Izvirni znanstveni članek/Article (1.01)
Bogoslovni vestnik/Theological Quarterly 85 (2025) 2, 447—468
Besedilo prejeto/Received:05/2025; sprejeto/Accepted:11/2025
UDK/UDC: 17.021.4:004.89
DOI: 10.34291/BV2025/02/Kirchschlaeger

Peter G. Kirchschlaeger

Securing a Future for Humanity and the Planet in Times of "AI": The Need for an International Data-Based Systems Agency (IDA) at the UN

Zagotavljanje prihodnosti za človeštvo in planet v času ‚umetne inteligence': potreba po Mednarodni agenciji za podatkovne sisteme (MAPS) v okviru OZN

Abstract: In 2024, the UN General Assembly has adopted a resolution aiming for "safe, secure and trustworthy artificial intelligence systems" and the "Pact for the Future" including a "Global Digital Compact" and a "Declaration on Future Generations". It is now urgent to implement and build on the UN General Assembly resolution and the "Pact for the Future". This article suggests that an International Data-Based Systems Agency (IDA) needs to be established urgently at the UN as a global platform for technical cooperation in the field of data-based systems (DS) fostering human rights, safety, security, a rules-based international order, sustainability, and peaceful uses of data-based systems (DS) with a multilateral approach, as well as a global supervisory and monitoring institution and regulatory authority in the area of data-based systems (DS) responsible for access to market approval.

Keywords: Artificial Intelligence (AI), Data-Based Systems (DS), Ethics, Human Rights, Sustainability, Responsibility, Human Rights-Based Data-Based Systems (HRBDS), UN, International Data-Based Systems Agency (IDA)

Povzetek: Leta 2024 je Generalna skupščina Združenih narodov sprejela resolucijo, katere cilj so »varni, zanesljivi in zaupanja vredni sistemi umetne inteligence« ter »Pakt za prihodnost«, ki vključuje »Globalni digitalni dogovor« in »Deklaracijo o prihodnjih generacijah«. Zdaj je nujno, da se resolucija Generalne skupščine Združenih narodov in »Pakt za prihodnost« izvajata in nadgradita. V prispevku predlagamo, da je treba v okviru ZN nujno ustanoviti Mednarodno agencijo za podatkovne sisteme (MAPS) kot globalno platformo za tehnično sodelovanje na področju podatkovnih sistemov (PS). Slednja naj krepi človekove pravice, varnost, zaščito, mednarodni pravni red, ki temelji na pravilih, trajnosti in miroljubni rabi sistemov na podlagi podatkov (PS) z večstranskim pristopom, pa tudi globalne nadzorne in spremljevalne institucije ter regulativne organe na področju podatkovnih sistemov (PS), odgovorne za dostop do tržne odobritve.

Ključne besede: umetna inteligenca (UI), podatkovni sistemi (PS), etika, človekove pravice, trajnost, odgovornost, podatkovni sistemi, temelječi na človekovih pravicah (PSČP), OZN, Mednarodna agencija za podatkovne sisteme (MAPS)

1. First Small Steps Are Achieved

So-called "artificial intelligence" (AI) possesses upsides and downsides from an ethical perspective. A global consensus has emerged that some action is needed to ensure that the ethical opportunities of "AI" are fostered, that all humans benefit from the ethical opportunities, and that the ethical risks are avoided or mastered. First small steps to ensure a future for humanity and the planet are achieved: The UN Human Rights Council unanimously adopted, on July 14, 2023, its latest resolution on "New and emerging digital technologies and human rights" (UN Human Rights Council 2023) which addressed for the first time the promotion and protection of human rights with explicit reference to "AI". 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 impact, and to ensure effective remedies, human oversight, and accountability.

On March 21, 2024, the UN General Assembly unanimously adopted a resolution "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development" (UN General Assembly 2024) on the promotion of "safe, secure and trustworthy" "artificial intelligence (AI)" systems that will also benefit sustainable development for all. It emphasizes: "The same rights that people have offline must also be protected online, including throughout the life cycle of artificial intelligence systems."

On September 22, 2024, the UN General Assembly adopted the "Pact for the Future" including a "Global Digital Compact" and a "Declaration on Future Generations" (United Nations 2024).

These are small steps in the right direction and represent first significant foundations for a future where all humans can live a life with human dignity and the planet enjoys sustainability. At the same time, it becomes clear if one poses the following test questions that these first small steps are not enough and increase the necessity to implement and build on them:

a) Can human rights-violating state actions or business practices be concretely stopped because of these steps? No.

b) Do multinational technology companies need to change anything regarding their human rights-violating business practices because of these steps? No.

c) Can states or multinational technology companies be held accountable for their human violations based on these steps? No.

These negative responses as well as the following realities undermine that more needs to be done:

- In the year 2025 it is still possible, e.g., to put on the market an app that sexualizes children's images (Lensa 2024; Heikkilä 2022), and the only thing that happens to this company is that it makes a lot of money from it. We need to do something about this! Urgent action is needed.
- In the year 2025 it is now possible with "social media" to destabilize within a few hours a peaceful, functioning, and wealthy country (e.g., the UK far-right riots 2024) (Cadwalladr 2024).
- In the year 2025 it is now possible with "generative AI" to destroy a politician or business leader with a "Hollywood-quality" deep fake (Verma and Zakrzewski 2024).
- In the year 2025, the energy-consumption (Ren and Wierman 2024) and water-consumption (Gordon 2024; Ren and OECD 2023) by "AI" is exploding.
- In the year 2025, the global "digital divide" is growing (International Labour Organization 2024).

"Artificial intelligence (AI)" has ethical implications: opportunities as well as risks. "AI" can be powerful for fostering human rights and sustainability, but can, at the same time, be dangerous in violating human rights and in its ecological negative impact. Elon Musk who needs to be criticized for many aspects from an ethical perspective but is definitely not someone who could be blamed for being against technological progress, has warned that "AI is far more dangerous than nukes [nuclear warheads]. Far. So why do we have no regulatory oversight? This is insane." (Clifford 2018)

The late Stephen Hawking pointed 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." (Kharpal 2017)

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

So-called "artificial intelligence" can be defined as striving by technical means to imitate or fulfill cognitive functions of human thought. In the course of an ethical critique of so-called "AI", it becomes clear that so-called "artificial intelligence" does not comprise the sum of human knowledge, nor is it objective, fair and neutral. It is only based on certain data, which increasingly includes the data that so-called "generative AI" (such as ChatGPT) generates itself. As a result, so-called "AI" is getting dumber and dumber.

From an ethical standpoint, the above-mentioned starting point regarding so-called "AI" is criticized because intelligence does not just consist in the solution of a cognitive task but also in the way it is pursued (Misselhorn 2018). In view of the nature of so-called "AI", doubts arise from an ethical perspective if the term is even adequate, because so-called "AI" strives to imitate human intelligence, but this is limited to a certain area of intelligence (e.g., certain cognitive capacities) (Dreyfus 1972; Dreyfus and Dreyfus 1986). Furthermore, it is to be assumed that so-called "AI" 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 (Turkle 2007). 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" (Manzeschke 2019).

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 (Kant 1974). 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 (Kirchschlaeger 2023). While a human can know that he or she does something ethically right or wrong, machines cannot identify the ethical quality by itself.

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 (Kirchschlaeger 2021).

Finally, sometimes ethics must go beyond principles, norms, and rules to be sensitive to the rule-transcending uniqueness of the concrete (Kirchschlaeger 2023). 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. "In general, concrete ethical, positive legal and many other norms that are generally applicable, although indispensable, are not sufficient to guarantee the basic humanity /…/. 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." (Virt 2007, 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 machines 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 and legal principles and norms; set a framework, goals, and limits for machines; 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 (DS)" would be more appropriate than "artificial intelligence" because this term describes what actually constitutes so-called "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 decisions ("black-box-problem") (Bathaee 2018; Castelvecchi 2016; Knight 2017a; 2017b), its wide vulnerability to systemic errors, its deep exposure for confusing causality with correla-

tion (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) (Iversen and Gergen 1997), and its high probability of biased and discriminatory data and algorithms leading to biased and discriminatory data-based evaluations, predictions, and decisions embrace its major disadvantages (Coeckelbergh 2020; UNESCO 2024). "Algorithms are opinions embedded in codes. They are not objective." (Demuth 2018) They are not neutral. They serve specific goals and purposes. They are highly complex math with enormous amount of data – nothing more.

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. Exclusive Responsibility of Humans

The previous considerations suggest that, due to the lack of moral capability of machines, humans have the exclusive responsibility for machines. DS cannot act as moral subjects due to their lack of autonomy as elaborated above. The definition of ethical and legal principles and norms, of a framework for DS as well as ethical decisions belong in the hands of humans. This is because the potential that technologies have in terms of ethical decisions and actions is nowhere near the moral capability of humans, as machines lack not only autonomy, but also vulnerability, conscience, freedom and responsibility that are essential for moral capability (Kirchschlaeger 2021). People also have a responsibility to examine, analyze, and evaluate technology-based innovations from an ethical point of view and not to realize everything that is technically feasible. It is also important to take into account the complexity of ethics, which includes the "dual use" problem, ambivalence, and the above-mentioned rule-transcending uniqueness of the concrete (Kirchschlaeger 2023). Due to their lack of moral capability, DS reach their limits here. A complete transfer of ethics to mathematics, programming or the training of machines proves to be impossible due to the complexity of ethics and the lack of moral capability of DS.

The ethical responsibility of humans for DS also means design responsibility: DS do not simply fall from the sky or are predetermined by nature, but it is up to humans to decide what and how DS are designed, developed, produced, deployed and used, or not used due to ethical concerns. The latter also includes the responsibility to use DS in a targeted manner where they serve humanity and the planet, and to develop DS in those areas of human intelligence where they already massively outperform humans – such as handling large amounts of data, data storage, computing power, and logical deduction. DS can substantially support human thinking by storing facts and perceiving complexity.

4. Sharing Ethical Benefits and Mastering Ethical Risks by Global Governance 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. DS can make an important contribution to the promotion of human rights. For example, DS can be used to design new drugs based on protein structures (Atz et al. 2024). With the help of DS, tumor diseases can be classified more precisely than before in terms of their threat level (Zhang et al. 2024). DS offer enormous ethical opportunities.

When dealing with the ethical opportunities and risks of DS, the following two errors of thought are all too often made in the ethical discourse, which should be avoided: Firstly, ethical dangers could supposedly be offset by ethical opportunities. Secondly, human rights violations are accepted if great benefits are expected from DS. These are two errors of reasoning, because a violation of human rights with or through DS can neither be outweighed by the positive ethical potential of DS in terms of human rights nor by a high return, since human rights are at stake, so that every single violation of human rights proves to be ethically unacceptable.

This makes the ethical risks all the more significant. Humans need to become active so that DS do not simply happen, but that humans shape it. This is necessary so that DS will not be reduced to an instrument serving pure efficiency (Kirchschlaeger 2025c) but can rise to their ethical potential. More importantly, there is a need for normative guidance to review the economic self-interests that run DS so far almost exclusively and to guide calls for global regulations and governance in the digital domain and in the sphere of DS to ensure that all humans and the planet share the ethical benefits of DS and that the ethical risks of DS are mastered.

This is of essential significance not only because of, but especially in a time where multinational technology-companies who are the decisive driving-force of this technological progress officially and explicitly step down from their ethical responsibilities. For example, Meta announces that it will stop its ethically informed moderation of the content (e.g., fighting racist and sexist hate speech as well as incitements to violence) on their social media-platforms (Isaac and Schleifer 2025). Google, for example, removes their "Responsible AI Principles" which it has so far used to engage in technologies violating human rights as well as in technologies for surveillance and for causing harm (e.g., DS for weapons) (Tiku and De Vynck 2025). These are just additional pieces of explicit evidence (Zuboff 2019) that self-regulation of multinational technology companies will not really happen nor work.

5. Existing Global Governance Initiatives

The debate about the international governance of DS (Kirchschlaeger 2024a; 2024d; 2024e) encompasses several declarations, recommendations, principles, and guidelines—the first generation of governance initiatives: "the sermons." Different

initiatives by states and civil society on national level exist (Agency for Digital Government 2024; Amnesty International 2025; Association for Computing Machinery 2017; Australian Human Rights Commission 2019; Beijing Academy of Artificial Intelligence [BAAI] 2019; Commission Nationale de l'Informatique et des Libertés [CNIL] et al. 2018; Fairness, Accountability, and Transparency in Machine Learning [FAT/ML] 2016; Future of Life Institute 2017; House of Commons of the United Kingdom – Science and Technology Committee 2018; Japanese Society for Artificial Intelligence [JSAI] 2017; Montréal Declaration 2018; Partnership on AI 2024; The Public Voice Coalition 2018; Swiss Federal Council 2020; Villani 2018). On a regional level, states and civil society have contributed to this debate (European Group on Ethics in Science and New Technologies 2012; 2014; 2015; 2018a; 2018b; High-Level Expert Group on Artificial Intelligence HLEG AI of the European Commission 2019; Council of Europe 2018a; 2018b; 2018c; 2019a; 2019b; The Charter of Digital Rights by European Digital Rights (EDRi) - network, 2014). On an international level, contributions by state and non-state actors enter the spotlight (CNIL et al. 2018; Partnership on AI 2016; The Public Voice Coalition 2018; UNESCO World Commission on the Ethics of Scientific Knowledge and Technology [UNESCO COMEST] 2017; Information Technology Industry Council [ITIC] 2017; Dutton 1990; Price 2018; D 64 Zentrum für Digitalen Fortschritt 2018; UNICEF 2017; Association for Computing Machinery's Committee on Professional Ethics 2018; WeGovNow 2020; The Future of Privacy Forum 2018; UNI Global Union 2017; Institute of Electrical and Electro-nic Engineers [IEEE] Standards Association 2024; Internet Governance Forum [IGF] 2014; 2019; ISO 2020; the "Bletchley Declaration 2023" (AI Safety Summit 2023), the "AI4People Summit Declaration," the "G7 International Guiding Principles on AI," the "AI Code of Conduct," the "Recommendation on the Ethics of Artificial In-telligence" by UNESCO, the OECD Principles on Artificial Intelligence (OECD 2019), the G20-Principles of AI, "Declaration of Principles" by the World Summit on the Information Society in Geneva 2003 (World Summit on the Information Society [WSIS] 2003). Contributions from professional ethics—e.g., the code of the Institute of Electrical and Electronic Engineers (IEEE Standards Association 2024), the code of the National Society of Professional Engineers (NSPE 2019) or the codes of the American Society of Mechanical Engineers (ASME n.d.) emerge together with other proposals. The main challenge with these declarations, recommendations, principles, and guidelines—with "the sermons"—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" (Nevejans 2016, 26). Beyond that, the European Par-liament and Council reached political agreement on the European Union's Artificial Intelligence Act ("EU AI Act") (European Parliament 2023). 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 for fundamental rights by DS as well as encouraging investments and innovation in the field of DS. In addition, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, De-mocracy and the Rule of Law was adopted on May 17, 2024, by the Committee of Ministers of the Council of Europe (Council of Europe 2024).

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" (Bradford 2023, 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) (Kaspersen and Wallach 2023) and the model of the International Civil Aviation Organization (ICAO) (Baker & McKenzie 2023; World Economic Forum 2022).

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) (Baker & McKenzie 2023) 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" (Türk 2023).

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 (Baker & McKenzie 2023).

6. Ten Main Global Dangers of DS

From an ethical perspective, the first global danger of DS is not doing anything about the ethical risks of DS: Products with DS that violate human dignity and human rights or destroy the planet can be put on the market without legal consequences. For example, if an app is sold online that sexualizes images of children as mentioned above, nothing happens to the company offering this app except that it makes a lot of money (Lensa 2024; Heikkilä 2022; Snow 2022). For several decades now, humanity has allowed business models that have human rights violations at their core to go unchallenged.

A second global danger posed by DS is the growing inequality and poverty around the world due to a widening "digital divide" (Kirchschlaeger 2021). Expected explosive increases in computing power (e.g. through quantum computing) will explode the already existing "digital divide" in some regions of the world there is no access to the internet, other places already benefit from large data centers.

A third global danger is the negative impact of DS on the climate and the environment, which threatens to destroy the planet (Kirchschlaeger 2021). The massive demand for both raw materials for the production of DS and for energy in the production and use of DS is still disproportionate to the contribution of DS to combating climate destruction.

A fourth global danger is the constant violation of the human right to privacy and data protection (Zuboff 2019). Whenever possible, people's data is stolen and sold to the highest bidder. The ongoing disregard for privacy and data protection is a massive attack on the freedom of all people.

A fifth global danger is that with DS, thanks to these huge amounts of data about people, one can "know" people better than people know themselves. This opens the door to economic and political manipulation. With DS, you can know to use a metaphor exactly which piano keys to press to make the music play, in other words, to get people to buy or choose or vote the way the hands controlling DS want them to (Nosthoff and Maschewski 2019).

A sixth global danger is the potential for disinformation. Due to global networking online, millions of people can be targeted with false information - so-called "fake news" - at the click of a mouse. Language and image generation systems (such as ChatGPT, Dall-E), for which the criterion of truthfulness, data protection and copyright play no role and which, figuratively speaking, are nothing more than cows chewing the cud (Kirchschlaeger 2024c), open up new horizons in this respect, as artificially created audio and video material with a high degree of perfection can be circulated globally as supposed realities. For example, words can be put into the mouths of political officials that they never actually said. Unfortunately, despite official clarifications from the politicians concerned, some of the untruths remain in people's minds - with "fake news" in images even more than in words (Kirchschlaeger 2021).

Linked to this, a seventh global danger is the possibility of targeted destabilization of societies through hate speech and incitement to violence spread on the internet - especially on social media - where masses can be mobilized at the click of a mouse (Cadwalladr 2024).

An eighth global danger is the risk to the mental and physical health of people (Mahari and Pataranutaporn 2024), especially children and adolescents, due to the effects of DS, including "social media". Among other things, the WHO is considering recognizing "social media" use as an addiction (Montag 2023) because "social media" has the following traits: "The internet is a giant hypodermic, and the content, including social media like Meta, are the psychoactive drugs." (Richtel 2023) It would be more adequate to call "social media" "anti-social media".

A ninth global danger is cyberattacks (Kirchschlaeger 2021). Growing digitalization and an increase in the use of DS in people's business and private lives, as well as in the systemic infrastructure of societies, mean an increase in vulnerability to cyberattacks. Cybercrime and the fight against it have a massive economic impact on national economies.

The military applications of DS represent a tenth global danger (Kirchschlaeger 2024b). Automated weapon systems are unable to distinguish between combatants and non-combatants in armed conflicts and to apply the principle of proportionality (Asaro 2008), which leads to more misconduct and crime (Sharkey 2011). Another argument against automated weapons is that they would lead to more wars (Kahn 2017) because, among other things, a lower number of victims would be expected (Wallach 2015) and thus a lower political price would have to be paid (Wagner 2014).

Facing these ten global threats, the limitations of the so-far three generations of governance initiatives become obvious if one approaches this risky reality with the test questions mentioned above for the already existing "sermons", "locals", and "international actors". The already existing "sermons", "locals", and "international actors" receive also 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.

7. The Need for Human Rights-Based DS and an International Data-Based Systems Agency IDA at the UN

In order to secure a future for humanity and the planet, in order to allow humans and the planet to flourish sustainably, and in order to guarantee that human rights

are respected globally not only *offline* but also *online* – and in the digital sphere and the domain of DS – the following concrete measures need to be implemented with urgency:

a) Human Rights-Based Data-Based Systems (HRBDS): Human rights-based data-based systems (HRBDS) are meant to ensure that human rights serve as the basis of DS. In other words, HRBDS seek to ensure that human rights – including rights of the child (Kirchschlaeger 2025a) are respected, protected, implemented, and realized within the entire life cycle of DS and the complete value-chain process of DS (in the design, the development, the production, the distribution, the use, or the non-use of DS because of human rights-concerns). HRBDS strives for protecting the powerless from the powerful.

   As a minimum ethical frame of reference, human rights could provide the necessary normative orientation that they only represent a minimum standard that guarantees all humans physical survival and a life in dignity – a life as human. As an ethical minimum standard, human rights do not ethically overtax technological progress and do not stifle it, as would be the case with a high ethos. Another argument in favor of human rights is the ethical justifiability of human rights and their universality with the principle of vulnerability (Kirchschlaeger 2013b), which underpins their global acceptance. The latter proves to be unrivalled in comparison to other value systems and catalogs of norms (Kirchschlaeger 2025b), as human rights are not anchored in a specific culture, tradition, religion or worldview, but serve as a basis for the lived diversity of cultures, traditions, religions or worldviews by protecting, among other things, the freedom of belief, conscience and religion of all people (Kirchschlaeger 2020).

   In addition, human rights encourage and promote innovation and technological progress by protecting freedom of thought, freedom of expression, and free access to information (Kirchschlaeger 2013a). Finally, they protect and promote the diversity that is constitutive for innovation processes by guaranteeing the self-determination of every person, which proves to be a source of diversity (Kirchschlaeger 2016; Kirchschlaeger 2020).

b) An International Data-Based Systems Agency (IDA): An International Data-Based Systems Agency (IDA) urgently needs to be established at the UN as a global platform for technical cooperation in the field of DS (Kirchschlaeger 2021; 2024a; 2024d; 2024e), fostering human rights, safety, security, a rules-based international order, sustainability, and peaceful uses of DS with a multilateral approach and by collaborating and joining forces with other UN agencies and 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, as well as a global supervisory and monitoring institution and regulatory authority in the area of DS responsible for access to market approval.

Given the areas of convergence between DS and nuclear technologies, the In-

ternational Atomic Energy Agency (IAEA) model would seem the most appropriate one for responsible global AI governance as it represents an UN-agency with "teeth".

IDA at the UN should fulfill the following three key-functions:

a) Providing an access to market approval-process which several other industries know since decades (e.g., the pharmaceutical industry) in order to avoid harm of humans and the environment; the access to market approval-process orchestrated by IDA should ensure that human rights-violating products (like, e.g., the above-mentioned app sexualizing pictures of children) do not even end up on the market. By this, an preventive impact is caused by IDA that the private sector does not even design and develop such human rights-violating DS knowing that they will not pass the access to market approval-process.

b) Monitoring and enforcing that human rights are not violated with or by DS;

c) Fostering international technical collaboration in the sphere of DS in order to enable humanity to reach faster and better the positive potential of DS.

While pursuing all three key functions, IDA could also rely on DS-based solutions to implement HRBDS.

The establishment of an IDA is feasible because humanity has already shown that we are able to avoid "blindly" pursuing and implementing things that are technical possible, but that we are also able to exercise caution when the welfare of humanity and the planet are at stake. For example, humans researched the field of nuclear technology but then humans substantially and massively limited research and development in the field of nuclear technology, in order to prevent even worse consequences. This restriction was successful mainly due to an international regime, concrete enforcement mechanisms, and thanks to the International Atomic Energy Agency (IAEA) at the UN - an UN-agency wish "teeth".

8. HRBDS and IDA Are Realistic

The realization of IDA is realistic because humanity has shown in the past that when the well-being of people and the planet is at stake, it can focus on what is technically desirable instead of blindly pursuing everything that is technically feasible. Humanity has also stepped up to foster ethical upsides and to stop or to master the ethical downsides of technological progress by establishing a binding global regulatory framework as well as an impactful global enforcement-institution.

E.g., humanity has pursued nuclear technology, developed the atomic bomb and even used it more than once. However, in order to prevent even worse things from happening, mankind then massively restricted the research and development of nuclear technology despite overwhelming opposition from state and non-state actors. It is mainly thanks to international guidelines, concrete enforcement mechanisms and the UN's International Atomic Energy Agency (IAEA) that nothing worse happened.

E.g., in the case of chlorofluorocarbons (CFCs), humanity also decided to ban ozone-depleting substances under the Montreal Protocol of 1987 (UN Environment Programme 1987) and to enforce the ban consistently. Here, too, there was considerable resistance, partly due to the vested interests of the private sector. This regulation and its uncompromising enforcement have led to the hole in the ozone layer slowly closing.

9. Broad Global Support for HRBDS and IDA

There is a growing international and interdisciplinary network of experts calling for the establishment of HRBDS and IDA (IDA 2024). The Elders, an independent group of world leaders founded by Nelson Mandela, has endorsed the recommendations for human rights-based DS and for the establishment of a global agency to monitor them. It has called upon the UN to take appropriate action. In their statement of May 31, 2023, 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." (The Elders 2023)

The idea of a human rights-based and legally binding regulatory framework as well as the establishment of an institution enforcing the global regulation enjoyed the support of Pope Francis (Francesco 2024).

His Holiness the Dalai Lama supports the establishment of IDA at the UN as well as human rights-based DS (HRBDS) because human rights-based data-based systems (HRBDS) and the establishment of an International Data-Based Systems Agency (IDA) at the UN foster the realization of human rights, peace, and sustainability in the sphere of data-based systems (DS) (IDA 2025).

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." (Nichols 2023) He has called for a new UN body like an "International Data-Based Systems Agency IDA" to tackle threats posed by artificial intelligence in the UN Security Council on July 18, 2023 (Guterres 2023).

UN High Commissioner for Human Rights Volker Türk has demanded at the UN Human Rights Council "urgent action" and proposed human-rights-based HRBDS and a coordinated global response towards an institutional solution like the cre-

ation of an "International Data-Based Systems Agency IDA" in his statement about AI and human rights on July 12, 2023 (Türk 23).

Also, some voices from multinational technology-companies – among others, Sam Altman (Founder of OpenAI) at the World Economic Forum WEF 2024 – have called for IDA (Santelli 2024).

Mustafa Suleyman (CEO of Microsoft AI, Co-Founder and former Head of applied AI at DeepMind, Co-Founder of Inflection AI) says:

> "We need our generation's equivalent of the nuclear treaty to shape a common worldwide approach – in this case not curbing proliferation altogether but setting limits and building frameworks for management and mitigation that, like the wave, cross borders. This would put clear limits on what work is undertaken, mediate among national licensing efforts, and create a framework for reviewing both. Where there is a clear scope for a new body or bodies is with technical concerns. A dedicated regulator that navigates contentious geopolitics (as much as possible), avoids overreach, and performs pragmatic monitoring function on broadly objective criteria is urgently needed. Think of something like the International Atomic Energy Agency." (Suleyman and Bhaskar 2023, 266)

10. Concluding Remarks

Time is running. Not only because of the enormous ecological footprint of DS. Not only because of the constantly occurring human rights violations with or by DS. Not only because of the at large scale manipulated democratic opinion-forming- and decision-making processes. Not only because of entire societies being threatened by destabilization with a targeted use of "fake news". Not only because of leaders and decision-makers in politics, society, and business fearing that their careers are put to an end by a "deep-fake". But also because of the following reality, which significantly reinforces the ethically negative potential of DS:

> "Washington and Beijing will soon discover that putting A.I. in the hands of every person and robot on the planet will superempower bad people to levels no law enforcement agency has ever faced. Remember: Bad guys are always early adopters! And without the United States and China agreeing on a trust architecture to ensure that every A.I. device can be used only for humans' well-being, the artificial intelligence revolution is certain to produce superempowered thieves, scam artists, hackers, drug dealers, terrorists and misinformation warriors. They will destabilize both America and China, long before these two superpower nations get around to fighting a war with each other." (Friedman 2025)

If we foster now human rights-based data-based systems (HRBDS) and if we

establish now the International Data-Based Systems Agency (IDA) at the UN, there will be a sustainable future for the planet and for humanity enjoying human dignity and human rights.

References

Agency for Digital Government. 2024. The Danish National Strategy for Artificial Intelligence. Agency for Digital Government. https://en.digst.dk/strategy/the-danish-national-strategy-for-artificial-intelligence/(accessed 30. 9. 2025).

AI Safety Summit. 2023. The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023. GOV.UK, 1.–2. 11. 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 30. 9. 2025).

American Society of Mechanical Engineers (ASME). n. d. Standards. ASME. https://www.asme.org/codes-standards (accessed 30. 9. 2025).

Amnesty International. 2025. The Toronto Declaration: Protecting the right to equality in machine learning. Toronto Declaration. https://www.torontodeclaration.org/declaration-text/english/ (accessed 30. 9. 2025).

Asaro, Peter. 2008. How just could a robot warbe? In: A. Briggle Brey and K. Waelbers, eds. Current Issues in Computing and Philosophy, 50–64. Amsterdam: IOS Press.

Association for Computing Machinery US Public Policy Council. 2017. Statement on Algorithmic Transparency and Accountability. ACM, 12. 1. https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_. (accessed 30. 9. 2025).

Association for Computing Machinery's Committee on Professional Ethics. 2018. 2018 ACM Code of Ethics and Professional Conduct: Draft 3. ACM, 22. 6. https://ethics.acm.org/2018-code-draft-3/ (accessed 30. 9. 2025).

Atz, Kenneth, Leandro Cotos, Clemens Isert, Maria Håkansson, Dorota Focht, and Mattis Hilleke. 2024. Prospective de novo drug design with deep interactome learning. Nature communications 15, no. 1:3408. https://doi.org/10.1038/s41467-024-47613-w

Australian Human Rights Commission. 2019. Discussion Paper: Human Rights and Technology. Human rights Gov.au, 17. 12. https://humanrights. gov.au/our-work/technology-and-humanrights/publications/discussion-paper-human-rights-and-technology (accessed 30. 9. 2025).

Baker McKenzie. 2023. International: Can a global framework regulate AI Ethics? Baker McKenzie, 8. 11. https://insightplus.bakermckenzie.com/bm/investigations-compliance-ethics/international-can-a-global-framework-regulate-ai--ethics (accessed 30. 9. 2025).

Bathaee, Yavar. 2018. The Artificial Intelligence-Black Box and the Failure of Intent and Causation. Harvard Journal of Law & Technology 31, no. 2:890–938. Beijing Academy of Artificial Intelligence. 2019.

Beijing Artificial Intelligence Principles. International Research Center for AI Ethics and Governance. https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/(accessed 30. 9. 2025).

Bradford, Anu. 2023. Digital Empires: The Global Battle to Regulate Technology. Oxford: Oxford University Press.

Cadwalladr, Carole. 2024. "A polarisation engine": how social media has created a 'perfect storm' for UK's far-right riots. The Guardian, 3. 8. https://www.theguardian.com/media/article/2024/aug/03/a-polarisation-engine-how-social-media-has-created-a-perfect-storm-for-uks-far-right-riots (accessed 30. 9. 2025).

Castelvecchi, Davide. 2016. Can we open theblack box of AI? Nature 538, no.7623:20–23.https://doi.org/10.1038/538020a

Clifford, Catherine. 2018. Elon Musk: "Mark my words — A.I. is far more dangerous than nukes". CNBC, 13. 3. https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html (accessed 30. 9. 2025).

Coeckelbergh, Mark. 2020. AI ethics. Cambridge: MIT Press.

Commission Nationale de l'Informatique et des Libertés (CNIL), European Data Protection Supervisor, Garante per la protezione dei dati personali. (2018). Declaration on Ethics and Data Protection in Artificial Intelligence: 40th International Conference of Data Protection and Privacy Commissioners. Privacy Conference 2018, 23. 10. https://www.privacyconference2018.org/system/files/2018-10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf. (accessed 30. 9. 2025).

Council of Europe. 2018a. Algorithms and human rights – Study on the human rights dimensions of automated data processing techniques and possible regulatory implications. Council of Europe. https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5 (accessed 30. 9. 2025).

–––. 2018b. European ethical charter on the use of artificial intelligence (AI) in judicial systems and their environment. Council of Europe. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (accessed 30. 9. 2025).

–––. 2018c. Recommendation of the Committee of Ministers: Guidelines to respect, protect and fulfil the rights of the child in the digital environment. Council of Europe. 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--recommendationcmrec20187-of-the-committee-of-ministers.html (accessed 30. 9. 2025).

–––. 2019a. Committee of Ministers: Declaration on the manipulative capabilities of algorithmic processes. Council of Europe, 13. 2. https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b (accessed 30. 9. 2025).

–––. 2019b. 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. Council of Europe, 25. 1. https://rm.coe.int/guidelines-on-artificial-intelligence-and-data--protection/168091f9d8 (accessed 30. 9. 2025).

–––. 2024. Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Council of Europe, 5. 9. https://rm.coe.int/1680afae3c (accessed 30. 9. 2025).

D 64 Zentrum für Digitalen Fortschritt. 2018. Grundwerte in der Digitalisierten Gesellschaft: Der Einfluss Künstlicher Intelligenz auf Freiheit, Gerechtigkeit und Solidarität. D-64. https://d-64.org/wp-content/uploads/2018/11/D64-Grundwerte-KI (accessed 30. 9. 2025).

Demuth, Yves. 2018. Die unheimliche Macht der Algorithmen. Beobachter, 26. 4. https://www.beobachter.ch/gesellschaft/multimedia/die--unheimliche-macht-der-algorithmen-169811 (accessed 30. 9. 2025).

Dreyfus, Hubert L. 1972. What computers can't do: The limits of artificial intelligence. Revised edition. New York: Harper & Row Publishers

Dreyfus, Hubert L., and Stuart E. Dreyfus. 1986. Mind over machine: The power of human intuition and expertise in the era of the computer. New York: Free Press.

Dutton, I. R. 1990. Engineering code of ethics. IEEE Potentials 9, no. 4:30–31. https://doi.org/10.1109/45.65865

European Group on Ethics in Science and New Technologies. 2012. Ethics of information and communication technologies. Publications Office of the European Union, 22. 1. https://op.europa.eu/en/publication-detail/-/publication/c35a8ab5-a21d-41ff-b654-8cd6d41f6794/language-en/format-PDF/source-77404276 (accessed 30. 9. 2025).

European Group on Ethics in Science and New Technologies. 2014. Ethics of security and surveillance technologies. Publications Office of the European Union, 20. 5. https://op.europa.eu/en/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-54fc3225c969 (accessed 30. 9. 2025).

European Group on Ethics in Science and New Technologies. 2015. The ethical implications of new health technologies and citizen participation. Publications Office of the European Union, 13. 10. https://op.europa.eu/en/publication-detail/-/publication/e86c21fa--ef2f-11e5-8529-01aa75ed71a1/language-en (accessed 30. 9. 2025).

European Group on Ethics in Science and New Technologies. 2018a. Future of work, future of society. Publications Office of the European Union, 19. 12. https://op.europa.eu/en/publication-detail/-/publication/9ee4fad5-eef7-11e9--a32c01aa75ed71a1/language-en/format-PDF/source-314882261 (accessed 30. 9. 2025).

European Group on Ethics in Science and New Technologies. 2018b. Statement on artificial intelligence, robotics and 'autonomous' systems. Publications Office of the European Union, 9. 3. https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en (accessed 30. 9. 2025).

European Digital Rights. 2014. The Charter of Digital Rights: A guide for policy-makers. EDRI. https://edri.org/wp-content/uploads/2014/06/EDRi_DigitalRightsCharter_web.pdf (accessed 30. 9. 2025).

European Parliament. 2023. Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. News European Parliament, 9. 12. https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for--trustworthy-ai (accessed 30. 9. 2025).

Fairness, Accountability, and Transparency in Machine Learning. 2016. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. FAT/ML. https://www.fatml.org/resources/principles-for-accountable-algorithms. (accessed 30. 9. 2025).

Francesco. 2024. Artificial Intelligence and Peace. Message of his holiness Pope Francis for the 57th world day of peace. Vatican, 1. 1. https://www.vatican.va/content/francesco/en/messages/peace/documents/20231208-messaggio--57giornatamondiale-pace2024.html (accessed 30. 9. 2025).

Friedman, Thomas L. 2025. The One Danger That Should Unite the U.S. and China. The New York Times, 3. 9. https://www.nytimes.com/2025/09/02/opinion/ai-us-china.html (accessed 30. 9. 2025).

Future of Life Institute. 2017. Asilomar AI Principles. Future of Life Institute, 11. 8. https://futureoflife.org/open-letter/ai-principles/ (accessed 30. 9. 2025).

Gordon, Cindy. 2024. AI is accelerating the loss of our scarcest natural resource: water. Forbes, 25. 2. https://www.forbes.com/sites/cindygordon/2024/02/25/ai-is-accelerating-the-loss-of--our-scarcest-natural-resource-water/(accessed 30. 9. 2025).

Guterres, António. 2023. Secretary-General's remarks to the Security Council on Artificial Intelligence. United Nations, 18. 7. https://www.un.org/sg/en/content/sg/speeches/2023-07-18/secretary-generals-remarks--the-security-council-artificial-intelligence (accessed 30. 9. 2025).

Heikkilä, Melissa. 2022. The viral AI avatar app Lensa undressed me—without my consent. MIT Technology Review, 12. 12. https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me--without-my-consent (accessed 30. 9. 2025).

High-Level Expert Group on Artificial Intelligence [HLEG AI] of the European Commission. 2019. Ethics guidelines for trustworthy artificial intelligence. European Comission, 6. 5. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence (accessed 30. 9. 2025).

House of Commons of the United Kingdom: Science and Technology Committee. 2017. Algorithms in Decision-Making: Report No. 4, UK. Parliament, 23. 5. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf (accessed 30. 9. 2025).

IDA. 2024. Idaonline. https://idaonline.ch/ (accessed 30. 9. 2025).

IDA. 2025. His Holiness the Dalai Lama Supports Human Rights-Based Data-Based Systems (HRBDS) and the Establishment of IDA at the UN. Idaonline, 7. 3. https://idaonline.ch/2025/03/07/his-holiness-the-dalai-lama-supports-human-rights-based-data-based-systems-hrbds-and-the-establishment-of-ida-at--the-un/ (accessed 30. 9. 2025).

Information Technology Industry Council. 2017. AI Policy Principles. ITIC. www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf (accessed 30. 9. 2025).

Institute of Electrical and Electronic Engineers Standards Association. 2024. The IEEE Global Initiative 2.0 on Ethics of Autonomous and Intelligent Systems. IEEE, 25. 3. https://standards.ieee.org/industry-connections/activities/ieee-global-initiative/ (accessed 30. 9. 2025).

International Labour Organization. 2024. Mind the AI Divide: Shaping a Global Perspective on the Future of Work. ILO, 26. 7. https://www.ilo.org/publications/major-publications/mind-ai--divide-shaping-global-perspective-future--work (accessed 30. 9. 2025).

Internet Governance Forum. 2014. The Global Multistakeholder for Dialogue on Internet Governance Issues. Internet Governance Forum, 8. 4. http://intgovforum.org/cms/2014/IGFBrochure.pdf (accessed 30. 9. 2025).

Internet Governance Forum. 2019. Best Practice Forum on Internet of Things, Big Data, Artificial Intelligence. Internet Governance Forum. https://www.intgovforum.org/multilingual/filedepot_download/8398/1915 (accessed 30. 9. 2025).

Isaac, Mike, and Theodore Schleifer. 2025. Meta to end fact-checking program in shift ahead of Trump term. The New York Times, 7. 1. https://www.nytimes.com/2025/01/07/technology/meta-fact-checking-facebook.html (accessed 30. 9. 2025).

ISO. 2020. Information Technology—Electronic Discovery—Part 3: Code of Practice for Electronic Discovery. ISO. https://www.iso.org/standard/78648.html (accessed 30. 9. 2025).

Iversen, Gudmund R., and Mary Gergen. 1997. Statistics: The conceptual approach. Berlin: Springer

Japanese Society for Artificial Intelligence. 2017. The Japanese Society for Artificial Intelligence Ethical Guidelines. AI-GAKKAI. https://www.

ai-gakkai.or.jp/ai-elsi/wp-content/uploads/sites/19/2017/05/JSAI-Ethical-Guidelines-1.pdf (accessed 30. 9. 2025).

Kahn, Leonard. 2017. Military Robots and the Likelihood of Armed Combat. In: Patrick Lin, Ryan Jenkins, Keith Abney, eds. Robot ethics 2.0: From autonomous cars to artificial intelligence, 274–287. New York: Oxford University Press.

Kant, Immanuel. 1974. Kritik der praktischen Vernunft; Grundlegung zur Metaphysik der Sitten. Werkausgabe. Vol. 7. Berlin: Suhrkamp

Kaspersen, Anja, and Wendell Wallach. 2023. Envisioning Modalities for AI Governance: A Response from AIEI to the UN Tech Envoy. Carnegie Council, 29. 9. https://www.carnegiecouncil.org/media/article/envisioning-modalities-ai-governance-tech-envoy#gaio (accessed 30. 9. 2025).

Kharpal, Arjun. 2017. Stephen Hawking says A.I. could be "worst event in the history of our civilization". CNBC, 6. 11. https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html (accessed 30. 9. 2025).

Kirchschlaeger, Peter G. 2013a. Human Rights as Ethical Basis for Science. Journal of Law, Information and Science 22, no.2:1–17.

–––. 2013b. Wie können Menschenrechte begründet werden? Ein für religiöse und säkulare Menschenrechtskonzeptionen anschlussfähiger Ansatz. Münster: LIT Verlag.

–––. 2016. Menschenrechte und Religionen: Nichtstaatliche Akteure und ihr Verhältnis zu den Menschenrechten Paderborn: Schöningh.

–––. 2020. Human Dignity and Human Rights: Fostering and Protecting Pluralism and Particularity. Interdisciplinary Journal for Religion and Transformation in Contemporary Society 6, no. 1:90–106. https://doi.org/10.30965/23642807-00601006.

–––. 2021. Digital transformation and Ethics: Ethical Considerations on the Robotization and Automation of Society and the Economy and the use of Artificial Intelligence. Baden-Baden: Nomos.

–––. 2023. Ethical Decision-Making. Baden-Baden: Nomos.

–––. 2024a. An International Data-Based Systems Agency IDA: Striving for a Peaceful, Sustainable, and Human Rights-Based Future. Philosophies 9, no. 3:73. https://doi.org/10.3390/philosophies9030073

–––. 2024b. Automatisierte Waffensysteme: Eine ethische Analyse. Stratos: Militärwissenschaftliche Zeitschrift der Schweizer Armee 24:24–40.

–––. 2024c. Für eine menschenwürdige und nachhaltige Zukunft. Datenbasierte Systeme (DS) statt Künstliche Intelligenz (KI). In: Claudia Hamm, ed. Automatensprache 71(3/2023 & 1/2024), 156–168. München: Carl Hanser Verlag.

–––. 2024d. Securing a Peaceful, Sustainable, and Humane Future Through an International Data-Based Systems Agency (IDA) at the UN. Data & Policy 6:e78. https://doi.org/10.1017/dap.2024.38.

–––. 2024e. The Need for an International Data-Based Systems Agency (IDA) at the UN: Governing "AI" Globally by Keeping the Planet Sustainably and Protecting the Weaker from the Powerful. Journal of AI Humanities 18:213–248.

–––. 2025a. Artificial Intelligence – an Analysis from the Rights of the Child Perspective. Berkeley Journal of International Law, 27. 1. https://www.berkeleyjournalofinternationallaw.com/post/artificial-intelligence-an-analysis-from-the-rights-of-the-child-perspective (accessed 30. 9. 2025).

–––. 2025b. Human Rights and Religions: Non-State Actors and Their Correlation with Human Rights. London: Palgrave Macmillan.

–––. 2025c. Ethics and Digital Transformation of Human Work: The Society-, Entrepreneurship-, Research-Time Model (SERT). London: Palgrave Macmillan.

Knight, Will. 2017a. The Dark Secret at the Heart of AI. MIT Technology Review, 11. 4. https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/ (accessed 30. 9. 2025).

–––. 2017b. The Financial World Wants to Open AI's Black Boxes. MIT Technology Review, 13. 4. https://www.technologyreview.com/2017/04/13/152590/the-financial-world-wants-to-open-ais-black-boxes/ (accessed 30. 9. 2025).

Lensa. 2024. Lensa AI. https://lensa.app (accessed 30. 9. 2025).

Mahari, Robert, and Pat Pataranutaporn. 2024. We need to prepare for "addictive intelligence". MIT Technology Review, 5. 8. https://www.technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/ (accessed 30. 9. 2025).

Manzeschke, Arne. 2019. Roboter in der Pflege: Von Menschen, Maschinen und anderen hilfreichen Wesen. EthikJournal 5, no. 1:1–11.

Misselhorn, Catrin. 2018. Maschinen mit Moral? Grundfragen der Maschinenethik. 5th edition. Ditzingen: Reclam

Montag, Christian. 2023. Was verändern Smart Phones in unserem Denkapparat? ISE-Ringvorlesung "Ethik der digitalen Transformation. Aktuelle Fragen". Lecture. Universität Luzern, 19. 10.

Montréal Declaration. 2018. Montréal Declaration for a Responsible Development of Artificial Intelligence. Déclaration de Montréal IA responsable. https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/04/UdeM_Decl-IA-Resp_LA-Declaration-ENG_WEB_09-07-19.pdf (accessd 30. 9. 2025).

Nevejans, Nathalie. 2016. European Civil Law Rules in Robotics: Study for the JURI Committee. European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf (accessed 30. 9. 2025).

Nichols, Michelle. 2023. UN chief backs idea of global AI watchdog like nuclear agency. Reuters Media, 12. 6. https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/ (accessed 30. 9. 2025).

Nosthoff, Anna-Verena, and Felix Maschewski. 2019. Die Gesellschaft der Wearables: Digitale Verführung und soziale Kontrolle. Berlin: Nicolai Publishing & Intelligence GmbH.

NSPE. 2019. Code of ethics for engineers. NSPE. https://www.nspe.org/resources/ethics/code-ethics (accessed 30. 9. 2025).

OECD. 2019. Recommendation of the Council on artificial intelligence. OECD Legal Instruments. https://legalinstruments.oecd.org/api/print?ids=648&lang=en (accessed 30. 9. 2025).

Partnership on AI. 2016. Our Tenets. https://partnershiponai.org/about/#tenets (accessed 30. 9. 2025).

Price, Antoinette. 2018. First International Standards committee for entire AI ecosystem: Industry recognizes standartization will be essential to broad adoption of AI. IEC e-tech, 15. 5. https://etech.iec.ch/issue/2018-03/first-international-standards-committee-for-entire-ai-ecosystem (accessed 30. 9. 2025).

Ren, Shaolei, and OECD. 2023. How much water does AI consume? The public deserves to know. OECD, 30. 11. https://oecd.ai/en/wonk/how-much-water-does-ai-consume (accessed 30. 9. 2025).

Ren, Shaolei, and Adam Wierman. 2024. The Uneven Distribution of AI's Environmental Impacts. Harvard Business Review, 15. 7. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts (accessed 30. 9. 2025).

Richtel, Matt. 2023. Is Social Media Addictive? Here's What the Science Says. The New York Times, 25. 10. https://www.nytimes.com/2023/10/25/health/social-media-addiction.html (accessed 30. 9. 2025).

Santelli, Filippo. 2024. Sam Altman: "In pochi anni l'IA sarà inarrestabile, serve un'agenzia come per l'energia atomica". La Repubblica, 18. 1. 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 30. 9. 2025).

Sharkey, Noel. 2011. The Automation and Proliferation of Military Drones and the Protection of Civilians. Law, Innovation and Technology 3, no. 2:229–240. https://doi.org/10.5235/175799611798204914

Snow, Olivia. 2022. 'Magic Avatar' App Lensa Generated Nudes From My Childhood Photos. WIRED, 7. 12. https://www.wired.com/story/lensa-artificial-intelligence-csem/?bxid=5cc9e15efc942d13eb203f10 (accessed 30. 9. 2025).

Suleyman, Mustafa and Michael Bhaskar. 2023. The Coming Wave Technology, Power, and the Twenty-first Century's Greatest Dilemma. London: The Bodley Head.

Swiss Federal Council. 2020. Leitlinien Künstliche Intelligenz für den Bund: Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung. Swiss Federal Council. https://www.sbfi.admin.ch/dam/de/sd-web/--uy97trD1VMc/Leitlinien%2520K%C3%BCnstliche%2520Intelligenz%2520-%2520DE%5B1%5D.pdf (accessed 30. 9. 2025).

The Elders. 2023. The Elders urge global co-operation to manage risks and share benefits of AI. The Elders, 31. 5. https://theelders.org/news/elders-urge-global-co-operation-manage-risks-and-share-benefits-ai (accessed 30. 9. 2025).

The Future of Privacy Forum. 2018. Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models. Future of Privacy Forum. https://fpf.org/wp-content/uploads/2018/06/Beyond-Explainability.pdf (accessed 30. 9. 2025).

The Public Voice Coalition. 2018. Universal Guidelines for Artificial Intelligence. EPIC, 23. 10. https://archive.epic.org/international/AIGuidleinesDRAFT20180910.pdf (accessed 30. 9. 2025).

Tiku, Nitasha, and Gerrit De Vynck. 2025. Google drops pledge not to use AI for weapons or surveillance. The Washington Post, 4. 2. https://www.washingtonpost.com/technology/2025/02/04/google-ai-policies-weapons-harm/ (accessed 30. 9. 2025).

Türk, Volker. 2023. Artificial intelligence must be grounded in human rights, says High Commissioner: Statement delivered by Volker Türk, UN High Commissioner for Human Rights, delivered at high level side event of the 53rd session of the Human Rights Council. OHCHR, 12. 7. https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be--grounded-human-rights-says-high-commissioner (accessed 30. 9. 2025).

Turkle, Sherry. 2007. Authenticity in the age of digital companions. Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems 8, no. 3:501–517. https://doi.org/10.1075/is.8.3.11tur

UN Environment Programme. 1987. Montreal Protocol on substances that deplete the ozone layer. United Nations. https://ozone.unep.org/treaties/montreal-protocol (accessed 30. 9. 2025).

UN General Assembly. 2024. Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development. United Nations, 11. 3. https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/78/L.49&Lang=E (accessed 30. 9. 2024).

UN Human Rights Council. 2023. Resolution adopted by the Human Rights Council on 14 July 2023: 53/29. New and emerging digital technologies and human rights. United Nations, 18. 7. https://undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2FRES%2F53%2F29&Language=E&DeviceType=Desktop&LangRequested=False (accessed 30. 9. 2025).

UNESCO. 2024. Ethics of Artificial Intelligence: The Recommendation. UNESCO. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics (accessed 30. 9. 2025).

UNESCO World Commission on the Ethics of Scientific Knowledge and Technology. 2017. Report of COMEST on Robotic Ethics. UNESCO, 14. 9. https://unesdoc.unesco.org/images/0025/002539/253952E.pdf (accessed 30. 9. 2025).

UNI Global Union. 2017. Top 10 Principles für Ethical Artificial Intelligence. The future world of work. www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (accessed 30. 9. 2025).

UNICEF. 2017. The State of the world's children 2017: Children in a digital world. New York: UNICEF. https://www.unicef.org/publications/index_101992.html. (accessed 30. 9. 2025).

United Nations. 2024. Summit of the Future 2024. United Nations. https://www.un.org/en/summit-of-the-future (accessed 30. 9. 2025).

Verma, Pranshu, and Cat Zakrzewski. 2024. AI deepfakes threaten to upend global elections. No one can stop them. The Washington Post, 23. 4. https://www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india/ (accessed 30. 9. 2025).

Villani, Cédric. 2018. For a Meaningful Artificial Intelligence: Towards a French and European Strategy. Privacy Conference, 23. 10. https://www.privacyconference2018.org/system/files/2018-10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf (accessed 30. 9. 2018).

Virt, Günter. 2007. Damit Menschsein Zukunft hat. In: Gerhard Marschütz and Gunter Prüller-Jagenteufelm, eds. Theologische Ethik im Einsatz für eine humane Gesellschaft. Würzburg: Echter.

Wagner, Markus. 2014. The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems. Vanderbild Journal of Transnational Law 47, no. 5:1371–1424.

Wallach, Wendell. 2015. A dangerous master: How to keep technology from slipping beyond our control. New York: Basic Books.

WeGovNow. 2020. Towards #WeGovernment: Collective and participative approaches for addressing local policy challenges. WeGovNow. https://wegovnow.eu// (accessed 30. 9. 2025).

World Economic Forum. 2022. How can the aviation industry make AI safer? Weforum, 25. 8. https://www.weforum.org/agenda/2022/08/how-can-aviation-industry-make-ai-safer/ (accessed 30. 9. 2025).

World Summit on the Information Society. 2003. Declaration of Principles: Building the Information Society: a global challenge in the new Millennium (WSIS-03/GENEVA/DOC/4-E). ITU, 12. 12. https://www.itu.int/net/wsis/docs/geneva/official/dop.html (accessed 30. 9. 2025).

Zhang, Xinyi, and Saradha Venkatachalapathy, and Daniel Paysan, and Paulina Schaerer, and Claudio Tripodo, and Caroline Uhler, and G. V. Shivashankar. 2024. Unsupervised representation learning of chromatin images identifies changes in cell state and tissue organization in DCIS. Nature communications 15, no. 1:6112. https://doi.org/10.1038/s41467-024-50285-1

Zuboff, Shoshana. 2019. The age of surveillance capitalism: The fight for a human future at the new frontier of power. First edition. New York: PublicAffairs.