GLOBAL TRENDS ANALYSIS

Amandeep Singh Gill

Aligning AI Governance Globally: Lessons from current practice

03 **2021**

"Dr. Frankenstein's crime was not that he invented a creature through some combination of hubris and high technology, but rather that he abandoned the creature to itself." (Latour 2012)

Imprint

Published by
Stiftung Entwicklung und Frieden/
Development and Peace Foundation (sef:)
Dechenstr. 2, 53115 Bonn, Germany
Bonn 2021

ISSN: 2568-8804

INTRODUCTION

Considering data and artificial intelligence (AI) as global commons could be crucial in ensuring that these key technologies of the 21st century benefit all of humanity. However, fragmented efforts of AI development and governance across the world risk diluting the effectiveness of a global commons approach. Recent experience of AI governance in the three domains of armed conflict, education and health clearly shows that the AI governance problem goes above and beyond unbiased data, transparency and explainability of algorithms. Systems thinking is needed to identify the limitations and trade-offs at each stage of AI development, reinforce human responsibility and accountability, and provide for post-deployment feedback into policy. Public officials need this lifecycle understanding of AI as well as new tools to audit AI systems nationally on an ongoing basis. Globally, a commons approach, shared vocabulary and values, benchmarks and digital public infrastructures can be powerful ways to align diverse approaches to AI governance and incentivise compliance.

FIGURE 1

NATIONAL AI POLICIES AND INITIATIVES (AS OF NOVEMBER 2021)

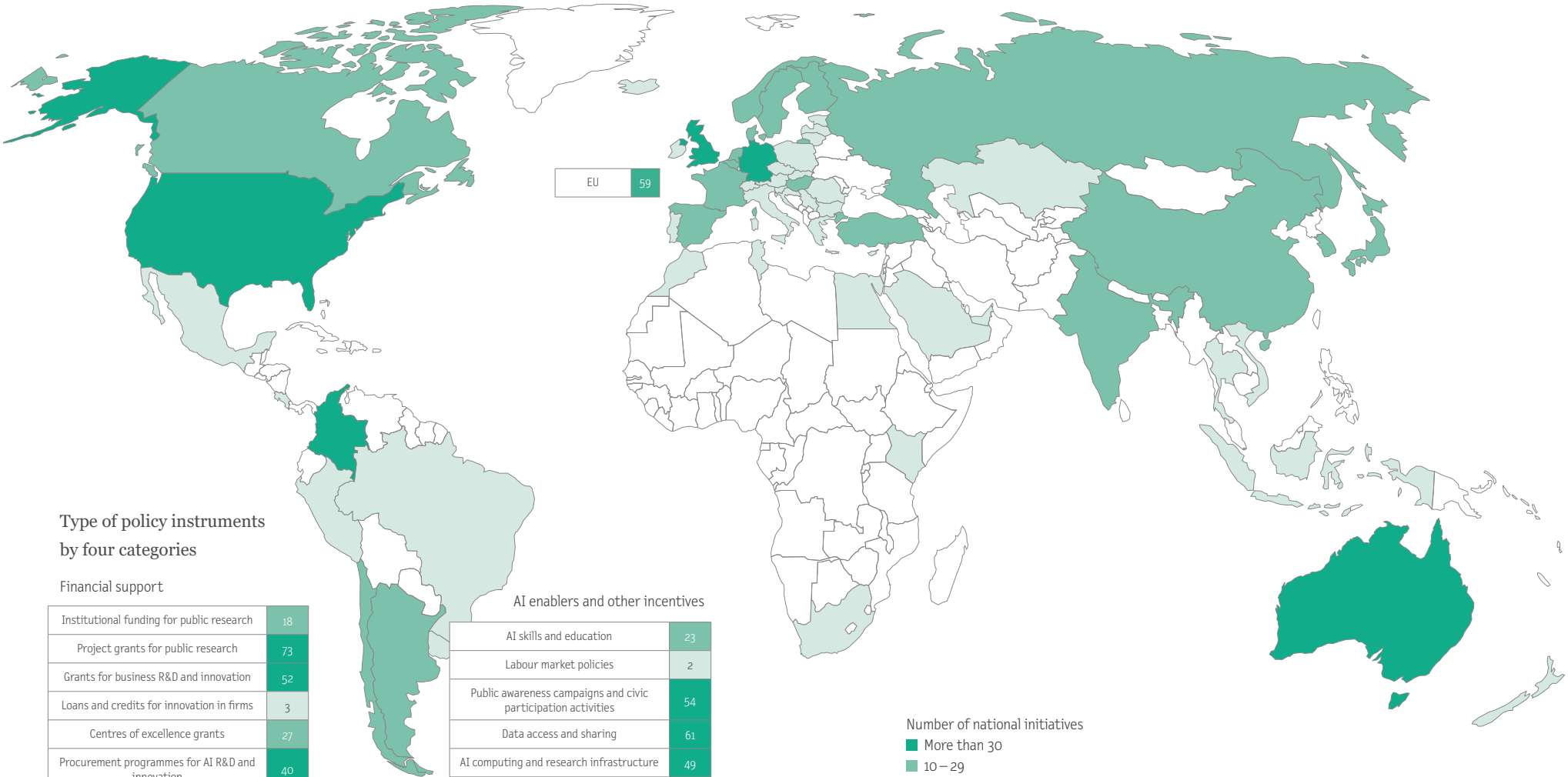
Number of initiatives per country

Governance

National strategies, agendas and plans	238
AI coordination and/or monitoring bodies	37
Public consultations of stakeholders or experts	135
AI use in the public sector	69

Guidance and regulation

Emerging AI-related regulation	163
Regulatory oversight and ethical advice bodies	56
Labour mobility regulation and incentives	11
Standards and certification for technology development and adoption	19



Type of policy instruments by four categories

Financial support

Institutional funding for public research	18
Project grants for public research	73
Grants for business R&D and innovation	52
Loans and credits for innovation in firms	3
Centres of excellence grants	27
Procurement programmes for AI R&D and innovation	40
Fellowships and postgraduate loans and scholarships	31
Equity financing	8
Indirect financial support	9

AI enablers and other incentives

AI skills and education	23
Labour market policies	2
Public awareness campaigns and civic participation activities	54
Data access and sharing	61
AI computing and research infrastructure	49
Networking and collaborative platforms	116
Knowledge transfers and business advisory services	21
Science and innovation challenges, prizes and awards	21

Source: https://oecd.ai/en/dashboards, 29.11.2021

1. UNDERSTANDING ARTIFICIAL INTELLIGENCE

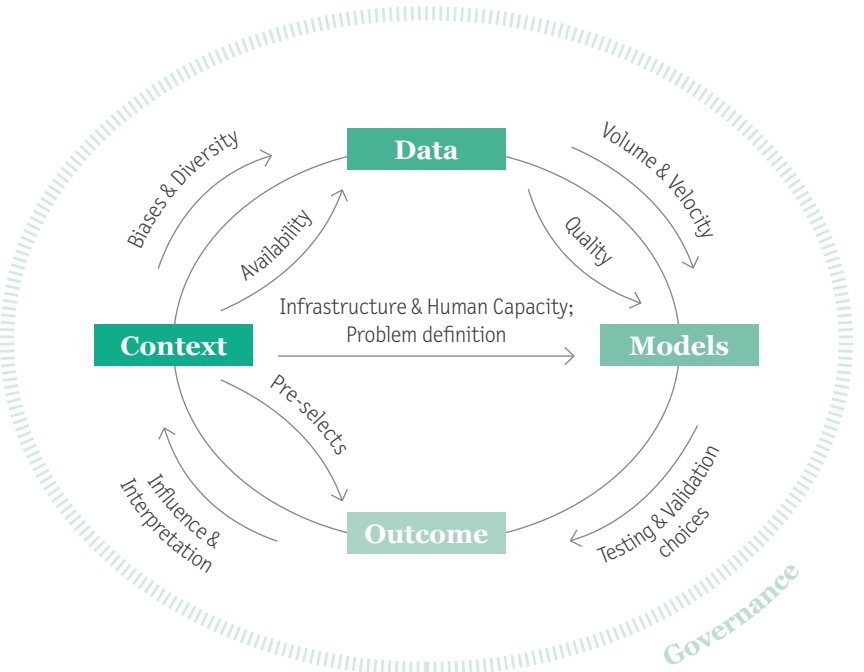
Traditional computer software uses quantitative measurements (data) and a set of rules (code) to reach an output. It is thus rules-based and deterministic. The data fits or it does not even though statistical methods allow for probabilistic assessments. In case of AI, the desired output becomes an input alongside the quantitative inputs (training data). This compares much to a baker using a bit of starter to raise dough. This inversion underlines an important dimension of the understanding of AI, namely that AI models are subjectively determined. They reflect human choices about the kind of world humans desire and are not artefacts determined solely by objective measurements. Once the desired outcome(s) is paired with data relevant to those outcomes – another step where human subjective choices are determinant – we get a model (code) as an output. This is our AI algorithm. We can feed new data inputs to this model, recognise patterns and predict with a certain degree of confidence whether the desired outcome will happen [see Figure 3].

Unlike old fashioned software, therefore, AI is non-deterministic and predictive. Crudely put, AI models are like the cognitive models of the world all humans build, which they then pair with incoming data to decide what actions to take. In Wartofsky's words, all cognitive artefacts we create "are models: representations to ourselves of what we do, of what we want, and of what we hope for. The model is not, therefore, simply a reflection or copy of some state of affairs, but beyond this, a putative mode of action" (Wartofsky 1979, p. xv).

This brings us to the second important aspect of understanding AI. Even though AI models are learning systems with data and desired outcomes as the teachers, they do not (yet) learn like humans do nor are they intelligent in the manner that humans are. Yes, AI can imitate some human cognitive functions such as perception very well but it is extremely narrow in its 'intelligence'. Many routine aspects of human intelligence such as creative and collaborative intelligence are beyond its reach. Further, despite more than hundred years since Spearman's two-factor model of intelligence, and despite incredible progress in understanding cognitive function, neuro-morphology and neurobiological mechanisms, we know very little about the famous 'g', general intelligence with its incredible variations across humans (Barbey 2018). To state the obvious, when we do not even know what human intelligence is, how can we claim that some human artefact approximates to it?

The third aspect of the proper understanding of AI is the interplay between context, data, models and outcomes [see Figure 2]. Context is what determines the problem to be solved through AI. Further it is crucial to the understanding of what types of data is relevant to that problem and whether and how we can access it. It thus sets limits to the depth and diversity of the data input. The way in which the problem is contextualised impacts the validity and reliability of the model. Same is true for the quality and representativeness of data. The models also reflect mathematical and computing ingenuity as well as choices and constraints around human resources and computing infrastructure. They have to be tested or in other words 'verified' (for the implemented model's fit with the conceptual model) and 'validated' (for implemented model outcomes to reflect the real-world problem that we set out to solve). The outcomes loop back into context. Without proper statistical understanding in context even the best model outcomes can be

FIGURE 2
AI and the interplay between context, data, models and outcomes

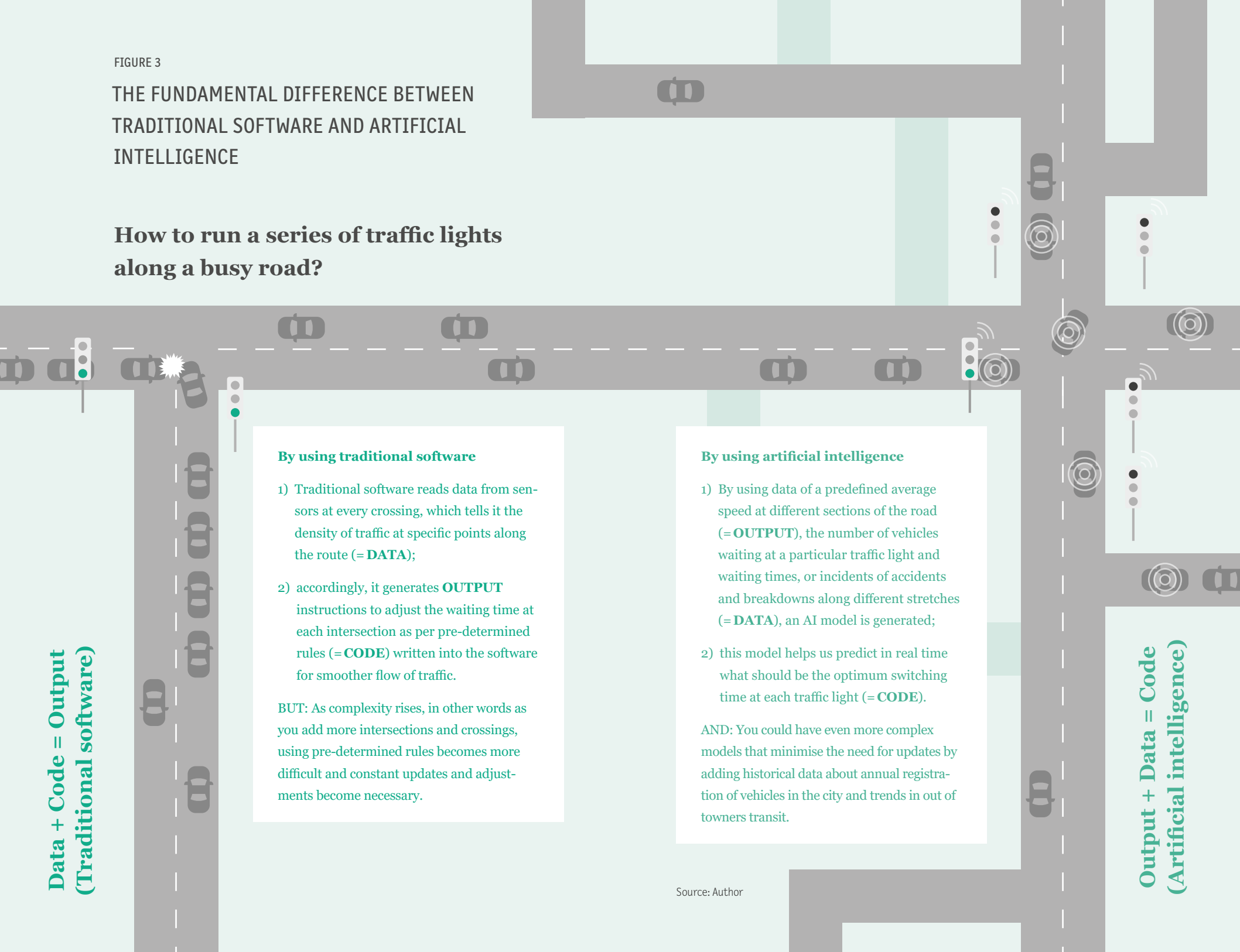


Source: Author

FIGURE 3

THE FUNDAMENTAL DIFFERENCE BETWEEN TRADITIONAL SOFTWARE AND ARTIFICIAL INTELLIGENCE

How to run a series of traffic lights along a busy road?



Data + Code = Output (Traditional software)

By using traditional software

1) Traditional software reads data from sensors at every crossing, which tells it the density of traffic at specific points along the route (= **DATA**);

2) accordingly, it generates **OUTPUT** instructions to adjust the waiting time at each intersection as per pre-determined rules (= **CODE**) written into the software for smoother flow of traffic.

BUT: As complexity rises, in other words as you add more intersections and crossings, using pre-determined rules becomes more difficult and constant updates and adjustments become necessary.

By using artificial intelligence

1) By using data of a predefined average speed at different sections of the road (= **OUTPUT**), the number of vehicles waiting at a particular traffic light and waiting times, or incidents of accidents and breakdowns along different stretches (= **DATA**), an AI model is generated;

2) this model helps us predict in real time what should be the optimum switching time at each traffic light (= **CODE**).

AND: You could have even more complex models that minimise the need for updates by adding historical data about annual registration of vehicles in the city and trends in out of towners transit.

Output + Data = Code (Artificial intelligence)

Source: Author

misinterpreted or misapplied. Importantly, each of these four dimensions can be shaped by governance to different degrees. Therefore, governance needs to be envisioned as an algorithmic systems intervention.

The first step to AI governance is thus proper understanding. It is best to think of AI as 'AI systems', which encompass problem definition, data models, data collection and curation, algorithms, testing and validation, impact and post-use assessments. AI's fundamental character is augmentation of human capabilities and not their emulation. Human intelligence, which we still understand poorly, certainly cannot be reduced to data flows in bits and bytes or confined to structures in a particular organ of the body. It is analytical but it is also emotional and ethical, somatic and haptic, and above all collective. With this understanding, we can proceed to look at AI's promise and limitations.

2. AI'S POTENTIAL AND ITS RISKS

AI offers significant advantages compared to traditional computing and data analysis. It can factor in more relationships and non-linearities compared with traditional statistical methods or computing tools. Instead of deriving insights only from historical data, AI can use real-time or future data to predict outcomes, based on learning from past data. Further, AI can handle diverse data such as numbers, images, videos and unstructured text more easily than traditional computing. This is a huge advantage in today's internet environment with myriad ways to engage users and to connect devices. Finally, AI systems offer more agility and opportunities to experiment than traditional software. In sum, it confers more data-driven problem-solving power on humans. In a world awash with data, AI helps humans avoid information overload and entrust routine decision making and predictions to machines.

There is another powerful socio-economic aspect of AI which is underappreciated. AI lowers entry barriers to expert domains. Simply by accessing historical data from oil wells and plugging it into smart AI models, people with little operational expertise of the oil sector can confidently make predictions about optimisation of oil production (Koroteev/Tekic 2021). This disrupts traditional ways of doing business, merges existing economic domains by drilling horizontally across them, and creates new growth opportunities. This is equally true of social problem solving in the context of national and international development. In recent years, a range of new actors have come into the Sustainable Development Goals (SDGs) arena with AI-based solutions for longstanding challenges in agriculture, health, education and the environment. The economic impact alone is expected to be significant and estimates of additional global GDP by 2030 vary from less than US$ 1.5tn to more than US$ 15tn (Szczepański 2019).

With this potential come significant risks. It is this Janus-faced nature of AI which dictates the need for its governance. Let us look at the pitfalls of AI with examples from three specific domains: war, education and health.

2.1 THE CASE OF ARMED CONFLICT

Technology embedded in weapons and related systems has been a crucial determinant of national security. Digital technologies are a feature of almost all modern weapon systems. Cyber weapons have risen in prominence as a separate class of weapons in the grey zone between peace and war, and between state and non-state actors. In the last few years, war planners have also started to plan the use of AI for a variety of functions including intelligence, training, defence and offence (Horowitz et al. 2018). This growing interest is natural, given the advantage AI offers with regard to the handling of vast amounts of data, fusion across different sensors and platforms, and augmenting the speed and accuracy of the human response in fast-moving multi-dimensional battlespace. However, weapon systems laced with AI create new challenges for human control and accountability for the use of lethal force. Concern has grown over the so-called lethal autonomous weapons systems or LAWS, the 'killer robots' and 'terminators' of Hollywood fame. This concern is partially legalistic: LAWS might escape the remit of the laws of war by occulting human accountability for the use of force in accordance with accepted principles of International Humanitarian Law. Examples are distinction (between civilian and military targets), proportionality and precaution. The concern is also ethical: Endowing machines with the ability to make life and death decisions militates against long-held notions of human dignity and agency. Finally, it is about international security: AI can create new asymmetries of power between technologically advanced countries and others, lower the barrier to the use of force and introduces uncertainty in unstable and contested regions (Gill 2019).

2.2 THE CASE OF EDUCATION

AI offers exciting opportunities for personalised learning and for shifting the focus from chalk and talk to more learner-centred approaches. AI can also facilitate the shift from classrooms to micro-learning moments on the go and from textbooks to Open Educational Resources (OERs) that can be remixed, reused, revised, redistributed, and retained digitally. While robots are unlikely to replace teachers anytime soon, AI-based virtual assistants are likely in the decades ahead. They can help teachers mark homework and track the progress of students individually. There are numerous risks, however, in ceding control over learning to machines. AI designed with the best intentions could end up disempowering teachers and students alike. It could multiply digital distraction further and entrench superficial thinking and loneliness. Personalised coaching could create new divides of haves and have-nots. Deploying virtual teaching assistants would not be possible without loads of training data and intrusive tracking of teachers and students. There are uncomfortable questions about data privacy and who owns children's data. Massive use of AI in education could tip us further into a stressed and surveilled society.

2.3 THE CASE OF HEALTH

Science and technology have been critical to progress on personal and public health. AI offers a paradigm-shifting opportunity to reinvent delivery of health services, reduce costs, personalise diagnosis and treatment, and to transform patient-doctor communication (Topol 2019). For instance, using 20 years of longitudinal patient data from 4.5 mn patients, cross-linked Electronic Medical Records and claims, the Clalit Research Institute in Israel has succeeded in predicting risks of renal failure among diabetics five years in advance (Balicer 2018). BloodCounts!, a Cambridge University based network, applies AI to analyse all data points from routinely performed Complete Blood Count (CBC) tests. They turn them into a broad surveillance network to detect infectious disease outbreaks without the need for any new instruments or reagents. The value of such advance warning to public health officials is inestimable in light of the COVID-19 pandemic. In resource poor settings, AI offers leapfrogging opportunities for universal health care (Wahl et al. 2018).

On the flipside, there are concerns on health data security, ownership, privacy and informed consent. Datasets used to develop AI for health can be biased and non-representative resulting in problematic outcomes: One study with two large cohorts showed that black patients had nearly three times the frequency of occult hypoxemia that was not detected by pulse oximetry as white patients (Sjoding et al. 2020). Tech 'solutionism' without sufficient regard for context and health worker engagement remains widespread. There is a lot of hype with claims about solutions that do not perform well in the real world and thus have not earned the trust of clinicians. For instance, while telemedicine surged with the COVID-19 pandemic as patients were confined to their homes, there were missteps with contact tracing apps, which failed to earn the trust of citizens (Lewis 2020). Again an AI-based outbreak monitoring platform was one of the first to alert the world on reports of the novel flu like disease in China (Niiler 2020). Nonetheless subsequent AI based predictions of disease spread and health system burden were of uneven quality and impact. Finally, there is concern that digital health and AI might only benefit certain sections of global society and further entrench existing imbalances in healthcare (Kickbusch et al. 2021).

3. THE PURPOSE(S) OF AI GOVERNANCE

The risk analysis, using three examples of AI applications, underlines the critical importance of AI governance. This is now broadly recognised and there are a variety of risk informed AI governance initiatives. A growing number of countries are adopting AI or related policies, like the EU's proposed regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (European Commission 2021b). Before we examine the specific aspects of these governance initiatives, it is worth pausing to reflect on the purpose(s) we should assign to AI governance as societies, and as the international community.

Let us take a concrete example. Sometimes back automatic faucets ran into a problem. They did not work that well with darker skin. The reason is that their sensors were trained with a limited set of data from people with white skin. It is not that the developers were racist or biased, they were simply blind to the lack of diversity in their datasets. This can be annoying. However, in another context the same underlying problem can be deadly. Pulse oximeters have been used during the Covid-19 pandemic to monitor levels of oxygen in infected people and decide whether to put them on oxygen at home or move them to an Intensive Care Unit in a hospital. For a variety of reasons,

they are likely to misdiagnose the level of oxygen in the blood three times for blacks compared to whites (Sjoding et al. 2020). Thus, a foundational purpose for the governance effort has to be prevention of harm. A degree of consumer protection is also implicit in harm prevention to protect naïve users against misleading claims, say of health benefits from AI applications.

Another discrete purpose for AI governance is the prevention of misuse of data and ensuring fair value to data owners for the use of their data. This implies not only protection of personal or public data against privacy and security breaches but also user agency over how and by whom their data is used. Our data is used by companies to sell us goods or services often without our full knowledge and consent. Social media posts can be used to profile individuals and target or manipulate them for political ends. AI based predictions of the risk of loan default based on personal data can be used to present a customer with options for a loan. In some scenarios, this could deny the customer a fair opportunity to be considered for a loan. Again, pharmacies have lots of what is called 'supply chain data', who buys what medicines, from whom, at what price and for which reason. This data has valuable insights hidden in it, for instance, on drug reactions or disease outbreaks (Bacry/Gaïffas 2020). Companies monetise this data but its value does not necessarily flow back to patients. This is not fair, particularly as the downside of leaked data is theirs to bear. Preventing misuse and promoting fair use of data is therefore an important part of the mission statement for governance.

A third discrete purpose is captured by 'missed use' and 'missing' data (Gill/Germann 2021). Let us look at 'missed use' with the example of the CBC machines that are used for 3.6 bn haemograms every year (University of Cambridge 2021). Data is generated but only a part of it is used by the prescribing physicians. The rest which is wiped out clean from the hard disks needs to be 'rescued' through regulatory facilitation. It could contain valuable insights about pandemic outbreaks or anti-microbial resistance. Likewise, emissions data related to different catalytic converters in cars could be shared for more optimal choices across the automobile industry. But commercial and proprietary considerations prevent the collaborative use of this data. There is plenty of such data particularly in high income settings that is sitting in silos or under commercial/public sector control and simply cannot be used for lack of enabling policy and regulations. Society at large bears an opportunity cost for the missed use of this data.

Then there is simply 'missing data' in both high income and low-resource settings, for instance disability faced by young people from COVID-related conditions (Briggs/Vassall 2021). What we do not count or cannot count has consequences for what we can or cannot do with AI. If policy and governance do not facilitate this redressal of data poverty and analysis for the larger good, an opportunity for using data for the public good is lost.

4. THE PRACTICE OF AI GOVERNANCE TODAY

We have seen in the previous two sections that AI governance is important and urgent for various reasons. As AI adoption has grown, several general governance initiatives have emerged alongside responses tailored to specific domains such as security, education and health. How do these governance initiatives contribute in general to the purpose of preventing harm and misuse of AI on the one hand, and preventing its missed use on the other? And how can the practice of AI governance meet the specific requirements of deployment in fields such as health?

Of a myriad of initiatives on general principles and codes of conduct, a few stand out. At the international level, the OECD's AI Principles, subsequently endorsed by the G20 at the Osaka Summit in June 2019, articulate five mutually-reinforcing values-based guidelines for responsible and trustworthy stewardship of AI [see Box] (OECD 2019).

The June 2019 Report of the UN Secretary-General's High-level Panel in its Recommendation 3C called for autonomous intelligence systems to be designed in ways that enable their decisions to be explained and for humans to be accountable for their use (UN Secretary-General's High-level Panel on Digital Cooperation 2019, p. 5). The Report called for practice to follow precept through audit and certification schemes. This is meant to monitor compliance of AI systems with engineering and ethical standards. These standards and principles such as transparency and non-bias to be developed through multi-stakeholder and multilateral approaches should be applicable in different social settings.

More recently, a Recommendation on the Ethics of Artificial Intelligence has been produced by a UN Educational, Social and Cultural Organization (UNESCO) Ad Hoc Expert Group (AHEG) of 24 global experts. Its text has been examined by Member States and adopted at the November 2021 Gener-

THE OECD AI PRINCIPLES

- AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.
- AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
- There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
- AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
- Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

Source: https://www.oecd.org/digital/artificial-intelligence/ai-principles/

al Conference (UNESCO 2021). Apart from a comprehensive set of principles, the text contains guidance on applicability of those principles to eleven policy areas and a few recommendations on monitoring and evaluation.

At the regional level, in the context of its work on Guidelines for Trustworthy AI, the European Union's High-Level Expert Group on Artificial Intelligence has identified seven key requirements that AI applications must respect. These are human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability (European Commission 2019). A checklist to assess whether these requirements are being fulfilled has also been proposed. It is to be noted that like the international guidance listed earlier, these EU Guidelines are non-binding and need to be adapted to context. The EU has taken the next step by proposing a legal framework on AI which links rules to a four-fold hierarchy of risk: unacceptable risk, high-risk, limited risk and minimal risk (European Commission 2021a).

Under the proposed EU legislation, an independent body would take care for high-risk AI products to meet aforementioned standards. 'Inter alia' quality considerations on data to reduce risks and discriminatory outcomes have to be satisfied. Furthermore, requirements for technical documentation and traceability have to be met, as well as transparency and provision of information to users, appropriate levels of human oversight, and high levels of cybersecurity, accuracy and robustness (European Commission 2021b).

In addition to governments and international organisations, the private sector, an extremely powerful actor in the digital domain, is active on industry standards and audits on AI governance. Facebook's Community Standards or content moderation algorithms are regulatory mechanisms with cross-border impact (Meta 2019). Microsoft has created pressure on regulation of facial recognition AI across the world with its own policy (Smith 2018). Technology associations such as the Institute of Electrical and Electronics Engineers (IEEE) have created independent groups to align terminology across different domains (IEEE Ethically Aligned Design, First Edition) or promote responsible industry development and procurement practices with regard to risky AI technologies (Bloch et al. n.d.). It has created the Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) to improve transparency, accountability, and reduction in algorithmic bias. Similarly, the World Economic Forum has fostered the Responsible AI (RAI) Certification Beta (April 2021) to address accountability, bias and fairness, data quality, explainability and interpretability, and robustness (Responsible Artificial Intelligence Institute 2021).

The initiatives described above are important for preventing misuse and missed use of AI. However, landing general guidance into regulations in a specific domain can be tricky, especially if the practitioners in that domain are used to governance mechanisms from the pre-AI era. Let us examine this challenge of meeting the requirements of deployment in a specific domain with the example of health.

In June 2021, the World Health Organization (WHO) released a report on Ethics and Governance of Artificial Intelligence for Health as well as six guiding principles for design and use (WHO 2021). These principles are similar to the OECD, UNESCO or EU principles, for instance in their emphasis on human autonomy, transparency, explainability, responsibility and accountability. Additionally, the WHO report brings in important reflections pertinent

to the health domain. For instance, it cautions against overestimating the benefits of AI for health, especially when it occurs at the expense of core investments and strategies required to achieve universal health coverage. It also cautions against subordinating the interests of patients and communities to powerful commercial or government interests. Special attention is turned on problematic reliance on datasets from high income countries for training AI solutions to be used in low- and middle-income settings.

A crucial issue in current regulatory approaches in specific domains such as health is what is it that the regulator approves. Is it a 'locked' algorithm that does not learn or change over time or is it a truly adaptive system that either operates in a pre-approved range or has to be brought back for another regulatory look after some time in the field? This is one of the reasons why the U.S. Food and Drug Administration (FDA) has proposed a total product lifecycle (TPLC) approach in its new Action Plan on Artificial Intelligence/ Machine Learning-Based Software as Medical Device (U.S. Food & Drug Administration 2021).

The main trends that emerge from this high-level survey of current AI governance practices are the following. First, only a small number of countries have adopted high-level AI strategy documents and governance initiatives [see Figure 1].

Second, prevention of harm and misuse remain the overriding purpose of governance efforts. There is less emphasis on missed use. In other words, regulation is largely seen as separate from development even though the EU has begun to describe its AI regulatory effort in the context of the region's ambitions on the digital economy. And in the domain of health, it has moved forward on creating a European Health Data Space to facilitate the sharing of data for public health, treatment, research and innovation. Likewise, the WHO's recent governance efforts are moving in parallel to the adoption of a digital health strategy and the creation of a new Department of Digital Health and Innovation to promote the responsible use of data and AI for health.

Within the prevention of misuse paradigm, the focus remains on data protection based on consent. Thus, it loses sight of data empowerment whereby the citizen is at the centre of data flows and can make informed decisions about data sharing (Nilekani 2018).

Third, national governance and policy interventions targeted at missed use and missing data are emerging but remain limited to a few countries. A good example is Finland's Act on the Secondary Use of Health and Social Data (552/2019), which entered into force on 01.05.2019. A new agency, Findata, has been created to collect, combine, pseudoanonymise or anonymise datasets and reduce complexity in the process of obtaining permissions to use data.

In contrast, there is a relative penury of policy efforts focused on missed use or missing data in low- and middle-income countries, which are potential rich but data poor. Fundamentally, this is about inclusion and equity. If vast numbers are 'uncounted' and have no ways to participate in the AI/data opportunity, it opens up a new digital divide on top of the existing divides on connectivity, content and devices. An extract from a recent report shows how research and development into digital health and AI is concentrated in a few countries in Asia, North America and Western Europe where datasets and AI expertise are available in plenty (I-DAIR 2021).

Finally, in contrast to the focus on broad values and principles, for instance at the international and the regional level, lesson-drawing, experience-sharing, and capacity-building for regulators and decision makers remain underexplored. Some national regulators have started to extend existing regulatory measures for devices and services to AI-based products. A few such as the UK Competition and Markets Authority (CMA) have begun to use terms like 'algorithmic systems' to imply that governance cuts across data, models, processes, algorithms, objectives and how people use the systems. This is also meant to underline the challenge that regulators will face in terms of training and techniques to assess AI systems on an ongoing basis (UK Competition & Markets Authority 2021).

5. THE WAY FORWARD: PRACTICING AND ALIGNING AI GOVERNANCE GLOBALLY

Having looked at the 'why' and 'what' of AI governance, we now turn to its 'how' with a selection bias for methods and methodologies that can work across different geographical settings.

The first important dimension going forward is a 'global commons' approach. In a variety of domains with transboundary impact such as outer

space, the oceans and biospheres, the international community has taken this approach to governance to facilitate international cooperation and action (Lambert et al. 2021). Recent high-level policy reflections such as the UN Secretary-General's High-level Panel on Digital Cooperation have underlined the utility of a global digital commons approach with both common rails and guard rails. These can spread the benefits of digital technologies more widely and prevent another 'tragedy of the commons' through misuse [see Figure 4].

As argued in the report of the UN Secretary-General's Panel, the common rails act to make the global digital ecosystem inclusive, stimulate innovation and scaling for achieving Sustainable Development Goals (SDGs). The guard rails make sure that no one gets left behind and social harm is curbed (UN Secretary-General's High-level Panel on Digital Cooperation 2019). Further,

FIGURE 4
Prerequisites and potentials of a global commons approach to AI governance

Source: UN Secretary-General's High-level Panel on Digital Cooperation 2019, p. 10f.

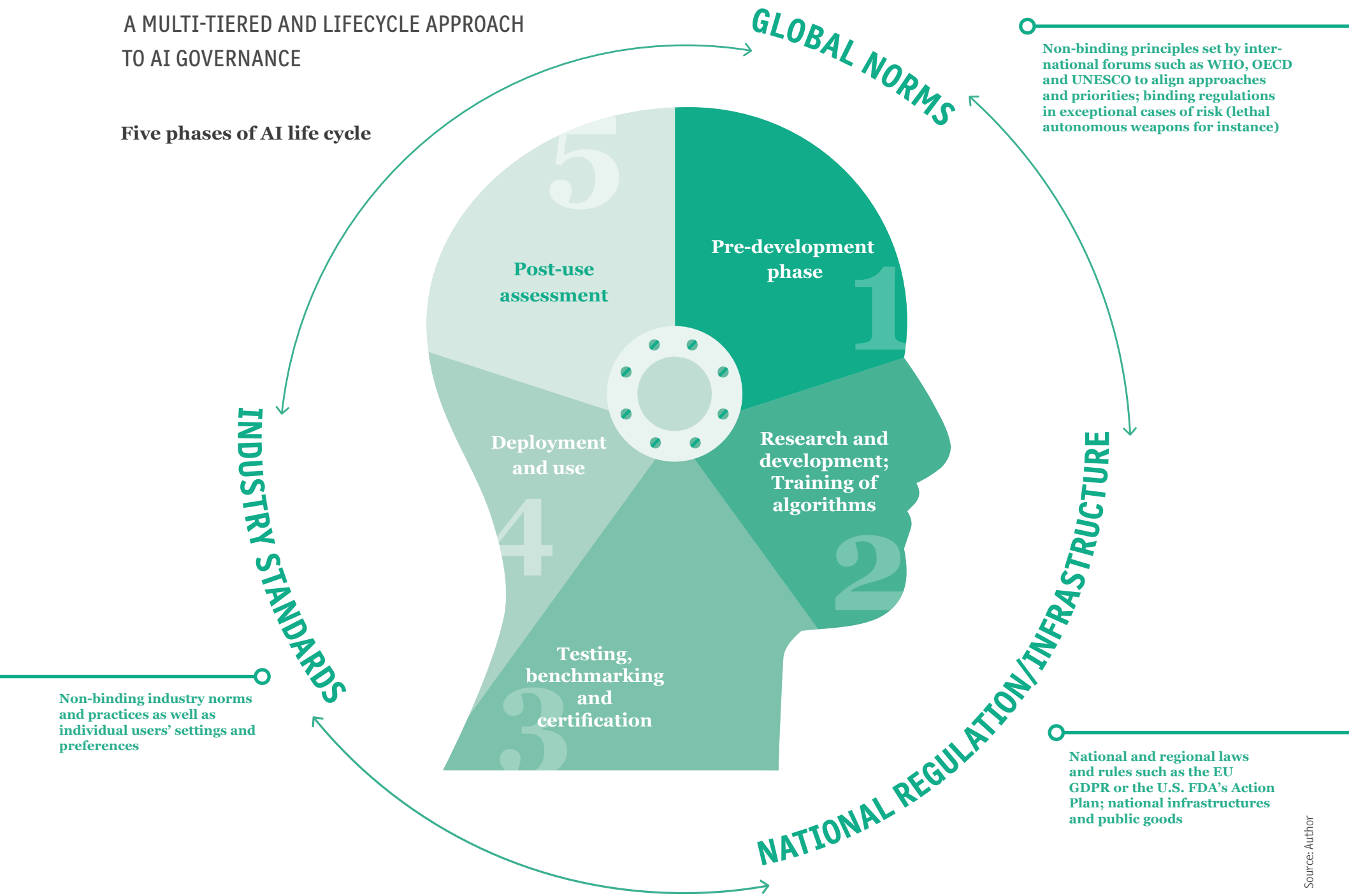
a global digital commons architecture can create dialogue on emerging issues and communicate use cases and problems to be solved to multiple stakeholders. The multi-stakeholder tracks or platforms constituting this architecture could also disseminate new data and evidence about the impact of artificial intelligence and other emerging technologies. It thus helps making discussions on governance more 'factful'.

The second dimension of 'how' is a values-informed diverse governance tool set. At the softer end of the spectrum, values and principles such as the one conceived recently at the OECD and at UNESCO can guide policy-making and implementation. They can channel international norms across nations and be more flexibly deployed across cultures and borders. They can extend governance into early/ambiguous parts of the technology development cycle. However, values and guiding principles need to be 'discovered' in each use-context and linked clearly to governance outcomes. They also need to be made visible to avoid potential manipulation by commercial or political interests ('ethics wash').

The third dimension is multi-tiered governance. Governing AI by centralising oversight over data and algorithms would be unwise either at the national or the international level. This could stifle innovation and create new risks to personal freedoms. Instead, governance measures should be applied in a tiered manner [see Figure 5]. Where necessary, say with regard to AI use in weapons systems, the governance response can be international. Where binding norms are hard to achieve, say in education and health, broad normative guidance from inter-governmental forums can be helpful in aligning national laws to a respectable mean. National or regional laws and normative frameworks to prevent harm and misuse have to be buttressed by digital architectures for data empowerment of citizens and by industry standards and practices. This counts within national and regional boundaries. One example is the 2016 EU General Data Protection Regulation. Especially industry standards and practices afford a separate opportunity to align global practices given their apolitical nature. Governance interventions at the macro, meso and micro levels can thus be mutually reinforcing to make the AI ecosystem safe, inclusive and publicly beneficial. This tiered approach also facilitates smart multi-stakeholder regulation by bringing together government, industry and civil society in an agile framework, which facilitates learning (Eising 2002).

FIGURE 5

A MULTI-TIERED AND LIFECYCLE APPROACH TO AI GOVERNANCE



Source: Author

The fourth dimension of the 'how' of AI governance is a systems approach. The AI governance problem goes above and beyond unbiased data only, or transparency and explainability of algorithms. Public officials need a lifecycle understanding of AI and need to prepare for a corresponding governance of such systems rather than simply algorithmic or data governance [see Figure 5]. They need to be able to identify the limitations and trade-offs at each stage of system development, reinforce human responsibility and accountability for use, and provide for redressal and adjustments to regulation post-deployment. Twentieth century policy tools would not suffice for a systems approach, public officials would also need new techniques to audit AI systems on an ongoing basis.

To sum it up, the practice of AI governance can be aligned across borders through a global commons umbrella, by the use of internationally agreed values and principles, and through a tiered systems approach. Global alignment of AI governance can be incentivised further by the use of common infrastructure and capacity development. A distributed infrastructure, shared for instance by collaborating scientists across different countries, would be helpful. It offers opportunities to architect good governance into technology development much as national digital public infrastructures can incorporate mechanisms for data empowerment of citizens.

6. CONCLUSION

The greatest challenge we face today with governing the digital economy is that it has grown quickly in different geographies without oversight or full understanding of its impact. Tech giants have presented policy makers with facts on the ground and accumulated incumbency power that constrains regulatory choices. They have indulged in jurisdiction shopping to give themselves a free hand. Regulators from diverse jurisdictions such as the EU, Asia and the United States are scrambling to catch-up.

Before the AI technology development and applications landscape presents another fait accompli to regulators, it will be important to align globally AI governance principles and best practices. Alignment does not mean that each country follows the same exact regulatory scheme. Nonetheless, it means that we recognise across borders that the AI domain is a global common and if there is harm and abuse it will muddy the waters for everyone. If data is missing or remains locked in silos, it will not contribute to collaborative development of AI, say by scientists trying to understand the next infectious pandemic.

A key step in aligning global AI governance is developing a common language, a shared vocabulary around technology and its impact as well as a clearly understood taxonomy of policy and governance responses. Academia and research institutions will have an important role in this regard.

Next, we must deploy values and principles to align governance across borders and shape choices of technologists at the early and ambiguous stages of technology development. Values have a way to resonate across cultures and can be rediscovered in context. Thus, they serve to engage diverse stakeholders who might have divergent perspectives and even a degree of mistrust to begin with. This is partly the reason why they have proven to be popular in multilateral settings where hard norms are difficult to craft and take relatively more time.

Another avenue for global alignment is institutionalising the exchange of governance innovations across jurisdictions. Sharing of best and worst practices can foster peer to peer learning and keep governance responses up to speed with technology adoption. UN forums in areas such as health (WHO) or education (UNESCO) can play a vital role in this regard. A foundation of good practices across diverse settings can also serve as a basis for capacity development programmes.

Shared metrics of risk and impact are other fruitful areas for global collaboration to build trust and align AI governance across borders. Harmonised benchmarks help regulators know that something works the way it is claimed to work, consistently and across different contexts.

Finally, building and sharing technology infrastructure across borders is a powerful way to incentivise good practices in AI benchmarking and governance. This is particularly true for researchers and innovators in the Global South who do not always have access to high performance computing for AI. Data security, system robustness and quality control can be mainstreamed through shared and distributed infrastructure, which has the additional advantage of making AI development more inclusive.

REFERENCES

BACRY, EMMANUEL/GAÏFFAS, STÉPHANE 2020: Machine Learning and Massive Health Data, in: Nordlinger, Bernard/Cédric, Villani/Rus, Daniela (eds.), Healthcare and Artificial Intelligence, Cham: Springer, pp. 23–31.

BALICER, RAN 2018: The Doctor Will See Your Future Now (Forbes, 16.04.2018), n.p. (https://www.forbes.com/sites/startupnationcentral/2018/04/16/for-predictive-medicine-its-back-to-the-future/?sh=1d29014f3525, 08.12.2021).

BARBEY, ARON K. 2018: Network Neuroscience Theory of Human Intelligence, in: Trends in Cognitive Science, Vol. 22/1, pp. 8–20.

BLOCH, EMMANUEL/CONN, ARIEL/GARCIA, DENISE/GILL, AMANDEEP/LLORENS, ASHLEY/NOORMA, MART/ROFF, HEATHER n.d.: Ethical and technical challenges in the development, use, and governance of autonomous weapons systems (Report by an independent group of experts convened by the IEEE Standards Association), n.p. (https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ethical-technical-challenges-autonomous-weapons-systems.pdf, 08.12.2021).

BRIGGS, ANDREW/VASSALL, ANNA 2021: Count the cost of disability caused by COVID-19, in: Nature, Vol. 593, pp. 502–505.

EISING, RAINER 2002: Policy Learning in Embedded Negotiations: Explaining EU Electricity Liberalization, in: International Organization, Vol. 56/1, pp. 85–120.

EUROPEAN COMMISSION 2019: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Building Trust in Human-Centric Artificial Intelligence (COM(2019) 168 final), Brussels (https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52019DC0168&from=EN, 08.12.2021).

EUROPEAN COMMISSION 2021a: Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in AI (Press Release), Brussels (https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682, 08.12.2021).

EUROPEAN COMMISSION 2021b: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206), Brussels (https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN, 08.12.2021).

GILL, AMANDEEP S. 2019: Artificial Intelligence and International Security: The Long View, in: Ethics and International Affairs, Vol. 33/2, pp. 169–179.

GILL, AMANDEEP S./GERMANN, STEFAN 2021: Conceptual and Normative Approaches to AI Governance for a Global Digital Ecosystem supportive of the UN Sustainable Development Goals (SDGs), in: AI and Ethics 06.05.2021 (https://doi.org/10.1007/s43681-021-00058-z, 08.12.2021).

HOROWITZ, MICHAEL/ALLEN, GREGORY C./SARAVALLE, EDOARDO/CHO, ANTHONY/FREDERICK, KARA/SCHARRE, PAUL 2018: Artificial Intelligence and International Security, Washington DC: Centre for New American Security (https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-AI-and-International-Security-July-2018_Final.pdf?mtime=20180709122303&focal=none, 08.12.2021).

I-DAIR 2021: The Global Research Map for digital health, n.p. (https://grm.i-dair.org/#/report, 08.12.2021).

KICKBUSCH, ILONA ET AL. 2021: The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world, in: The Lancet, Vol. 398/10312, pp. 1727–1776.

KOROTEEV, DMITRY/TEKIC, ZELJKO 2021: Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future, in: Energy and AI, Vol. 3/March 2021 (https://doi.org/10.1016/j.egyai.2020.100041, 08.12.2021).

LAMBERT, DOMINIQUE/REICHBERG, GREG/THELISSON, EVA 2021: Human Fraternity in Cyberspace: Ethical Challenges and Opportunities (Caritas in Veritate Foundation), n.p. (forthcoming).

LATOUR, BRUNO 2012: Love Your Monsters: Why We Must Care for Our Technologies As We Do Our Children, The Breakthrough Institute (https://thebreakthrough.org/journal/issue-2/love-your-monsters, 08.12.2021).

LEWIS, DYANI 2020: Where Covid Contact-Tracing Went Wrong, in: Nature, Vol. 588, pp. 384-388.

META 2019: Writing Facebook's Rulebook, n.p. (https://about.fb.com/news/2019/04/inside-feed-community-standards-development-process/, 08.12.2021).

NIILER, ERIC 2020: An AI Epidemiologist Sent the First Warnings of the Wuhan Virus (Wired, 25.01.2020), n.p. (https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/, 08.12.2021).

NILEKANI, NANDAN 2018: Data to the People. India's Inclusive Internet, in: Foreign Affairs, Vol. 97/5, pp. 19–26.

OECD 2019: Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.

RESPONSIBLE ARTIFICIAL INTELLIGENCE INSTITUTE 2021: Responsible Artificial Intelligence (RAI) Certification Beta, n.p. (https://assets.ctfassets.net/rz1q59puyoaw/1myaH22mA16YoeIXQND3qv/7974df6bd0973e65f100d327b93129a2/Whitepaper.pdf, 08.12.2021).

SJODING, MICHAEL W./DICKSON, ROBERT P./IWASHYNA, THEODORE J./GAY, STEVEN E./VALLEY, THOMAS S. 2020: Racial Bias in Pulse Oximetry Measurement, in: New England Journal of Medicine, Vol. 383/25, pp. 2477–2478.

SMITH, BRAD 2018: Facial recognition technology: The need for public regulation and corporate responsibility (Microsoft Blog, 13.07.2018), n.p. (https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/, 08.12.2021).

SZCZEPAŃSKI, MARCIN 2019: Economic impacts of artificial intelligence (AI), Briefing to the European Parliament (European Parliamentary Research Service, PE 637.967), n.p. (https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf, 08.12.2021).

TOPOL, ERIC 2019: Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, New York: Basic Books.

UK COMPETITION & MARKETS AUTHORITY 2021: Algorithms: how they can reduce competition and harm consumers, n.p. (https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers, 08.12.2021).

UN SECRETARY-GENERAL'S HIGH-LEVEL PANEL ON DIGITAL COOPERATION 2019: The Age of Digital Interdependence, Report of the UN Secretary-General's High-level Panel on Digital Cooperation, n.p. (https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf, 08.12.2021).

UNESCO (UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION) 2021: Draft Text of the Recommendation on the Ethics of Artificial Intelligence (SHS/IGM-AIETHICS/2021/APR/), n.p. (https://unesdoc.unesco.org/ark:/48223/pf0000376713, 08.12.2021).

UNIVERSITY OF CAMBRIDGE 2021: BloodCounts! Breakthrough in disease detection (University of Cambridge News, 25.06.2021), n.p. (http://www.haem.cam.ac.uk/blog/bloodcounts-breakthrough-in-disease-detection/, 08.12.2021).

U.S. FOOD & DRUG ADMINISTRATION 2021: Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan, n.p. (https://www.fda.gov/media/145022/download, 08.12.2021).

WAHL, BRIAN/COSSY-GANTNER, ALICE/GERMANN, STEFAN/SCHWALBE, NINA 2018: Artificial Intelligence (AI) and Global Health: how can AI contribute to health in resource-poor settings?, in: BMJ Global Health (http://dx.doi.org/10.1136/bmjgh-2018-000798, 08.12.2021).

WARTOFSKY, MARX W. 1979: Models. Representation and the Scientific Understanding, Dordrecht: Springer.

WHO (WORLD HEALTH ORGANIZATION) 2021: Ethics and Governance of Artificial Intelligence for Health: WHO guidance, Geneva.

THE AUTHOR

AMANDEEP SINGH GILL
Project Director/CEO of the International Digital Health & AI Research Collaborative (I-DAIR), and Professor of Practice, The Graduate Institute of International and Development Studies, Geneva

PREVIOUS ISSUES

All issues are available free of charge at http://www.sef-bonn.org/en

GLOBAL TRENDS. ANALYSIS 02|2021

Let's speak law!
A call for a legally embedded multilateralism
Heike Krieger
December 2021, 25 pages

For some time now, multilateralism, which is embedded in international law, has come under pressure. However, protracted turbulences and ambivalences which sometimes point in diametrically opposed directions create space for political actors. In GLOBAL TRENDS. ANALYSIS 2|2021, Heike Krieger calls on EU member states to promote favourable trends for stabilising the international order. To this end, they should prefer a legally embedded type of multilateralism over informal network structures of the like-minded. This will require these states to act consistently, credibly and compliantly and to continuously negotiate for shared understandings of international law, in particular with the Global South.

GLOBAL TRENDS. ANALYSIS 01|2021

Freeing Fiscal Space:
A human rights imperative in response to COVID-19
Ignacio Saiz
May 2021, 27 pages

Inequality between states has been magnified by the COVID-19 pandemic. The economic consequences have been particularly devastating in countries of the Global South. The resources they can mobilise to respond to the crisis are, however, totally inadequate. This makes it all the more important that the wealthier countries and the international financial institutions cooperate by lifting the barriers their debt and tax policies impose on the fiscal space of low- and middle-income countries. Such cooperation is not only a global public health imperative. It is also a binding human rights obligation, as Ignacio Saiz explains.

GLOBAL TRENDS. ANALYSIS 03|2020

Tech power to the people!
Democratising Cutting-edge Technologies to Serve Society
Renata Ávila Pinto
December 2020, 27 pages

The technologisation and digitisation of public services is advancing rapidly. However, the hoped-for increase in efficiency and cost reduction is associated with the risks of discrimination and surveillance. The Guatemalan human rights lawyer Renata Ávila Pinto therefore calls for the design of tech interventions in the public sector to be guided more strongly by human rights, democratic rules and the objectives of sustainable development. This requires a greater degree of independence from big tech companies, participatory design and testing in collaboration with the communities the technologies are intended to serve.

Institut für Entwicklung und Frieden (INEF),
Universität Duisburg-Essen, Duisburg

sef: Stiftung Entwicklung und Frieden
Development and Peace Foundation

INEF

GLOBAL TRENDS. ANALYSIS

examines current and future challenges in a globalised world against the background of long-term political trends. It deals with questions of particular political relevance to future developments at a regional or global level. GLOBAL TRENDS. ANALYSIS covers a great variety of issues in the fields of global governance, peace and security, sustainable development, world economy and finance, environment and natural resources. It stands out by offering perspectives from different world regions.