Centre on Regulation in Europe

TOWARDS AI ADEQUACY

OPERATIONALISING THE PRINCIPLES UNDERPINNING GLOBAL GOVERNANCE OF AI SYSTEMS

September 2024

Gianclaudio Malgieri

GLOBAL GOVERNANCE FOR THE DIGITAL ECOSYSTEMS: PHASE TWO

Issue Paper

Towards AI Adequacy: Operationalising the Principles Underpinning Global Governance of AI Systems

Gianclaudio Malgieri

September 2024

As provided for in CERRE's bylaws and procedural rules from its “Transparency & Independence Policy”, all CERRE research projects and reports are completed in accordance with the strictest academic independence.

The project, within the framework of which this report has been prepared, received the support and/or input of CERRE member organisations. However, these bear no responsibility for the contents of this report. The views expressed in this CERRE report are attributable only to the authors in a personal capacity and not to any institution with which they are associated. In addition, they do not necessarily correspond either to those of CERRE, or of any sponsor or of members of CERRE.

info@cerre.eu – www.cerre.eu

TABLE OF CONTENTS

About CERRE

Providing top quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network and digital industries. CERRE's members are regulatory authorities and operators in those industries as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach;
- the widely acknowledged academic credentials and policy experience of its team and associated staff members;
- its scientific independence and impartiality;
- the direct relevance and timeliness of its contributions to the policy and regulatory development process applicable to network industries and the markets for their services.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological and social environment. CERRE's work also aims at clarifying the respective roles of market operators, governments and regulatory authorities, as well as at strengthening the expertise of the latter, since in many Member States, regulators are part of a relatively recent profession.

About the Author

Gianclaudio Malgieri is an Associate Professor of Law & Technology and a Board Member at eLaw – Center for Law and Digital Technologies. He serves as the Co-Director of the Brussels Privacy Hub, Free University of Brussels (VUB) and as an Affiliated Researcher at the Augmented Law Institute of the EDHEC Business School (Lille, France). He is an Associate Editor of Computer Law and Security Review, an External Ethics Expert of the European Commission, and an Advisory Board member of EPIC.org. He also coordinates VULNERA, the International Observatory of Vulnerable People in Data Protection. His field of research and teaching are data protection law, privacy, AI regulation, digital law, consumer protection in the digital market, data sustainability, and intellectual property law.

With contributions from **Niti Chatterjee** (CIPP/E, L.L.M. University of Leiden), and **Liubomir Nikiforov** (VUB Brussels).

Introduction

In the rapidly evolving landscape of Artificial Intelligence (AI), the development and deployment of AI technologies heralds unprecedented opportunities for advancement in science (biology, physics, material sciences, etc.), engineering, health, and education. This potential, however, will not be realized extemporaneously. Progress requires well-crafted policies and effective governance. The introduction of AI systems also introduces complex challenges that span technical, ethical, legal, and societal dimensions. CERRE's report Generative AI: Global Governance and the Risk-Based Approach[1] takes this first step of this by articulating some of the risks that have driven regulatory scrutiny.

In the case of AI, the quest for regulatory convergence is driven by the recognition that AI technologies transcend national and jurisdictional boundaries, necessitating an international consensus on fundamental principles and standards to address global challenges, promote innovation and trust in AI technologies, and ensure that the benefits of AI are realized equitably across societies.

Structured to offer a comprehensive understanding of existing AI governance frameworks in the EU, critically evaluate their effectiveness and gaps, and propose pathways towards achieving globally coherent AI regulation, this document aims to contribute to the ongoing dialogue among policymakers, industry leaders, academics, and civil society. It explores the possibility of developing **AI Adequacy guidelines based on ethics, transparency, accountability and common human values and rights.**

Note that throughout this paper, the term "AI regulation" is used to encompass the formal adoption of legal statutes or universal standards for AI, and in essence is an important subset of the term "AI governance", an all-encompassing term that involves policymaking, voluntary commitments from industry, and the eventual adoption of standards or statutes.

The imperative for robust regulatory frameworks to govern the multifaceted impacts of AI has become increasingly pronounced.

In the above mentioned CERRE report, the concept of **AI Adequacy** was introduced and explored, as it emphasizes the harmonization of standards for AI systems and the legislative frameworks that regulate their application. This project closely analysed the Hiroshima Approach, as well as compared different approaches across key countries, including the US, China, India, Japan, etc. These jurisdictions either have or do not have at all sectorial and segmented legislation on the subject, or they prefer to adopt voluntary technical standards instead of legal statutes. In the CERRE report, we recommended that the G7 adopt a "risk-based approach" to regulating AI systems. In order to operationalise this, it will be important to harmonise regulatory interventions and introduce accountability mechanisms that at least bring confidence to regions like the European Union where AI laws already exist.

AI Adequacy emerges as a guiding principle, advocating for a dual focus: ensuring that AI systems themselves are inherently safe, transparent, and accountable, and that national and international legislative environments are equipped to effectively oversee these technologies. This document

[1] "Generative AI : Global Governance and the Risk-Based Approach", URL : https://cerre.eu/publications/generative-ai-global-governance-and-the-risk-based-approach/

endeavours to bridge the gap between technological advancements and regulatory frameworks. The present document aims to outline a vision for a regulatory convergence that balances the promise of AI with ethical oversight and societal wellbeing.

By synthesizing insights from diverse sources, this report seeks to map the current landscape of AI regulation, identifying commonalities and divergences that could inform the creation of a harmonized global governance framework.

Legislative Foundation for AI Adequacy in the EU

The need of AI Adequacy is inspired, at least, in part by two pieces of EU legislation, the General Data Protection Regulation (GDPR) and the newly approved European Union (EU) AI Act. Those two pivotal legislative frameworks are widely considered as benchmarks for AI regulation. Under Article 45 of the GDPR, the European Commission is empowered to determine whether a non-EU country provides an adequate level of data protection to facilitate the cross-border flow of data, while ensuring that the data of rights holders under the GDPR is protected.

The enforcement architecture of the EU AI Act is an example of *ex-ante* justification for AI systems, whereby the respective system is deemed adequate according to specific, delineated principles (transparency, accountability, human oversight, accuracy, security, etc.).[2] These principles serve to build trust among citizens as well as ensure "trustworthiness" – which refers to the inherent qualities of the AI system. Justification, or "justified trust" will need to be built up between industry players and governments, framework to determine adequacy of AI.

In addition, it will be important for countries to devise tools for the implementation of specified "AI adequacy principles". AI Adequacy should be developed to specify measures for maintaining data quality, as well as to responsibly deploy advanced AI products across G20 countries.

The GDPR contributed to the consolidation of the "Brussels effect" where EU regulation influences foreign regulatory regimes and policies. The explanation and analysis of the effect, however, escapes the scope of this document. Nonetheless, it is important to recall that the GDPR is the fruit of a decades long international consensus evolution,[3] coupled with preceding case law, doctrine thereof and the experience from the Directive 95/46/EC, which is the precursor of the regulation we know today. Therefore, the GDPR is the emanation of historical lessons learnt, experience and fundamental human values, which other countries have also chosen to adopt. All of this demonstrates the universality of the GDPR as a principle-based piece of legislation applicable to a variety of contexts. Although the EU AI Act is still in its infancy, it has the potential to repeat the same effect as the GDPR. This is because the AI Act is one of the few comprehensive laws regulating AI systems with a wide geographic applicability (across the EU). Another factor which contributes to the likelihood that the AI Act produces its own "Brussels effect" is that it fundamentally aims to achieve legal certainty and prevent fragmentation of the EU market while at the same time striking basic rules for the use of a

[2] Malgieri, G., Pasquale, F. (2023), "Licensing high-risk artificial intelligence: Toward ex ante justification for a disruptive technology", URL: https://www.sciencedirect.com/science/article/pii/S0267364923001097

[3] While it is not the objective to site all the cases which contributed to the current evolution, which led to the adoption of the GDPR, here we can cite two international documents as evidence of the consensus around the principles and rules which drive global community and on which the GDPR is based: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, 1980 , Council of Europe, "Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data" (Convention 108), 1981

rapidly evolving technology, establishing a layered risk-based approach.[4,5] As a result, it is probable that the underpinning design of the AI Act facilitates international convergence, as its design promotes replicability on an international scale.

In addition, the GDPR, with its focus on data protection and privacy, provides insights into the ethical underpinnings necessary for AI governance. To complement this, the EU AI Act marks a significant step forward in establishing comprehensive regulations centred around risk assessment, transparency, and human oversight, in the same vein as other platform regulation like the EU Digital Services Act (DSA).

These EU-based frameworks, alongside contributions from international organizations, national governments, industry groups, and academia, constitute the core of our exploration into AI governance.

[4] European Commission, 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 final) 2021

[5] European Commission, “Inception Impact Assessment for a Proposal for a Legal Act of the European Parliament and the Council Laying down Requirements for Artificial Intelligence”

Principles Underpinning AI Regulation

The following sections analyse the most relevant principles for the regulation of AI systems (e.g., as defined in the EU AI Act, or what is referred to as "advanced AI systems" in the Hiroshima Process). The GDPR and the AI Act are pivotal in identifying which principles that emerge from the EU should underpin AI Regulation. These findings will be complemented by other international legal sources from organisations such as the G7, the OECD, the Council of Europe, and the UN. The aim is to provide guiding principles for our AI Adequacy model for global convergence on AI regulation.

Principle-driven regulation of AI first arose at the multilateral level at the OECD. As national governments increasingly started developing policy on AI, the OECD principles have become a global reference for "trustworthy AI".[6] In addition, the OECD has been working with governments to adapt to evolutions in AI legislation, for example, during the AI Act negotiations, the OECD updated its definition of AI. Recently, it has been working on a "Revised Recommendation on Artificial Intelligence", which was approved in the OECD Digital Policy Committee on 4 April 2024 and adopted on the 3 May 2024.[7] This recommendation articulates five key principles, which overlap with our proposed framework.[8]

To start, the principles which the present paper identifies are as follows, each will be discussed in turn:

- Lawfulness
- Purpose limitation
- Accuracy (integrity)
- Human agency and oversight
- Technical robustness and safety
- Privacy (data minimisation, storage limitation, confidentiality)
- Transparency
- Openness
- Accountability
- Data governance
- Diversity
- Fairness and non-discrimination
- Societal and environmental wellbeing

A. Lawfulness

AI governance should be guided by democratic standards such as the rule of law. The rule of law in the context of AI encompasses the principles of lawfulness, fairness, transparency, and accountability. This is why the principle of lawfulness is either implicit or explicit. When it comes to legal norms regulating the relationship between public authorities and citizens, the principle of lawfulness is

[6] See the report "State of OECD AI Principles: Four Years On". URL: https://www.oecd-ilibrary.org/docserver/835641c9-en.pdf?expires=1713863568&id=id&accname=guest&checksum=985E506588C3BA2366A54BF0DF5C0F53

[7] https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

[8] The five principles outlined in the draft are (1) inclusive growth, sustainable development and wellbeing; (2) human-centred values and fairness; (3) transparency and explainability; (4) robustness, security and safety; and (5) accountability.

implicit because every public authority’s action emanates from an implicit premise of lawfulness whose source might be constitutional or contextual such as national security protection for example. This principle could be also explicit, and it usually is, especially when the given norm governs the transactions between different legal subjects in a horizontal relationship such as the one between parties in a contract. The GDPR adopts precisely the latter approach.

The GDPR codifies several decades of international legislation and case law in the domains of data protection and privacy. It establishes the rights of data subjects and the rules that enable the collection, processing, and use of their personal data. It also establishes the role and liability of data controllers and processors. The GDPR, however, goes further in developing the legal norm by embedding a normative compass, which guides each stakeholder’s actions. The mandated actions to be followed are provided in Article 5, which stipulates the principles concerning the processing of personal data.[9]

Article 5 has a central role as a guide for all data processing because it encompasses an extended definition of the principle of the rule of law. It mandates that data processing should be carried out lawfully, fairly, and transparently. While the latter two elements are expanded further in the same Article, the EU lawmaker expressly developed the notion of lawfulness in Article 6.

Article 6 establishes the normatively permitted preconditions under which both public and private parties may carry out data processing. Hence, the principle of lawfulness delimitates the contours of what is legally allowed by limiting and outlining the requirements for how subjects to the GDPR can process personal data. Every processing falling outside of these contours is a violation of Art. 8 EU Charter and is therefore, sanctionable.[10]

Given that AI systems and the foundational models underlying them make use of huge data sets, which might contain personal data, and whose further deployment and use by citizens also implies that substantially more personal data feeds the operability of the AI system or model (through “learning”), AI governance should follow closely the guardrails enclosed in the GDPR, in particular cohesion with the principle of lawful data processing.[11] On a global level, the principle of lawful processing is recognised as a requisite for the rule of law principle. The UN AI Advisory Body Interim Report,[12] in its preliminary recommendations, refers to AI governance as grounded in the rule of law (Guiding Principle 5). In further detail, the OECD AI Principles stipulate that AI should respect the rule of law (Principle 1.2).[13]

[9] Any legal norm contains an "algorithm" (or a "set of logical instructions") which follows a set logic in fixed steps: (1) hypothesis; (2) provision and (3) sanction. At the same time, principles act as a metaphorical "north pole" to achieve specific objectives related to fundamental rights, human dignity, etc.)

[10] Felix Bieker, ‘EU Data Protection Legislation’ in Felix Bieker (ed), *The Right to Data Protection: Individual and Structural Dimensions of Data Protection in EU Law*, vol 34 (1st edn, TMC Asser Press 2022) 22 <https://doi.org/10.1007/978-94-6265-503-4_2> accessed 5 January 2024.

[11] It is worth examining how the EU AI Act approaches the debate on data processing. The Act does not include a list of the legal grounds for data processing (AI Act, Recital 41). However, it expressly refers to the GDPR, for example, in Recitals 5aa, 7, Art. 3 paras 33d, 44a, 44c and be, Arts. 29 and 29a.

[12] UN Advisory Body on Artificial Intelligence, ‘Interim Report: Governing AI for Humanity’ (United Nations 2023) Interim <https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf> accessed 13 February 2024

[13] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, (8 November 2023) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 13 February 2024

It follows that there is international consensus that the rule of law is a pivotal element of AI governance. This means that the use of AI tools should be subjected to specific legal requirements which begin with the lawful grounds which permit the incision into the private sphere of citizens by AI actors as they process personal data.

B. Purpose Limitation

This principle examines two types of purpose limitation: the first concerns **training data underlying AI models**, and the second relates to **the AI system itself**, specifically to mitigate against unintended use.

Article 5(1)(b) of the GDPR sets out the principle of purpose specification. It prescribes purpose specification (that personal data must be processed for specified, explicit, and legitimate purposes) and compatible use (prohibition on further processing in any manner incompatible with the initial purpose of processing). It also exempts the use of personal data for scientific, historical research, or statistical purposes that are not considered to be incompatible with the initial purposes.

In an AI systems context, adherence to this principle ensures that AI systems that use personal data need to identify a precisely defined purpose (objective) that meets the test of legitimacy and compatibility at every stage of development, training, and deployment. Interestingly, drawing from national approaches, the CNIL advocates the identification of a primary objective at the design stage of AI systems.[14] Further, to meet the threshold of explicit purposes – the CNIL's recent guidance[15] on GDPR vis-à-vis AI systems clarifies that the purpose of processing must be known and understandable to the data subject.[16] However, the exemption for scientific and research purposes could serve as an essential pathway for innovation in AI. The notion of purpose limitation is also directly linked to that of explainability. As per the HLEG guidelines, the "capabilities and purpose of AI systems must be openly communicated" to meet the threshold of explainable AI.

Interestingly, while purpose limitation as a concept is not explicitly included within the paradigm of the AI Act, the categorisation of high-risk AI systems is based on the 'intended purpose' of their usage (refer to Article 7(2)(a) of the AI Act). For instance, this includes AI systems used for remote biometric identification and recruitment (refer to Annex III of the AI Act). However, as a necessary corollary, other use cases of the same AI system remain outside the purview of the additional obligations applicable to high-risk systems, given the focus on specific 'intended' purposes to identify high-risk systems.

Further, the elevated obligations for "general purpose" AI systems also draw out the AI Act's focus on adhering to the purpose limitation principle by additional guardrails towards systems that may not meet the GDPR's threshold. Another interesting aspect of purpose limitation stems from the definition of an AI system (Article 3(1), AI Act) which provides that such systems work towards "implicit or explicit objectives". The corresponding Recital 6 observes that the objectives of the AI system may be different from the intended purpose of the AI system in a specific context – which also conveys the dichotomies inherent in AI systems and the possibility of potential tension with the GDPR's purpose limitation

[14] AI: 'Ensuring GDPR Compliance' <https://www.cnil.fr/en/ai-ensuring-gdpr-compliance> accessed 13 February 2024.

[15] IA : 'Définir Une Finalité' <https://cnil.fr/fr/definir-une-finalite-0> accessed 15 April 2024

[16] Directorate-General for Parliamentary Research Services (European Parliament), Francesca Lagioia and Giovanni Sartor, '*The Impact of the General Data Protection Regulation on Artificial Intelligence*' (European Parliament 2020) <https://data.europa.eu/doi/10.2861/293> accessed 13 February 2024.

principle. The notion of purpose limitation as a principle emerges not in the limited GDPR interpretation of the purpose of the AI system, but rather from an AI Act perspective of "objectives" and "use" in a high-risk context.

C. Accuracy (Integrity)

Accuracy in data protection is one of the fundamental principles, denoting that personal data must always be accurate and, where necessary, up to date. The GDPR also requires reasonable steps to be taken to rectify records without undue delay to avoid misleading or inaccurate processing. Therefore, to satisfy the data protection understanding of accuracy, the input that is fed into a system and its output, in terms of personal data processing, must be accurate. However, in an AI context, accuracy is often understood as the accuracy of such systems itself, or as the UK's ICO calls it – '*statistical accuracy*', to refer to the ability of the AI system to provide statistically accurate results.[17]

The EU High Level Expert Group published a voluntary list of guidelines for trustworthy AI (HLEG Guidelines[18]) and include a similar notion of accuracy, requiring an AI system to ensure its output is accurate, for example, by providing correct judgements, predictions, recommendations, or decisions based on data or models. The rationale for accuracy as a key AI principle is linked to correcting or mitigating unintended harms. The EU AI Act also highlights the risks linked to inaccuracy, such as discrimination and biased output. It requires high-risk AI systems to include a particular focus on accuracy (refer to Art. 15 AI Act, which places specific obligations on providers). High-risk AI systems must meet the threshold of appropriate accuracy throughout their lifecycle in the form of performance metrics linked to robustness and cybersecurity, amongst others. The European Commission is also expected to develop collaborative benchmarking for such metrics to overcome the unique challenges in measuring and ensuring accuracy, where attempts are often hindered by regulators facing difficulties with the proper identification and assessment of providers' practical realities. The additional obligation of setting up a quality management system (see Art. 17 of the AI Act, regarding obligations of providers) which includes, amongst other elements, quality control and quality assurance of high-risk AI systems, is also a critical element which will further permeate the accuracy principle into these high-risk AI systems.

Having said that, **it is interesting to note that the GDPR and the AI Act complement each other, wherein the GDPR's accuracy principle ensures the accuracy of the data that is fed into the AI system and the AI Act's formulation of accuracy aims to ensure a consistent and fair outcome in the decisions/recommendations provided by the high-risk system**. However, it is worth noting that accuracy as a principle is not similarly relevant for AI systems which are not covered by Annex III of the AI Act (i.e., are defined as "high risk" systems). Since it is important to include suitable safeguards to prevent inaccuracy,[19] it is pertinent to answer whether protection 'by design' may be needed from the inception of all AI systems, irrespective of its intended purpose. Another relevant question here is

[17] '*What Do We Need to Know about Accuracy and Statistical Accuracy*?' (19 May 2023) <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/what-do-we-need-to-know-about-accuracy-and-statistical-accuracy/> accessed 15 February 2024.

[18] See https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[19] AEPD, 'Artificial Intelligence: Accuracy Principle in the Processing Activity' (31 May 2023) <https://www.aepd.es/en/prensa-y-comunicacion/blog/artificial-intelligence-accuracy-principle-in-processing-activity> accessed 15 February 2024.

how to interpret the GDPR's provisions considering new technological developments and whether its provisions will need to be updated and reviewed accordingly.

In this context, it is also important to refer to the cybersecurity-related obligations of the EU AI Act (Article 15), which are not just relevant from the perspective of safety and technical robustness but are also an important asset in relation to meeting the threshold of accuracy. It is well-known that cybersecurity attacks can lead to breaches of confidentiality, integrity, and availability of information.[20] Further, cybersecurity attacks may leverage AI assets, leading to data poisoning or even interference with trained models, resulting in an adverse impact on the integrity of the results produced by the AI system and consequently affecting the system's 'accuracy'.[21]

D. Human Agency and Oversight

The permeation and integration of AI technologies into a progressively wider range of domains necessitates a deliberate emphasis on preserving human autonomy and preventing adverse outcomes. To this end, human oversight must constitute a fundamental principle in the management and deployment of AI systems. The specific characteristics and application domains of an AI system will make evident the appropriate nature and scope of oversight over the system in question. Critical to this oversight are the adaptability, precision, and comprehensibility of AI technologies.

This aligns with the GDPR which mandates that individuals must retain the right to not be subjected exclusively to automated decision-making processes, especially those that significantly impact them legally or in other substantial ways (as delineated in Article 22 of the GDPR).

Effective oversight can be facilitated through various governance strategies, including the implementation of human-in-the-loop frameworks. The choice of oversight model depends on the specific application and the desired balance between autonomy and control, with an understanding that less human oversight necessitates more rigorous testing and governance protocols.

Furthermore, the legislative framework underpinning AI development and deployment, notably the AI Act, underscores the importance of human oversight. Article 14 and Recital 48 of the AI Act stipulate that high-risk AI systems must be designed to enable effective human oversight.

Prior to their market introduction or operational deployment, AI systems must be assessed to identify and integrate necessary oversight measures. These measures should ensure that AI systems operate within defined constraints, remain subordinate to human directives, and that individuals tasked with oversight possess the requisite skills, training, and authority.

In summary, ensuring human oversight in AI systems is not just a regulatory compliance requirement but a foundational ethical principle. It safeguards human autonomy, mitigates risks, and enhances the trustworthiness of AI applications. As AI technologies continue to evolve, the principles of oversight and governance must be dynamically adapted to ensure they remain effective and relevant. Public authorities play a crucial role in this process, requiring the authority and resources to execute their

[20] See for example pg 6 of ENISA's 2023 Threat Landscape Report ; (19 October 2023) https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023; Accessed on 15 February 2024.

[21] See for example pg 13-14 of OpenAI's ChatGPT4 System Card ; (23 March 2023) https://cdn.openai.com/papers/gpt-4-system-card.pdf; Accessed on 15 February 2024.

oversight responsibilities effectively, ensuring that AI serves the public good while adhering to ethical and legal standards.

E. Technical Robustness and Safety

The conceptualisation of technical robustness and safety in the context of AI systems takes inspiration from product liability. The HLEG Guidelines also include this principle as one of the critical parameters to ensure trustworthiness in AI systems. In this context, it is important to mention that technical robustness has two aspects – one which lies in cybersecurity and relates to the technical resilience demonstrated by the AI system against unlawful interference by third parties. The other aspect lies in the ability of the system to minimise unintended harm. In this context, technical robustness refers to a societal perspective and denotes safeguards against harm in general.

Notably, both of these aspects tie in with the notion of safety – one from a cyber-safety perspective and the other from the perspective of general guardrails against harms (such as misrepresentation, discrimination, etc.). In this context, it is important to highlight the critical role played by technical solutions as well as the inclusion of human oversight. The AI Act references both, highlighting the importance of having detailed documentation (see Art. 13 and 15 for obligations of deployers), setting out elements which need inclusion and explanation, as well as calling for the establishment of a risk management system for the lifecycle of high-risk AI systems (see Art. 9).

Additionally, the EU AI Act also underlines the importance of human oversight for AI systems by insisting on appropriate human-machine interface tools and requiring a human-in-the-loop to ensure oversight commensurate with the risks, level of autonomy, and context of use of the AI system (see for example Article 14). Undeniably, these are essential considerations in the preservation of safety and security stemming from the deployment of AI systems for mass use.

F. Privacy (Data Minimisation, Storage Limitation, and Confidentiality)

AI Governance should be based on the respect for human rights. Various international agreements which guarantee digital rights aim to protect individuals from interference in their private lives, restrict both private and public surveillance, and ensure citizens have control over their data. Following the Council of Europe Guide to Human Right for Internet Users,[22] human rights should apply equally online and offline. The Universal Declaration of Human Rights, Article 12, establishes that everyone has the right to privacy. In the EU context, privacy and data protection are fundamental rights (Article 7 and 8 of the Charter of Fundamental Rights of the EU). Moreover, the UN Internet Governance Forum Internet Rights and Principles Coalition,[23] as well as the European Declaration on Digital Rights and Principles,[24] include data protection and privacy among the human rights protected online.

[22] '*Council of Europe and Internet - Human Rights, Democracy and Rule of Law*' <https://edoc.coe.int/en/internet/6078-leaflet-council-of-europe-and-internet.html> accessed 14 February 2024.

[23] 'Internet Rights and Principles Coalition – Committed to Making Internet Work for Human Rights' (Internet Rights and Principles Coalition) <https://internetrightsandprinciples.org/> accessed 20 February 2024.

[24] 'European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future' (2022) <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles> accessed 22 April 2024.

On a technical level, the requirements to ensure the protection of individuals' privacy and personal data are data minimisation, storage limitation, integrity, and confidentiality.

First, AI actors across the technology's development, deployment, and use value chain should implement the necessary measures to **limit the amount of personal data collected** and reduce the privacy and data protection risks related to unauthorized data uses or access.

Second, **storage limitation**, which is the maximum time beyond which data should be deleted or anonymised, could emerge as another legal requirement for adequate AI governance. Given the risk of personal data exposure as it is collected to build the model, or as users reveal this personal data while using this AI tool, personal data should ideally be kept for a specific amount of time and not indefinitely. This functionality of purpose and storage limitation of AI data will eventually become salient as markets for copyrighted and "clean" data continue to develop.

Finally, **integrity and confidentiality** ensure that where data is collected it is managed with the security measures which are necessary to prevent any unauthorised access to - or corruption of - the personal. This requirement suggests that AI technologies should be protected from malicious attacks on software but also on a hardware level.

G. Transparency

Transparency stands as a cornerstone for robust AI governance, addressing the challenge of processing vast amounts of data that surpass human capabilities for comprehension and verification. Recognized by the GDPR, the EU AI Act, and the OECD AI principles, the requirement for transparency in AI encompasses expressly defined mandates. According to Felix Bieker in "The Right to Data Protection",[25] transparency should not only shed light on the general architecture of algorithms but also provide detailed insights into how an individual's data is processed. This includes, as stipulated by the GDPR in Articles 13-14, information about the identity of the data controller, the purposes of data collection, the lawful grounds for processing, the recipients of the data, and the rights available to individuals (Art. 12-22).

Furthermore, the manner in which information is conveyed to data subjects is critical; it must be accessible and understandable, ensuring that the average AI user can easily acquire and comprehend this information. The EU AI Act, particularly in Article 13, underscores the importance of transparency provisions for high-risk systems, mandating that information provided to users must enable them to interpret the system's output and use it appropriately. This includes providing instructions for use that are concise, complete, correct, and clear, ensuring relevance, accessibility, and comprehensibility for users.[26]

Transparency thus serves as a mechanism through which individuals can exercise control over their data and hold accountable those who lawfully process it. It is indispensable for the exercise of citizens' rights, as detecting, verifying, or proving data infringements becomes infeasible without individual oversight. Through facilitating individual and public oversight mechanisms, transparency not only

[25] Bieker (n 7) 25.

[26] Maxwell, W. and Dumas, B., (2023) Meaningful XAI Based on User-Centric Design Methodology. CERRE.

empowers individuals but also enforces accountability, contributing to trust in AI technologies and mitigating information asymmetries.

H. Openness

The concept of openness in AI systems complements transparency by advocating for the disclosure and sharing of AI technologies, algorithms, data, and findings, along with openness by regulatory authorities to take feedback and inputs from representatives of affected groups. Openness involves making AI research, development, and deployment processes accessible to a wider community, including researchers, practitioners, and the public, through e.g., deployment strategies like the open sourcing of foundational models. This approach encourages collaborative innovation, peer review, and ethical scrutiny, fostering a culture of shared responsibility and continuous improvement. Open AI models and systems can also help small and medium enterprises to scale up and innovate. Open AI systems can facilitate a deeper understanding of AI's impact, promote inclusiveness, and ensure that AI advancements benefit a broader spectrum of society.

Openness allows the governance of AI to move towards a more democratic, participatory, and equitable direction, where accountability is not just institutionally enforced but collectively assured as “adequate”. In addition, openness can act as a foundational principle for governance frameworks, for example, through incorporating a multi-stakeholder approach which includes a diverse community of voices, including those most affected or vulnerable groups. This multi-stakeholder and inclusive approach to enforcement is extremely important as only then will clear obligations for different actors involved in the AI technology stack be clarified, avoiding an opaque or self-referential regulatory system. Further, openness allows for the contestability of processes, with designed mechanisms in the law for openness to contestation. Further, improved cooperation between different actors involved in the AI life cycle can allow less powerful actors to be involved in the decision-making and in the design process. Given the potential for harm, the inclusion of these marginal voices in decision making should be a core priority for promoting openness.

In summary, the two above principles of transparency and openness in AI systems are pivotal in establishing a governance model that not only informs and empowers individuals but also encourages a communal approach to ethical AI development and usage. Together, these principles form the foundation for a governance framework that ensures AI technologies are developed and deployed in a manner that is ethical, transparent, accountable, and aligned with human values and rights.

I. Accountability

Accountability is an essential component in securing the rule of law in AI governance.[27] The principle of accountability relates to the responsibility of the AI provider, not only to ensure adherence to the principles of AI governance, but also to be able to demonstrate their compliance materially.[28] The accountability principle supposes that all principles which guide AI governance are not only considered

[27] Margot E Kaminski, ‘*Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability*’ (2019) 92 Southern California Law Review 1529 <https://papers.ssrn.com/abstract=3351404> accessed 22 April 2024.

[28] Susan von Struensee, ‘*Analyzing Dilemmas Posed by Artificial Intelligence and 4IR Technologies Requires Using All Available Models, Including the Existing International Human Rights Framework and Principles of AI Ethics*’ (25 June 2021) <https://papers.ssrn.com/abstract=3874279> accessed 22 April 2024.

by the processing party or balanced against competing principles but actually implemented in practical terms, the proof thereof lays in the hands of all actors involved throughout the AI system's value chain.

Accountability provisions are contained in the GDPR, the EU AI Act, and the OECD AI principles. Pursuant to Art. 5 (2) GDPR, accountability is one of the guiding principles of data processing. Similarly, in the EU AI Act, Article 11 and Annex IV contain references to accountability in the form of detailed documentation requirements about compliance with the Regulation's provisions. Also in the AI Act is Article 17, concerning the management systems of high-risk AI, which mandates that an "accountability framework setting out the responsibilities of the management and other staff…" should be elaborated.

The accountability principle guarantees the operability of the other principles and is an essential part of the rule of law within the context of AI governance. Without proper checks and balances, through transparency requirements and proof of compliance measures, the principles discussed in this paper fall short of meeting the challenges posed by AI in addition to not meeting the rule of law standard. Hence, the accountability principle bolsters the need for an authority that enforces, follows, and investigates potential infringements of the norms embodying this principle.

J. Data Governance

The OECD's report on "Going Digital to Advance Data Governance for Growth and Wellbeing"[29] defines data governance as arrangements comprising technical, policy, regulatory or institutional measures that affect each stage of data usage, including creation, collection, storage, use, protection, access, sharing, and deletion. The HLEG guidelines characterise data governance as measures taken to preserve the quality and integrity of data and facilitate access.

Data governance is a critical component of trust and reliability, validating the output generated by AI systems. The development of a robust data governance system has the potential to harmonize best practices and provide certainty for business development and trust vis-à-vis consumers. Conversely, uncertainties around proper data governance lead to inefficiencies and frequent data breaches.[30]

Given the importance of the concept, the AI Act underscores data governance as a key principle, especially concerning high-risk AI systems. It obligates providers to ensure that data governance and management practices are appropriate in the context and intended purpose of the AI system, beginning with design choices through each step, including examination of possible biases (see Article 10). The concept is also intrinsically embedded in the obligations around quality management systems (Article 17, AI Act) – as noted in Recital 44, high-quality datasets for training, validation, and testing also require the implementation of an appropriate data governance system and, therefore – the two concepts (data governance and quality management) have overlapping consequences.

Another example of the interplay of legislative obligations is where the AI Act proposes to build on the GDPR's conceptualisation of data governance (such as preserving the confidentiality and integrity of data noted in Article 5(1)(f)). In essence, the AI Act's obligations continue to apply in parallel with the

[29] OECD, '*Going Digital to Advance Data Governance for Growth and Wellbeing*' (Organisation for Economic Co-operation and Development 2022) <https://www.oecd-ilibrary.org/science-and-technology/going-digital-to-advance-data-governance-for-growth-and-well-being_e3d783b0-en> accessed 12 March 2024.

[30] '*Data Governance and the Board: Risk Advisory*' (Deloitte Singapore) URL: https://www2.deloitte.com/sg/en/pages/risk/articles/data-governance-and-the-board.html; 22 February 2024.

GDPR to the extent personal data is used. In this context, the incremental obligations set out around the use of special categories of data under Article 10(5) of the AI Act have been included to ensure the higher level of safeguards required under the GDPR are equally applicable in the AI Act.

K. Diversity

Equality as a fundamental right includes respect for cultural, linguistic, and religious diversity.[31] In relation to AI systems, diversity requires not just heterogeneous and comprehensive data sets but also gender balance and diversity in engineering teams. The HLEG guidelines also highlight this, noting that hiring from diverse backgrounds encourages diverse opinions. Additionally, developing AI systems to be user-centric and inclusive in design can assist in reaching a wider group of users, regardless of age, gender, abilities, or characteristics.

In the context of the AI Act, Recital 14a also refers to the involvement of diverse actors, to promote gender equality, accessibility, and cultural diversity. Recital 53a also contains a direct reference to the United Nations Convention on the Rights of Persons with Disabilities[32] and calls upon providers of AI systems to ensure compliance with accessibility requirements, including Directive (EU) 2016/2102 and Directive (EU) 2019/882 as well as universal design principles. Specific obligations have also been placed on the AI Office to ensure diverse perspectives and consultation with civil society and other stakeholders, for drawing up Codes of Practice and development of inclusive and diverse design (Recital 60s and Art. 69 AI Act).

L. Fairness and Non-Discrimination

Given that bias is one of the foremost concerns around the development of AI systems, the HLEG Guidelines identified fairness and non-discrimination as inextricably linked principles that lay the foundation of a trustworthy AI system. Fairness in the processing of personal data is also a key component of the GDPR (Article 5) and discrimination has often been identified as a direct risk to the rights and freedoms of natural persons, flowing from unfair use of personal data (Recital 75 GDPR).

The concept of fairness reappears in the AI Act. For example, it appears in the context of data governance, through examination and mitigation of unfair bias and discrimination prohibited by Union law (Article 10), awareness of automation bias and over-reliance on outputs produced by high-risk AI systems in the context of human oversight (Article 11), or bias in the feedback loop through improved accuracy and robustness (Article 12), etc.

Aside from manipulation, one of the direct results of unfair practices is discrimination. Therefore, amongst others, AI systems that may be used for social scoring or biometric categorisation systems have been categorised as prohibited systems. The aim is to prevent the perpetration of discrimination by such systems. The AI Act also links the concept of discrimination to Union law more broadly (refer to Article 10), thus ensuring that historical patterns of discrimination based on sexual orientation, gender, age, disabilities, race, and ethnicity, etc., are mitigated.

[31] Article 22, European Union, *Charter of Fundamental Rights of the European Union (2007/C 303/01)*, C 303/1, 14 December 2007.

[32] Convention on the Rights of Persons with Disabilities (adopted on 13 December 2006, entry into force on 3 May 2008), UNTS Volume Number, 2515 (p.3). 2515 UNTS 3

M. Societal and Environmental Wellbeing

The principle of societal and environmental wellbeing is key in the context of AI governance given that this is a technology capable of changing established economic and societal relations.[33] The OECD AI principles already establish that AI actors "should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes" for users as well as for the planet.

This principle suggests that underrepresented or vulnerable groups should be included in the AI governance architecture. The EU AI Act includes several provisions that take into account the impact on "vulnerable groups" or "groups of persons" in general. For example, Annex IV of the EU AI Act includes a provision (2(b)) related to the inclusion of "persons or groups of persons on which the system is intended to be used". Societal impact can also be achieved by including consideration of the people who work on creating the AI systems. This would suppose that this regulation applies is in the domain of working conditions for those who participate in the development of AI systems.

There is a need to ensure the energy grid can cope with the high levels of energy required to deploy this technology at scale, with special consideration for the environmental impact of building these AI systems. The OECD AI principles and the EU AI Act include provisions requiring environmental sustainability assessment measures (Article 84, EU AI Act). Hence, a principle for societal and environmental wellbeing is necessary to ensure that AI systems contribute to a more inclusive, equal, and sustainable society.

[33] Xinyue Hao and Emrah Demir, 'Artificial Intelligence in Supply Chain Decision-Making: An Environmental, Social, and Governance Triggering and Technological Inhibiting Protocol' (2023) 19 Journal of Modelling in Management 605 <https://doi.org/10.1108/JM2-01-2023-0009> accessed 22 April 2024., David Rolnick, et al. 2022. Tackling Climate Change with Machine Learning. ACM Comput. Surv. 55, 2, Article 42 (February 2023), 96 pages. <https://doi.org/10.1145/3485128>, accessed 30 April 2024

Critical Reflections

The intersection of AI regulation and the complexities surrounding global governance, present a nuanced landscape that requires a critical examination of existing guidelines and frameworks. Although the guidance from CNIL is admittedly valuable, its sufficiency in addressing the intricate challenges of large AI systems, particularly concerning the reuse of personal data and the principle of purpose limitation (see principle B above) must be questioned. Such challenges are accentuated in the context of AI systems, with the phenomenon of function creep (i.e., the expansion of a system or technology beyond its original purposes) posing significant risks to privacy and ethical standards. In 2023 significant advancements in AI regulation were made, notably through the political agreement on the AI Act in Europe and collaborative efforts at the G20 and GPAI summits. These developments underscored a growing consensus on the need for safe and trustworthy AI. However, the pursuit of global regulatory convergence faces obstacles, such as the digital divide, cultural differences, and economic priorities, especially in the global South. India's focus on digital innovation, despite its privacy legislation, illustrates the complex trade-offs between economic advancement and privacy concerns.

From a political perspective, the enforcement of AI principles such as accuracy and cybersecurity is heavily reliant on infrastructure, posing additional challenges for countries in the Global South with limited internet penetration. The disparate approaches to AI risk management, as seen in the frameworks of the US NIST[34] and ISO/IEC 23894:2023(E),[35] reflect cultural and geographic differences, which evidence the international standardization community's drive for inclusion and collaboration across stakeholders for an effective development of trustworthy AI.

The infrastructural prerequisites in terms of technical preparedness but also administrative capacity for adhering to the AI Act's standards highlight the difficulty in establishing globally interoperable standards. The potential lack of representation from the global South in these discussions raises concerns about the inclusivity and feasibility of a universal AI governance framework.

AI's capacity to disrupt economic and social structures, as evidenced by its implications for labour dynamics and social unrest,[36] necessitates a nuanced approach to policymaking. The challenges and opportunities presented by AI automation vary significantly across economies, with vulnerable groups in less prepared economies facing increased risks of displacement and unrest. Policymakers are confronted with the choice between adopting a fragmented approach, allowing for greater flexibility and autonomy in policy implementation, or pursuing a universal approach that emphasizes regulatory convergence, a level playing field for all the actors throughout the development or deployment chains, and the protection of human rights.

The argument for a universal approach is strengthened by the need for common principles that guide the ethical development and deployment of AI, ensuring fairness, transparency, and accountability. Such a global framework would support developing actors, reduce moral outsourcing, and provide a

[34] https://www.nist.gov/

[35] https://cdn.standards.iteh.ai/samples/77304/cb803ee4e9624430a5db177459158b24/ISO-IEC-23894-2023.pdf

[36] Daron Acemoglu, Pascual Restrepo, The wrong kind of AI? Artificial intelligence and the future of labour demand, *Cambridge Journal of Regions, Economy and Society*, Volume 13, Issue 1, March 2020, Pages 25–35, <https://doi.org/10.1093/cjres/rsz022>, accessed 30 April 2024

collective response to the challenges posed by AI. Prioritizing principles such as lawfulness, privacy, accountability, transparency, and sustainability is crucial for establishing a democratic and ethical AI governance structure that respects both technological advancements and human rights.

In summary, the critical examination of AI regulation and governance reveals a complex interplay of technological, ethical, and political factors. Achieving regulatory convergence and developing a universally accepted framework for AI governance requires inclusive dialogue, collaborative efforts, and a commitment to ethical principles that transcend geographic and cultural boundaries.

Operationalising AI Adequacy Principles

The principles of AI adequacy, as outlined above, emphasize the importance of lawfulness, purpose limitation, transparency, accountability, data governance, privacy and security, safety, agency oversight, diversity, non-discrimination as well as sustainability, and societal wellbeing in AI governance. These principles are essential in creating a regulatory framework that balances technological advancements with ethical oversight and societal wellbeing. Their implementation by governmental or industry players will have a crucial role in becoming guidelines for responsible development and deployment of AI systems.

The following categorisation of the principles aims to provide concrete and substantive direction in the pursuit of achieving global convergence. This analysis therefore contains various sources, including General Data Protection Regulation (GDPR), the European AI Act, and work from international organizations such as the OECD.

The table below represents in a summarized version the principles discussed above. We gathered the principles into three main axes, grouping them based on similarity, and recommending tools for the operationalization of those principles based on this grouping. The axes are **rule of law** (red), **proportionality** (green), and **human-centricity** (yellow). The operationalisation of these principles represents the ideal implementation of the broader principle of shared responsibility, which is also reflected in our analysis below.

Key Axes of the Principles

Lawfulness	Rule of law refers to the principles of **lawfulness, fairness, transparency, and accountability**. GDPR Art. 5(1) as a guide for all data processing contains an extended definition of the principle of "rule of law". The EU AI Act explicitly refers to GDPR in several places, analogous to "lawful data processing".
Fairness and Accuracy	**Fairness and non-discrimination** are inextricably linked principles (HLEG Guidelines). The EU AI Act addresses mitigation measures for bias, overreliance, and improved accuracy (Articles 10-12). Fairness is a key component of the GDPR (Art. 5) and requires reasonable steps taken to rectify records without undue delay, to avoid misleading or inaccurate processing.
Openness	**Openness and open AI systems** include an emphasis on improved accessibility to AI (for e.g. lower obligations for open-source, R&D), while transparency obligations that not only help end users but also

	deployers. Article 13 of the EU AI Act also is a tool to control the use of data and to hold accountable those who lawfully acquired it.
Accountability	Accountability relates to the responsibility of the AI provider not only to ensure adherence to the principles of AI governance but also to be able to demonstrate material compliance.
Purpose Limitation	There are two aspects relevant for AI: limitation of data used in input/training, and of the deployment of the AI system itself. While purpose limitation is not explicitly included within the EU AI Act, the categorisation of high-risk AI systems is based on the 'intended purpose' of their usage (Art. 7(2)(a) AI Act).
Data Governance	EU AI Act underscores data governance, especially concerning high-risk AI systems, obliging providers to ensure that data governance and management practices are appropriate to the intended purpose, beginning with the examination of biases in design choices (Art. 10). Data governance is also intrinsically embedded in the obligations on quality management systems (Art. 17 AI Act) and builds upon the GDPR's interpretation.
Privacy and Security	OECD AI principles protect fundamental rights, including privacy. Requirements for the protection of individuals' privacy and personal data include data minimisation, storage limitation, integrity, and confidentiality.
Safety, Agency Oversight	Safety is central to AI governance, both product safety and guardrails against harms (misrepresentation, discrimination, etc.). Technical solutions and the inclusion of human oversight help ensure safety and human agency.
Diversity & non-discrimination	Discrimination is considered a direct risk to the rights and freedoms of individuals across a range of legislative frameworks. In relation to AI systems, the avoidance of bias requires not just diverse and comprehensive data sets but also gender balance and diversity in development teams. The EU AI Act refers to the involvement of diverse actors, to promote gender equality, accessibility, and cultural diversity, and refers to other EU directives and UN Conventions.

Sustainability & Societal Wellbeing	OECD AI principles already establish a requirement that AI actors strive for beneficial outcomes for users as well as for the planet, and along with EU the AI Act includes provisions on environmental sustainability and on "vulnerable" groups (Art. 84, AI Act).

Table 1

Essential elements (toolkit) for operationalising principles of AI global governance:

1. Rule of law:

The GDPR emphasizes the importance of transparency, which extends to AI given the legislative interplay described above, requiring that individuals be informed about how their data is being used. It also requires an additional level of transparency when it comes to automated systems where users should be informed about the logic involved and the consequences they may bear. This principle is also reflected in the EU AI Act, which mandates that high-risk AI systems provide information to users that enables them to interpret the system's output. The principles of openness and transparency have been demonstrated above, and consequently, so have the underpinning principles of lawfulness, fairness, and accuracy, as well as accountability. Hence, when we refer to open AI systems (see table above), we precisely mean systems which are subject to the standards outlined in the abovementioned core principles (in red). Nevertheless, all this does not mean that private parties rights such as Intellectual property should be forsaken. Quite the opposite a balance between interests is possible.

Strong checks and balances through open AI systems do not and should not entail a derogation of any existing intellectual property rights as a driving factor of innovation and investments in automated systems. Thus, while fostering open development and deployment of automated systems, it should be ensured that involved actors throughout the research, development, production, supply and deployment chain are capable of reaping the benefits of their investments. This includes specific policy developments in the realm of intangible rights such as trade secrets (algorithms), trademarks (design) and copyright (software). Although further discussion of the necessity to regulate intellectual property vis-à-vis principle-based AI development goes beyond the scope of this document, we underline the importance of the balance between the overarching interests of an open AI system and the sustainability of the investments made by business actors.

Openness is a fundamental component of "rule of law" and is also included in this axis. On a global level, the heterogeneity of standards and statutes invites the inclusion of this principle in globally accepted documents. The currently under development G7 AI Toolkit and Compendium of Digital Public Services is an example of such an international policy document which has the potential to foster openness in AI on a larger scale. Other publicly available outputs resulting from global efforts to provide guidelines on the matter are the OECD AI Principles, the UN AI Advisory Body Interim Report, and the Hiroshima AI Process. Despite being relatively recent, the advancement of technology as well as the regulatory progress in some jurisdictions (for example, the AI Act) evidences the recognition that the principles making up AI openness are integral to the amendment of those documents and are indispensable for the negotiations within those initiatives.

While general guidelines are the core of a principle-based AI, standardization work in fora such as the International Organization for Standardization (ISO) can develop those principles into applicable bylaws for industry and public actors. The inclusion of these principles in a uniform and standardised way ensures a global acceptance and application, thus providing a converging and certain background for innovation and economic growth.

2. Proportionality

The EU AI Act categorizes high-risk AI systems based on the intended purpose of their usage, with the bigger the risks, the bigger the obligations and duty of care. This principle is dynamic and assessed through participative methods, such as the G7 Hiroshima Principles. Proportionality in AI governance means that the level of oversight, regulation, and compliance requirements should be commensurate with the risks posed by the AI system. In line with the previous section, the proportionality axis suggests a differentiated approach towards different risk levels that these AI systems pose, in order to ensure not only higher level of data governance where it is most needed (high-risk systems) and thus focusing compliance efforts where the impact on users is higher,[37] but also to make sure that industry and government efforts are focused. This means that actors throughout the AI's development and deployment chain would profit from reduced costs and compliance efforts. While the specific efficiency or economic implications escape from the scope of this document, we suggest that the proportionality principle should be an inherent part of the current global initiatives around AI because of the advantages discussed throughout this text.

This objective, however, could be achieved by means of a global consensus converging towards the necessity to focus AI compliance efforts. Further development of the G7 Toolkit as well as the OECD Toolkit should establish a methodology for assessing the risks posed by AI systems based on their intended purpose. The toolkit should also provide guidelines for dynamic risk assessment and encourage participative methods, such as the G7 Hiroshima Principles, to ensure that the level of oversight and regulation is proportional to the risks posed by the advanced AI system. The example of the AI Act could serve as a guideline for this evolution which policymakers could integrate and build on.

3. Shared responsibility

The EU AI Act stipulates that high-risk AI systems must be designed to enable effective human oversight, with responsibility placed on providers to ensure that data governance and management practices are appropriate for the context and intended purpose of the AI system. This principle also involves clear, enforceable obligations differentiating various actors through precise allocation of compliance responsibilities and potential liabilities. This axis plays a crucial role in the enforcement of the discussed principles, as it would foster an environment where private and public actors are able to allocate liabilities clearly and thus ensure an environment of legal and economic certainty. AI system deployers and developers would benefit from a unified set of operational rules, while users could expect consistent accountability for identical adverse effects, regardless of the location or actors involved.

Shared responsibility in AI governance acknowledges that multiple stakeholders, including governments, industry players, civil society organizations, and academic institutions, have a role to

[37] Such as users' privacy vis-à-vis the adverse effects of automated systems

play in ensuring ethical AI development and deployment. Therefore, common local, regional, and international collaborative initiatives should be supported. For the sake of this analysis, we have examined initiatives such as the G7 AI Principles and the OECD AI Policy Observatory. They provide platforms for fostering dialogue and coordination among different stakeholders and interest groups. The UN backed Internet Governance Forum's Policy Network on Artificial Intelligence (PNAI) initiative represents another example thereof. As mentioned before, ISO standardization efforts and the ensuing international collaboration is another example. In addition, the World Economic Forum's Global AI Action Alliance brings together governments, companies, and civil society organizations to address the ethical and societal implications of AI. Governments can facilitate collaboration that encourages information sharing and cooperation amongst stakeholders. Looking forward into the future of AI governance, we could envisage an organization replicating the successful model and multi-stakeholder character of the Internet Corporation for Assigned Names and Numbers (ICANN), which for the past couple of decades manages to balance public and private interests throughout times of revolution (internet), evolution (internet expansion), and crisis (wars and humanitarian crisis).

4. Human centricity

Human oversight is paramount for two main objectives. First, human oversight is a precondition for the reduction of the adverse effects AI deployment may entail, such as skewed decisions based on distorted and biased data, as well as discrimination. Second, human oversight secures a human application of fundamental rights and their assessment. This is the rationale behind the "human involvement" requirement for automated systems' decisions in the GDPR and fundamental rights impact assessment in the AI Act, beyond the technical compliance requirements of the conformity assessments. Third, human oversight fosters trust in AI systems vis-à-vis the larger adoption of the technology.

In order to effectively ensure human involvement, individuals tasked with oversight should possess adequate skills, training, and authority. This ensures that AI systems operate within defined constraints, remain subordinate to human directives, and that individuals can exercise control over the development and deployment chains. Despite some persistent challenges related to the issue of preventing human bias reinforcement, initiatives adopting a human-centric approach allow for the effective implementation of the previous set of principles, because the involvement of individuals in the development and deployment of AI supposes an increased level of accountability and responsibility allocation, thus clearly in accordance with the previous set of principles.

Methodologies for incorporating human-centric design principles include user-centred design approaches, participatory stakeholder engagement processes, and impact assessments that consider the potential social, economic, and ethical implications of AI technologies. We consider that the development of the OECD and G7 Toolkits, the Hiroshima Process, as well as similar initiatives, should not only ensure an international forum for discussion of AI principles, but should also serve to establish requirements for a human-centric approach governed in accordance with a globally accepted set of values and principles, enhancing existing consensus. Along with public efforts, organizations such as the ISO, the Institute of Electrical and Electronics Engineers (IEEE), and the International Telecommunication Union (ITU) develop standards and guidelines for AI ethics, safety, and interoperability which already include this approach. For example, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems provides a framework for ethical AI design and development.

Conclusion

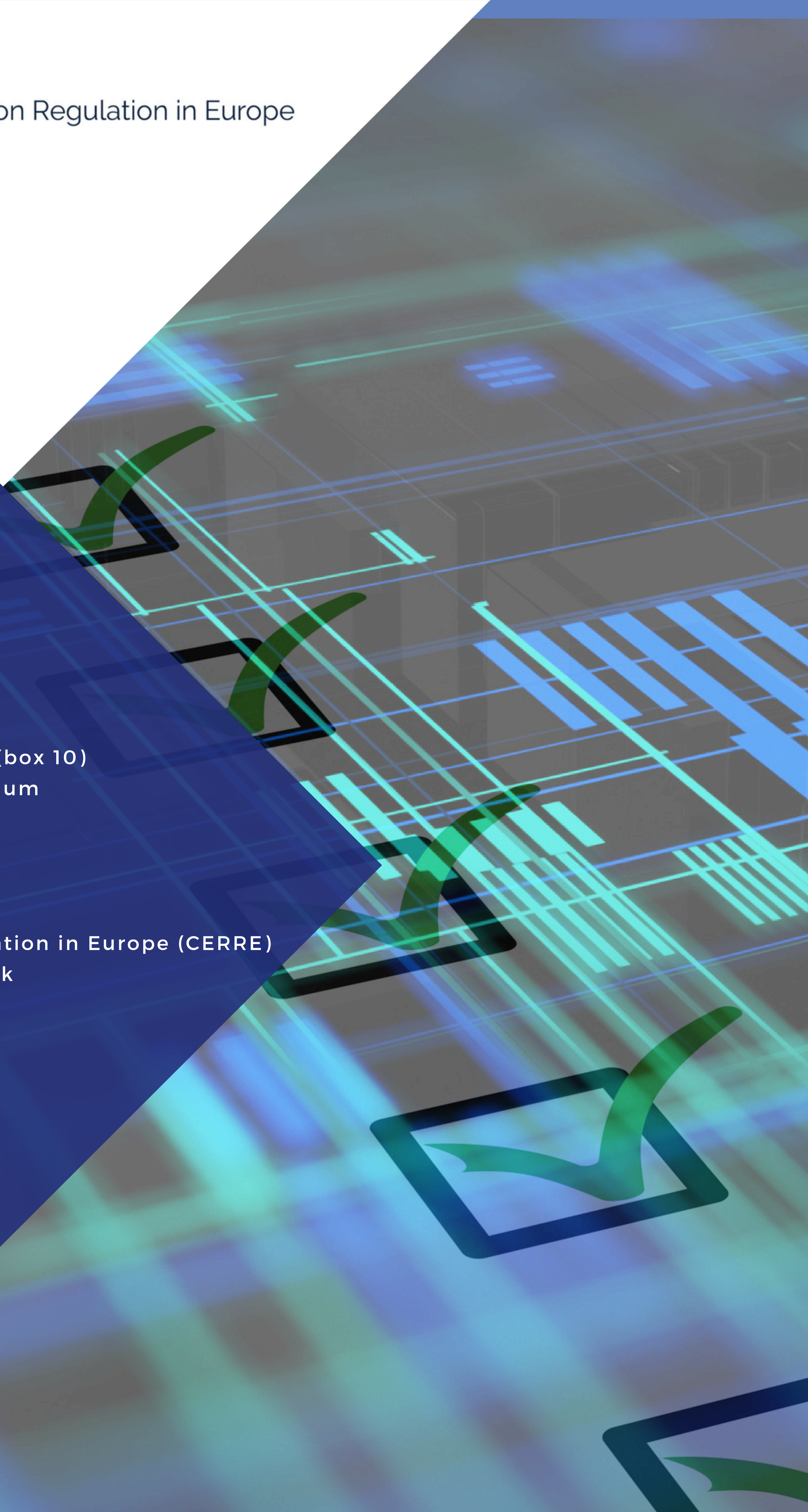
This report aims to provide a comprehensive overview of the principles that are essential for the operationalisation of AI systems in a manner that respects the rule of law, proportionality, human-centricity, and establishes a shared responsibility. These principles are derived from existing regulations such as the General Data Protection Regulation (GDPR) and the European AI Act, as well as international organizations like the OECD. The four axes identified in the text are intended to guide the development of globally accepted documents and policies that foster openness, proportionality, human-centricity, and shared responsibility in AI governance. The principles aim to ensure a global convergence towards a more transparent, accountable, and ethical AI ecosystem. The report also emphasizes the importance of balancing the overarching interest of an open AI system with the sustainability of investments made by business actors. It suggests that the principles should be included in globally accepted documents and standardized to ensure a uniform and standardized application, which would provide a converging and certain background for innovation and economic growth. The text also highlights the need for proportionality in AI governance, which means that the level of oversight, regulation, and compliance requirements should be commensurate with the risks posed by the AI system. The text also emphasizes the importance of shared responsibility, which involves clear, enforceable obligations, differentiating various actors through precise allocation of compliance responsibilities and potential liabilities. The text concludes by emphasizing the need for collaboration among stakeholders, including governments, industry players, civil society organizations, and academic institutions, to ensure ethical AI development and deployment.

However, the path towards a holistic AI adequacy requires a careful accommodation of the technological prowess of AI and its societal implications. The discourse herein underscores the pivotal role of regulatory convergence in fostering an environment where AI can be developed and deployed ethically, transparently, and equitably. The principles identified above could help mitigate some of the risks— from the intricacies of data usage and purpose limitation to the disparities in technological infrastructure and the divergent priorities between economic advancement and privacy — and these illustrate the complexity of crafting a universally acceptable AI regulatory framework.

The juxtaposition of various initiatives, and the critical analysis of their effectiveness reveal a consensus on the necessity of a collaborative and inclusive approach in order to achieve a consistent AI adequacy. Such an approach must not only acknowledge but also actively address the disparities and nuances inherent in the global landscape. The pursuit of universal principles for AI governance, while daunting, is imbued with the potential to establish a balanced ecosystem that respects human rights, fosters innovation, and ensures equitable benefits from AI advancements.

In conclusion, the path towards an AI adequacy is iterative and evolving. It requires the engagement of stakeholders across the spectrum — from policymakers and industry leaders to academics and civil society. The collective endeavour to align on core principles and standards, while navigating the complex interplay of technical, ethical, and political factors, is fundamental to realizing the promise of AI. In this journey, the shared commitment to ethical principles, transparency, and inclusivity stands as the cornerstone for a future where AI serves as a catalyst for positive change, enhancing human capabilities while safeguarding the rights and dignity of individuals across the globe.

cerre Centre on Regulation in Europe



Avenue Louise 475 (box 10)
1050 Brussels, Belgium
+32 2 230 83 60
info@cerre.eu
www.cerre.eu

Centre on Regulation in Europe (CERRE)
CERRE Think Tank