordinated global response towards an institutional solution like the creation of an “International Data-Based Systems Agency IDA” in his statement about AI and human rights on 12 July 2023 [83].

The UN Human Rights Council unanimously adopted, on 14 July 2023, its latest resolution on “New and emerging digital technologies and human rights” which included for the first time an explicit reference to AI, and the promotion and protection of human rights. The Resolution emphasizes that new and emerging technologies with an impact on human rights “may lack adequate regulation”, highlighted the “need for effective measures to prevent, mitigate and remedy adverse human rights impacts of such technologies” and stressed the need to respect, protect and promote human rights “throughout the lifecycle of artificial intelligence systems”. It called for frameworks for impact assessments related to human rights, for due diligence to assess, prevent, and mitigate adverse human rights impacts, and to ensure effective remedies, human oversight, and accountability [117].

On 21 March 2024, the UN General Assembly unanimously adopted a resolution “Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development” on the promotion of “safe, secure and trustworthy” “artificial intelligence (AI)” systems that will also benefit sustainable development for all. It empha-

sizes: “The same rights that people have offline must also be protected online, including throughout the life cycle of artificial intelligence systems” [1].

Also, some voices from multinational technology companies—among others, Sam Altman (Founder of OpenAI, which developed ChatGPT)—have called for IDA [118,119]. Now is the time.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Conflicts of Interest: The author declares no conflicts of interest.

Notes

- ¹ Dorian Selz, CEO and Founder of Squirro, made this observation at a workshop at the ETH Zurich on 10 April 2019.

References

1. UN General Assembly. Resolution Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development. 24 March 2024. Available online: <https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/78/L.49&Lang=E> (accessed on 13 April 2024).
2. Zuboff, S. *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*; PublicAffairs: London, UK, 2019.
3. Weizenbaum, J. Not without Us. *ETC A Rev. Gen. Semant.* **1987**, *44*, 42–48. [CrossRef]
4. Ohly, L. *Ethik der Robotik und der Künstlichen Intelligenz*; Theologisch-Philosophische Beiträge zu Gegenwartsfragen 22; Peter Lang: Berlin, Germany, 2019.
5. Kirchsclaeger, P.G. Artificial Intelligence and the Complexity of Ethics. *Asian Horiz.* **2020**, *14*, 587–600.
6. Tzafestas, S.G. *Roboethics: A Navigating Overview*; Intelligent Systems, Control and Automation: Science and Engineering; Springer: Cham, Switzerland, 2016.
7. Jansen, P.; Broadhead, S.; Rodrigues, R.; Wright, D.; Brey, P.; Fox, A.; Wang, N. A Report for the Sienna Project, an EU H2020 Research and Innovation Program under Grant Agreement. European Commission, 13 April 2018. Available online: <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b9f93f94&appId=PPGMS> (accessed on 13 April 2024).
8. Coeckelbergh, M. *AI Ethics*; The MIT Press Essential Knowledge; MIT Press: Cambridge, MA, USA, 2020.
9. Misselhorn, C. *Grundfragen der Maschinenethik*; Reclam: Stuttgart, Germany, 2018.
10. Dreyfus, H.L. *What Computers Can't Do: The Limits of Artificial Intelligence*; MIT Press: New York, NY, USA, 1972.
11. Dreyfus, H.L.; Dreyfus, S.E. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*; Free Press: New York, NY, USA, 1986.
12. Turkle, S. Authenticity in the age of digital companions. *Interact. Stud.* **2007**, *50*, 501–517. [CrossRef]
13. Manzeschke, A. Roboter in der Pflege: Von Menschen, Maschinen und anderen hilfreichen Wesen. *Ethik J.* **2019**, *5*, 1–11. Available online: https://www.ethikjournal.de/fileadmin/user_upload/ethikjournal/Texte_Ausgabe_2019_1/Manzeschke_1_Nov_FINAL.pdf (accessed on 13 April 2024).
14. Kant, I. *Grundlegung zur Metaphysik der Sitten*. In *Werkausgabe*, 7th ed.; Weischedel, W., Ed.; Suhrkamp: Frankfurt am Main, Germany, 1974; Volume 7.
15. Kirchsclaeger, P.G. *Ethical Decision-Making*; Nomos: Baden-Baden, Germany, 2023.
16. Kirchsclaeger, P.G. *Digital Transformation and Ethics. Ethical Considerations on the Robotization and Automation of Society and the Economy and the Use of Artificial Intelligence*; Nomos: Baden-Baden, Germany, 2021.
17. Virt, G. Damit Menschsein Zukunft hat. In *Theologische Ethik im Einsatz für eine humane Gesellschaft*; Marschuetz, G., Pruellergenteufel, G.M., Eds.; Echter: Wuerzburg, Austria, 2007.
18. Knight, W. The Dark Secret at the Heart of AI. *MIT Technology Review*. 11 April 2017. Available online: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/> (accessed on 13 April 2024).
19. Bathaee, Y. The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harv. J. Law Technol.* **2018**, *31*, 889–938.
20. Knight, W. The Financial World Wants to Open AI's Black Boxes. *MIT Technology Review*. 13 April 2017. Available online: <https://www.technologyreview.com/s/604122/the-financial-world-wants-to-open-ais-black-boxes/> (accessed on 13 April 2024).
21. Castelvechi, D. Can We Open the Black Box of AI? *Nature* **2016**, *538*, 20–23. [CrossRef] [PubMed]
22. Iversen, G.R.; Gergen, M. *Statistics: The Conceptual Approach*; Springer Undergraduate Textbooks in Statistics; Springer: New York, NY, USA, 1997.

23. UNESCO. *Elaboration of a Recommendation on the Ethics of Artificial Intelligence*. 2020. Available online: <https://en.unesco.org/artificial-intelligence/ethics> (accessed on 13 April 2024).
24. Demuth, Y. Die Unheimliche Macht der Algorithmen. *Beobachter*, 26 April 2018. Available online: <https://www.beobachter.ch/digital/multimedia/big-data-die-unheimliche-macht-der-algorithmen> (accessed on 13 April 2024).
25. Clifford, C. Elon Musk: Mark My Words—A.I. Is Far More Dangerous than Nukes. *CNBC*. 13 March 2018. Available online: <https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html> (accessed on 13 April 2024).
26. Kharpal, A. Stephen Hawking Says A.I. Could Be Worst Event in the History of Our Civilization. *CNBC*, 6 November 2017. Available online: <https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html> (accessed on 13 April 2024).
27. Montreal Declaration. For a Responsible Development of AI. 2018. Available online: <https://montrealdeclaration-responsibleai.com/the-declaration/> (accessed on 13 April 2024).
28. Association for Computing Machinery US Public Policy Council. Statement on Algorithmic Transparency and Accountability. 12 January 2017. Available online: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf (accessed on 13 April 2024).
29. Danish Ministry of Finance, Agency for Digitalisation. Denmark's National Strategy for Artificial Intelligence. Available online: <https://en.digst.dk/policy-and-strategy/denmark-s-national-strategy-for-artificial-intelligence/> (accessed on 13 April 2024).
30. Commission Nationale de l'Informatique et des Libertés (CNIL); European Data Protection Supervisor (EDPS); Garante per la Protezione dei Dati Personali. Declaration on Ethics and Data Protection in Artificial Intelligence. In Proceedings of the 40th International Conference of Data Protection and Privacy Commissioners (ICDPP), Brussels, Belgium, 23 October 2018; Available online: https://www.privacyconference2018.org/system/files/2018-10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf (accessed on 13 April 2024).
31. Japanese Society for Artificial Intelligence (JSAI). The Japanese Society for Artificial Intelligence Ethical Guidelines. 2017. Available online: <https://www.ai-gakkai.or.jp/ai-elsi/wp-content/uploads/sites/19/2017/05/JSAI-Ethical-Guidelines-1.pdf> (accessed on 13 April 2024).
32. Villani, C. *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*. 2018. Available online: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf (accessed on 13 April 2024).
33. House of Commons of the United Kingdom—Science and Technology Committee. Algorithms in Decision-Making. Report No. 4, UK. 2017. Available online: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf> (accessed on 13 April 2024).
34. Swiss Federal Council. *Leitlinien Künstliche Intelligenz für den Bund. Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung*; Swiss Federal Council: Bern, Switzerland, 2020.
35. Engineering and Physical Sciences Research Council (EPSRC). Principles of Robotics. 2010. Available online: <https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/> (accessed on 13 April 2024).
36. Fairness, Accountability, and Transparency in Machine Learning (FAT/ML). Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. 2016. Available online: www.fatml.org/resources/principles-for-accountable-algorithms (accessed on 13 April 2024).
37. Australian Human Rights Commission. Human Rights and Technology. Discussion Paper; 2019. Available online: https://tech.humanrights.gov.au/?_ga=2.211445781.1641337062.1609843370-1930064430.1609843370 (accessed on 13 April 2024).
38. Future of Life Institute. Asilomar AI Principles. 2017. Available online: <https://futureoflife.org/ai-principles/> (accessed on 13 April 2024).
39. Partnership on AI. Tenets. 2016. Available online: <https://partnershiponai.org/about/#tenets> (accessed on 13 April 2024).
40. Austrian Council on Robotics and Artificial Intelligence. Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz. 2018. Available online: https://www.bmk.gv.at/dam/jcr:f2f7a973-8aa4-4be8-9a6b-0c7c44e73ce4/white_paper_robotikrat.pdf (accessed on 13 April 2024).
41. The Public Voice Coalition. Universal Guidelines on Artificial Intelligence (UGAI). *Brussels*. 23 October 2018. Available online: <https://thepublicvoice.org/ai-universal-guidelines/> (accessed on 13 April 2024).
42. Amnesty International; Access Now. The Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems. Declaration launched at RightsCon 2018, Toronto, Canada. 16 May 2018. Available online: <https://www.torontodeclaration.org/#:~:text=The%20Toronto%20Declaration,-Protecting%20the%20right&text=It%20calls%20on%20governments%20and,to%20equality%20and%20non-discrimination> (accessed on 13 April 2024).
43. Beijing Academy of Artificial Intelligence (BAAI). Beijing AI Principles. 2019. Available online: <https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/> (accessed on 13 April 2024).
44. European Group on Ethics in Science and New Technologies. Ethics of Information and Communication Technologies. 2009. Available online: <https://op.europa.eu/en/publication-detail/-/publication/c35a8ab5-a21d-41ff-b654-8cd6d41f6794/language-en/format-PDF/source-77404276> (accessed on 13 April 2024).
45. European Group on Ethics in Science and New Technologies. Ethics of Security and Surveillance Technologies. 2014. Available online: <https://op.europa.eu/en/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-54fc3225c969> (accessed on 13 April 2024).

46. European Group on Ethics in Science and New Technologies. The Ethical Implications of New Health Technologies and Citizen Participation. 2015. Available online: <https://op.europa.eu/en/publication-detail/-/publication/e86c21fa-ef2f-11e5-8529-01aa75ed71a1/language-en/format-PDF/source-77404221> (accessed on 13 April 2024).
47. European Group on Ethics in Science and New Technologies. Statement on Artificial Intelligence, Robotics and «Autonomous» Systems. 2018. Available online: <https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en> (accessed on 13 April 2024).
48. European Group on Ethics in Science and New Technologies. Future of Work, Future of Society. 2018. Available online: <https://op.europa.eu/en/publication-detail/-/publication/9ee4fad5-eef7-11e9-a32c-01aa75ed71a1/language-en/format-PDF/source-314882261> (accessed on 13 April 2024).
49. High-Level Expert Group on Artificial Intelligence HLEG AI of the European Commission. Ethics Guidelines for Trustworthy Artificial Intelligence. 2019. Available online: <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence> (accessed on 13 April 2024).
50. Council of Europe. Algorithms and Human Rights—Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications. 2018. Available online: <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> (accessed on 13 April 2024).
51. Council of Europe. Recommendation of the Committee of Ministers: Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment. 2018. Available online: <https://edoc.coe.int/en/children-and-the-internet/7921-guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-the-digital-environment-recommendation-cmrec20187-of-the-committee-of-ministers.html> (accessed on 13 April 2024).
52. Council of Europe. Consultative Committee of the Convention for the Protection of Individuals with Regard to the Processing of Personal Data (Convention 108): Guidelines on Artificial Intelligence and Data Protection. 2019. Available online: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8> (accessed on 13 April 2024).
53. Council of Europe. Committee of Ministers: Declaration on the Manipulative Capabilities of Algorithmic Processes. 2019. Available online: https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b (accessed on 13 April 2024).
54. Council of Europe. European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment. Available online: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c> (accessed on 13 April 2024).
55. European Union. Charter of Fundamental Digital Rights of the European Union. 2018. Available online: <https://digitalcharta.eu/> (accessed on 13 April 2024).
56. European Digital Rights (EDRi). The Charter of Digital Rights. A Guide for Policy-Makers. 2014. Available online: https://edri.org/wp-content/uploads/2014/06/EDRi_DigitalRightsCharter_web.pdf (accessed on 13 April 2024).
57. UNESCO COMEST. Report of COMEST on Robotic Ethics. World Commission on the Ethics of Scientific Knowledge and Technology. 2017. Available online: <https://unesdoc.unesco.org/images/0025/002539/253952E.pdf> (accessed on 13 April 2024).
58. Information Technology Industry Council (ITIC). AI Policy Principles. 2017. Available online: www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf (accessed on 13 April 2024).
59. Dutton, I.R. Engineering code of ethics. *IEEE Potentials* **1990**, *9*, 30–31. [CrossRef]
60. Price, A. First International Standards Committee for Entire AI Ecosystem. *IE e-tech* **2018**, *3*. Available online: <https://etech.iec.ch/issue/2018-03/first-international-standards-committee-for-entire-ai-ecosystem> (accessed on 13 April 2024).
61. D 64 Zentrum für Digitalen Fortschritt. Grundwerte in der Digitalisierten Gesellschaft. Der Einfluss Künstlicher Intelligenz auf Freiheit, Gerechtigkeit und Solidarität. Available online: <https://d-64.org/wp-content/uploads/2018/11/D64-Grundwerte-KI.pdf> (accessed on 13 April 2024).
62. UNICEF. The State of the World’s Children: Children in a Digital World. 2017. Available online: https://www.unicef.lu/site-root/wp-content/uploads/2017/12/SOWC2017_ENG_lores.pdf?_adin=132415900 (accessed on 13 April 2024).
63. Association for Computing Machinery’s Committee on Professional Ethics. 2018 Code, Draft 3. 2018 ACM Code of Ethics and Professional Conduct: Draft 3, 2017. Available online: <https://ethics.acm.org/2018-code-draft-3/> (accessed on 13 April 2024).
64. WeGovNow. Towards #WeGovernment: Collective and Participative Approaches for Addressing Local Policy Challenges. 2020. Available online: <https://wegovnow.eu/> (accessed on 13 April 2024).
65. The Future of Privacy Forum. Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models. 2018. Available online: <https://fpf.org/wp-content/uploads/2018/06/Beyond-Explainability.pdf> (accessed on 13 April 2024).
66. UNI Global Union. Top 10 Principles für Ethical Artificial Intelligence. 2017. Available online: www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (accessed on 13 April 2024).
67. Institute of Electrical and Electronic Engineers (IEEE) Standards Association. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available online: <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html> (accessed on 13 April 2024).
68. Internet Governance Forum (IGF). The Global Multistakeholder for Dialogue on Internet Governance Issues. 2014. Available online: <http://intgovforum.org/cms/2014/IGFBrochure.pdf> (accessed on 13 April 2024).
69. Internet Governance Forum (IGF). Best Practice Forum on Internet of Things, Big Data, Artificial Intelligence. 2019. Available online: https://www.intgovforum.org/multilingual/filedepot_download/8398/1915 (accessed on 13 April 2024).
70. ISO. Information Technology—Electronic Discovery—Part 3: Code of Practice for Electronic Discovery. 2020. Available online: <https://www.iso.org/standard/78648.html> (accessed on 13 April 2024).

71. AI Safety Summit. The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023. Available online: <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023> (accessed on 13 April 2024).
72. OECD. Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. 2019. Available online: <https://legalinstruments.oecd.org/api/print?ids=648&lang=en> (accessed on 13 April 2024).
73. World Summit on the Information Society. Declaration of Principles, Building the Information Society: A Global Challenge in the New Millennium. World Summit on the Information Society: Geneva, Switzerland, 12 December 2003. Available online: <https://www.itu.int/net/wsis/docs/geneva/official/dop.html> (accessed on 13 April 2024).
74. Institute of Electrical and Electronic Engineers (IEEE). Ethically Aligned Design Version 2. Available online: <https://standards.ieee.org/industry-connections/ec/ead-v1.html> (accessed on 13 April 2024).
75. National Society of Professional Engineers (NSPE). Code of Ethics for Engineers. Available online: <https://www.nspe.org/resources/ethics/code-ethics> (accessed on 13 April 2024).
76. American Society of Mechanical Engineers (ASME) Standards. Available online: <https://www.asme.org/codes-standards> (accessed on 13 April 2024).
77. Nevejans, N. *European Civil Law Rules in Robotics. Study for the JURI Committee*; European Parliament: Strasbourg, France, 2016. Available online: [https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU\(2016\)571379_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf) (accessed on 13 April 2024).
78. European Parliament. Artificial Intelligence Act: Deal on Comprehensive Rules for Trustworthy AI. 2023. Available online: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai> (accessed on 13 April 2024).
79. Bradford, A. *Digital Empires. The Global Battle to Regulate Technology*; Oxford University Press: Oxford, UK, 2023.
80. Carnegie Council for Ethics in International Affairs. Envisioning Modalities for AI Governance: A Response from AIEI to the UN Tech Envoy. Artificial Intelligence & Equality Initiative. 2023. Available online: <https://www.carnegiecouncil.org/media/article/envisioning-modalities-ai-governance-tech-envoy#gaio> (accessed on 13 April 2024).
81. Baker & McKenzie International. Can a Global Framework Regulate AI Ethics? *Insight Plus*, 8 November 2023. Available online: <https://insightplus.bakermckenzie.com/bm/investigations-compliance-ethics/international-can-a-global-framework-regulate-ai-ethics> (accessed on 13 April 2024).
82. WEF. How Can the Aviation Industry Make AI Safer? 2022. Available online: <https://www.weforum.org/agenda/2022/08/how-can-aviation-industry-make-ai-safer/> (accessed on 13 April 2024).
83. Türk, V.; Office of the High Commissioner for Human Rights. Artificial Intelligence Must Be Grounded in Human Rights, Says High Commissioner. High Level Side Event of the 53rd Session of The Human Rights Council. 12 July 2023. Available online: <https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner> (accessed on 13 April 2024).
84. Andrzejewski, C. “Team Jorge”: In the Heart of a Global Disinformation Machine. *Forbidden Stories*. 15 February 2023. Available online: <https://forbiddenstories.org/story-killers/team-jorge-disinformation/> (accessed on 13 April 2024).
85. Kirchsclaeger, P.G. *Wie Können Menschenrechte Begründet Werden? Ein für Religiöse und Säkulare Menschenrechtskonzeptionen Anschlussfähiger Ansatz*; ReligionsRecht im Dialog 15; LIT-Verlag: Muenster, Germany, 2013.
86. Kirchsclaeger, P.G. Human Rights as an Ethical Basis for Science. *J. Law Inf. Sci.* **2013**, *22*, 1–17.
87. Kirchsclaeger, P.G. Das ethische Charakteristikum der Universalisierung im Zusammenhang des Universalitätsanspruchs der Menschenrechte. In *Gleichheit und Universalität*; Ast, S., Mathis, K., Haenni, J., Zabel, B., Eds.; Archiv für Rechts- und Sozialphilosophie 128; Franz Steiner: Stuttgart, Germany, 2011; pp. 301–312.
88. Council of Europe, Council of Europe and Artificial Intelligence. 2024. Available online: <https://www.coe.int/en/web/artificial-intelligence> (accessed on 13 April 2024).
89. United Nations Human Rights Council. Resolution New and Emerging Digital Technologies and Human Rights. No. 41/11. 13 July 2023. Available online: <https://www.ohchr.org/en/hr-bodies/hrc/advisory-committee/digital-technologiesand-hr> (accessed on 13 April 2024).
90. Kirchsclaeger, P.G. Ethics and Human Rights. *Ancilla Iuris* **2014**, *59*, 59–98.
91. Laaff, M. Ok, Zoomer. *Die Zeit*. 31 March 2020. Available online: <https://www.zeit.de/digital/2020-03/videokonferenzen-zoom-app-homeoffice-quarantaene-coronavirus> (accessed on 13 April 2024).
92. Partnership on AI. Our Pillars. 2017. Available online: <https://partnershiponai.org/about/#pillars> (accessed on 13 April 2024).
93. Roff, H.M. The Strategic Robot Problem: Lethal Autonomous Weapons in War. *J. Mil. Ethics* **2014**, *13*, 211–227. [CrossRef]
94. Kahn, L. Military Robots and the Likelihood of Armed Combat. In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*; Lin, P., Abney, K., Jenkins, R., Eds.; Oxford University Press: New York, NY, USA, 2017; pp. 274–287.
95. Leys, N. Autonomous Weapon Systems and International Crises. *Strateg. Stud. Q.* **2018**, *12*, 48–73.
96. Singer, P. *Wired for War: The Robotics Revolution and Conflict in the 21st Century*; Penguin Press: New York, NY, USA, 2009.
97. Pappenberger, M. Schattenkriege im 21. Jahrhundert. Die Automatisierung des Krieges durch Drohnen und Roboterwaffen. *Forum Paz.* **2013**, *2*, 38–44.
98. Grut, C. The Challenge of Autonomous Lethal Robotics to International Humanitarian Law. *J. Confl. Secur. Law* **2013**, *18*, 5–23. [CrossRef]

99. Wallach, W. *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control*; Basic Books: New York, NY, USA, 2015.
100. Wagner, M. The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems. *Vanderbilt J. Transnatl. Law* **2014**, *47*, 1371–1424.
101. Broeckling, U. Heldendämmerung? Der Drohnenkrieg und die Zukunft des militärischen Heroismus. *BEHEMOTH A J. Civilis.* **2015**, *8*, 97–107.
102. Hood, J. The Equilibrium of Violence: Accountability in the Age of Autonomous Weapons Systems. *Int. Law Manag. Rev.* **2015**, *11*, 12–40. [CrossRef]
103. Kaufmann, S. Der ‘digitale Soldat’. Eine Figur an der Front der Informationsgesellschaft. In *Forschungsthema: Militär*; Apelt, M., Ed.; Springer: Wiesbaden, Germany, 2010; pp. 271–294.
104. The Montreal Protocol. About Montreal Protocol. UN Environment Programme (UNEP). 1987. Available online: <https://www.unep.org/ozonaction/who-we-are/about-montreal-protocol> (accessed on 13 April 2024).
105. International Atomic Energy Agency (IAEA). The International Legal Framework for Nuclear Security. 2011. Available online: <https://www.iaea.org/publications/8565/the-international-legal-framework-for-nuclear-security> (accessed on 13 April 2024).
106. International Atomic Energy Agency (IAEA). *IAEO Basiswissen. Den Beitrag nuklearer Technik zur Gesellschaft maximieren und ihre friedliche Verwendung verifizieren*; IAEA: Vienna, Austria, 2013.
107. Stoiber, C.; Baer, A.; Pelzer, N.; Tonhause, W. *Handbook on Nuclear Law*; IAEA: Vienna, Austria, 2003.
108. Rockwood, L. *Legal Framework for IAEA Safeguards*; IAEA: Vienna, Austria, 2013.
109. El Baradei, M.; Nwogugu, E.; Rames, J. International law and nuclear energy: Overview of the legal framework. *IAEA Bull.* **1995**, *3*, 16–25.
110. Sharma, S.K. The IAEA and the UN family: Networks of nuclear co-operation. *IAEA Bull.* **1995**, *3*, 10–15.
111. IDA. Supporters of IDA. 2024. Available online: <https://idaonline.ch/supporters-of-ida/> (accessed on 13 April 2024).
112. The Elders. The Elders Urge Global Co-Operation to Manage Risks and Share Benefits of AI. 2023. Available online: <https://theelders.org/news/elders-urge-global-co-operation-to-manage-risks-and-share-benefits-ai> (accessed on 13 April 2024).
113. Pope Francis. Artificial Intelligence and Peace. Message of Pope Francis for the 57th World Day of Peace. 2024. Available online: <https://www.vatican.va/content/francesco/en/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html> (accessed on 13 April 2024).
114. Guterres, A. UN Chief Backs Idea of Global AI Watchdog Like Nuclear Agency. 12 June 2023. Available online: <https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/> (accessed on 13 April 2024).
115. Secretary-General Urges Broad Engagement from All Stakeholders towards United Nations Code of Conduct for Information Integrity on Digital Platforms. Available online: <https://press.un.org/en/2023/sgsm21832.doc.htm> (accessed on 13 April 2024).
116. Guterres, A. Secretary-General Urges Security Council to Ensure Transparency, Accountability, Oversight, in First Debate on Artificial Intelligence. 18 July 2023. Available online: <https://press.un.org/en/2023/sgsm21880.doc.htm> (accessed on 13 April 2024).
117. UN Human Rights Council. Resolution New and Emerging Digital Technologies and Human Rights. No. 41/11. 13 July 2023. Available online: <https://documents.un.org/doc/undoc/gen/g23/146/09/pdf/g2314609.pdf?token=dLWzJnULXDGnJTLOJg&fe=true> (accessed on 13 April 2024).
118. Euronews. OpenAI’s Sam Altman calls for an International Agency Like the UN’s Nuclear Watchdog to Oversee AI. *Euronews*, 7 June 2023. Available online: <https://www.euronews.com/next/2023/06/07/openais-sam-altman-calls-for-an-international-agency-like-the-uns-nuclear-watchdog-to-over> (accessed on 13 April 2024).
119. Santelli, F. Sam Altman: In Pochi Anni l’IA Sarà Inarrestabile, Serve Un’agenzia Come Per L’energia Atomica. *La Repubblica*, 18 January 2024. Available online: https://www.repubblica.it/economia/2024/01/18/news/sam_altman_in_pochi_anni_lia_sara_inarrestabile_serve_unagenzia_come_per_lenergia_atomica-421905376/amp/ (accessed on 13 April 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.