PART 2: THE EMERGENCE OF REGIONAL SOLUTIONS

6. The Incipient Latin American Approach to AI Governance: Highlighting Data Governance Issues through Emerging Supervisory Authorities

Pablo Trigo Kramcsák, Bárbara Lazarotto and Rocco Saverino

Abstract. Influenced by global trends, particularly the European Union's (EU) digital regulations, Latin American countries are starting to incorporate artificial intelligence (AI) rules into their data protection frameworks while exploring comprehensive AI laws.
This paper examines the emerging AI regulations in Latin America (LatAm), highlighting diverse approaches in countries such as Brazil and Chile, where the establishment of specialised AI regulatory bodies reflects the region's awareness of the complex issues these technologies present. The analysis emphasises data governance as a key factor in shaping AI oversight. As LatAm refines its approach to AI regulation, the region is well-positioned to contribute to the global discourse on AI governance.
Keywords: AI regulation, personal data protection, supervisory authorities.

Introduction

AI systems have rapidly emerged as a transformative technology. As these models evolve and their applications expand, coherent regulatory responses have become urgent. Around the world, countries are racing to establish AI regulations, often drawing inspiration from landmark legal frameworks like the EU's AI Act.

In LatAm, the journey toward AI governance has begun. However, these efforts remain in the early stages, marked, inter alia, by integrating AI governance into existing data protection frameworks, adopting a risk-based approach (classifying AI systems into different risk categories), creating new supervisory authorities, and emphasising data governance challenges.

This work analyses the region's adaptation to and engagement with global trends in AI regulation, with particular attention to data governance and supervisory authorities. It analyses emerging AI laws and regulatory frameworks in Brazil and Chile to present the region's challenges and opportunities in building an effective AI governance model.

6.1. Global Influence: The European Union's AI Regulatory Framework

The EU AI Act represents a pioneering legal framework, distinguished by its comprehensive, human-centric, and risk-based approach (Kusche, 2024). This Act has significantly shaped global discussions on AI governance. The EU's influence extends beyond its borders, primarily through what is known as the "Brussels Effect" (Bradford, 2020), where its regulations, such as the General Data Protection Regulation (GDPR), have set global standards that other regions often emulate (Greenleaf, 2021).

Like the GDPR, the AI Act is designed with extraterritorial reach (Hacker, 2023), meaning its impact is felt even in countries not part of the EU. This is particularly relevant for LatAm, where countries have historically aligned their data protection laws with the European legal approach.[48]

[48] See, e.g., Gadoni Canaan, 2023.

The AI Act's emphasis on data governance, transparency, and accountability is expected to have a similar influence on regional AI regulations. However, while the AI Act is setting the pace for global AI governance, the extent to which Latin American countries will replicate this model remains uncertain.

This uncertainty is closely linked to the situation in Europe, where each state's approach to relying on existing data protection authorities (DPAs) or establishing new AI authorities does not contribute to a harmonised framework. One of the most challenging aspects of enforcing the AI Act is the role of DPAs alongside AI authorities, particularly considering the potential variance in the structure of competent authorities from country to country. Even during the proposal stage of the AI Act, there was an apparent broadening of the supervisory framework within the GDPR (Chamberlain & Reichel, 2023). Given the close connection between data and AI systems, cases of overlapping and confusion regarding the competency of DPAs or AI authorities are always possible.

6.2. The Rise of AI Authorities in Latin America

Latin American countries are beginning to establish their own AI frameworks, which have been influenced by the EU[49] but tailored to their specific contexts. One of the critical aspects of these emerging frameworks is the creation of supervisory authorities responsible for overseeing AI systems. These efforts are still nascent, and there is considerable variation in how countries perceive these authorities.

For instance, Brazil's AI Law Proposal No. 2338/2023 outlines the creation of a National System of Regulation and Governance of Artificial Intelligence (SIA), which includes a network of authorities such as the Brazilian Data Protection Authority (ANPD), state AI regulators, and other entities responsible for AI certification and self-regulation. This multifaceted approach reflects Brazil's recognition of the complexity of AI governance and the need for a collaborative framework involving multiple stakeholders. More recently, the Brazilian Data Protection Authority (ANPD) issued an opinion on the bill, emphasising that the overlap between Brazil's General Data Protection Law (LGPD) and the AI governance framework could not be overlooked. Therefore, the ANPD should play a leading role in AI governance. After that, the bill was modified, and the ANPD was designated as SIA's coordinating authority.[50]

In parallel, Chile is advancing its AI governance model through Bill No. 16821-19, which proposes establishing an AI Technical Advisory Council to guide the Ministry of Science, Technology, Knowledge, and Innovation. This council will be complemented by the Data Protection Agency, which will enforce the AI law once it is established under forthcoming legislation to modernise Chile's data protection framework.

These examples illustrate LatAm's varied approaches to AI governance, where existing data protection authorities are being reconfigured to take on AI oversight or new bodies are being created altogether.

[49] The major influence is Spain, with very close links to Latin American countries. Indeed, it is a member of the Ibero-American Data Protection Network and the Permanent Secretary. Furthermore, Spain was the first country to establish an AI authority independent of the existing data protection authorities: the Spanish Artificial Intelligence Supervisory Agency (AESIA).

[50] Análise preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial. Available at https://www.gov.br/anpd/pt-br/assuntos/noticias/analise-preliminar-do-pl-2338_2023-formatado-ascom.pdf

However, Latin American countries are exploring a broader spectrum of adaptation, ranging from close emulation of the EU model to more independent strategies.

6.3. Data Governance: A Central Issue in AI Regulation

Data governance is a crucial component of AI regulation, given that AI systems rely heavily on data for their development and deployment. The EU AI Act underscores the importance of data governance, emphasising transparency, accountability, and the protection of fundamental rights. This focus on data is mirrored in the emerging AI regulations in LatAm, where data protection remains a central concern.

In many Latin American countries, AI regulation efforts are closely linked to their data protection approaches, reflecting the influence of the GDPR. The GDPR's significant impact on crucial aspects of AI systems, such as big data processing, profiling, and automated decision-making, should be noted.

For example, Brazil's LGPD, which mimics the GDPR (Erickson, 2019), plays a significant role in the country's AI governance framework. The LGPD's principles of transparency, purpose limitation, adequacy, necessity, prevention, data quality, non-discrimination and accountability are expected to extend to AI systems (Belli et al., 2023), ensuring that they operate within a framework that prioritises protecting personal data.

Chile's approach to AI governance also highlights data protection issues. A yet-to-be-established Data Protection Agency will oversee the proposed AI law. This approach underscores the importance of data governance in AI regulation, as the effectiveness of AI oversight will largely depend on the robustness of the underlying data protection framework.

Nonetheless, effective AI regulation faces steep data governance hurdles (fragmented data protection laws, uneven institutional capacity, and the struggle to balance innovation with fundamental rights). Additionally, integrating AI governance into existing frameworks raises a critical question: Do current data protection authorities have the expertise and resources to oversee AI systems effectively?

6.4. Challenges and Opportunities in Latin American AI Governance

Developing AI governance frameworks in LatAm presents challenges and opportunities. On the one hand, the region can draw on the experiences of other regions, such as the EU.[51] On the other hand, Latin American countries must navigate a complex landscape of political, economic, and social factors that are very different from those of European countries, which may hinder the implementation of such frameworks.

The Latin American social and political environment makes these countries vulnerable to abuse through AI systems. For instance, facial recognition technology in Brazil is often used as a security measure due to the country's high number of crimes. However, this technology often infringes on individuals' human and fundamental rights, a concern that must be carefully addressed in Latin American AI laws.[52]

Therefore, one of the main challenges is the need for coordination among different regulatory bodies. The creation of multiple supervisory authorities, as seen in Brazil's AI Law Proposal, can lead to

[51] See, e.g., Novelli et al. 2024.
[52] See Ramiro & Cruz, 2023.

fragmentation, inefficiency, and potential rights violations if these authorities do not work together effectively. Ensuring that these bodies have clear mandates and mechanisms for coordination and collaboration will be crucial for the success of AI governance in the region.

Another challenge is the need for sufficient resources and expertise. Many Latin American countries face limited institutional capacity, which could weaken the effectiveness of AI regulation. Developing the necessary expertise within supervisory authorities, particularly in the technical aspects of AI, will be highly relevant for effective oversight. Moreover, securing the financial support to sustain these authorities is a critical challenge, particularly in countries with tight public budgets.

Despite these challenges, Latin American countries have significant opportunities to shape AI governance proactively. Adopting a risk-based approach, like the EU AI Act, allows the development of AI regulations that balance innovation with the protection of fundamental rights. Additionally, integrating AI governance into existing data protection frameworks enables the region to leverage its data protection experience, ensuring AI systems operate transparently and accountably.

LatAm also could contribute to the global discourse on AI governance by developing regulatory models that reflect its unique social, economic, and cultural contexts. While the region may draw inspiration from the EU, it is well-positioned to innovate and develop flexible and structured approaches that address AI's specific challenges and opportunities in the Majority World. For instance, the region's emphasis on social justice and human rights could lead to developing AI regulations that prioritise protecting vulnerable populations and promoting equitable access to AI technologies.

6.5. The Path Forward: Toward a Coherent AI Governance Framework

As Latin American countries continue to develop their AI governance frameworks, several key issues must be addressed to ensure the effective regulation of AI. First, there is a need for greater harmonisation of AI regulations across the region. While the diversity of approaches reflects the different contexts of each country, a more coordinated approach could help address cross-border issues and promote regional collaboration in AI governance. Harmonisation does not necessarily mean uniformity but rather the alignment or convergence of key principles and standards to ensure a consistent approach to regional AI regulation.

Second, data protection authorities' role in AI governance must be clearly defined. It is essential to ensure these authorities are equipped with the necessary expertise and resources to effectively oversee and regulate the data processing aspects of AI systems. This may require capacity-building initiatives, increased funding, and the development of new regulatory tools and methodologies specific to AI.

Third, there is a need for greater public engagement and transparency in developing new AI governance frameworks. AI regulation should not be a top-down process; instead, it should involve a broad range of stakeholders, including civil society, industry, academia, and the public. Public engagement can help build trust in AI systems and ensure that AI regulations reflect the values and concerns of society. Additionally, transparency in the regulatory process can help ensure that AI governance is accountable and that the decisions made by policymakers and regulatory authorities are open to public scrutiny.

Finally, Latin American countries should consider the potential for regional cooperation in AI governance, being necessary to improve the incentives and conditions that allow collaboration in this

area, for example, overcoming the transaction costs associated with AI governance and regulation (Contreras, 2024). Initiatives such as creating a regional AI governance framework could help coordinate regional efforts and promote sharing best practices. Regional cooperation could also enhance the region's ability to engage in the global discourse on AI governance and ensure that Latin American perspectives are represented on the world stage.

6.6. CONCLUSION

The incipient Latin American approach to AI governance reflects the region's recognition of the importance of regulating AI systems in a manner that aligns with global standards while addressing local needs and contexts. While still in its early stages, this approach is marked by a growing awareness of the critical role that data governance plays in the effective oversight of AI technologies. Drawing from the foundation established through data protection laws, Latin American countries are starting to establish supervisory authorities capable of addressing the unique challenges posed by AI.

However, the region faces significant challenges, including the need for greater coordination among regulatory bodies, developing specialised expertise, and allocating sufficient resources to support effective oversight. Additionally, integrating AI governance into existing frameworks raises important questions about DPAs' capacity to manage the complexities of AI regulation.

LatAm has substantial potential to shape the AI governance debate. A proactive and regionally coordinated approach would enable the region to contribute significantly to the global regulatory conversation while safeguarding citizens' rights, emphasising principles such as social justice, equity, and human rights.

References

Belli, L, Curzi, Y. & Gaspar, W. B. (2023). AI regulation in Brazil: Advancements, flows, and need to learn from the data protection experience, Computer Law & Security Review 48, 105767, https://doi.org/10.1016/j.clsr.2022.105767.

Bradford, A. (2020). The Brussels Effect: How the European Union Rules the World. New York: Oxford University Press.

Chamberlain, Johanna, & Reichel, Jane. (2023). Supervision of Artificial Intelligence in the EU and the Protection of Privacy. FIU Law Review, 17(2), 267-286.

Contreras, P. (2024). International Convergence and Own Paths: Regulation of Artificial Intelligence in Latin America. Actualidad Jurídica Iberoamericana 21, 468-493.

Erickson, A. (2019). Comparative Analysis of the EU's GDPR and Brazil's LGPD: Enforcement Challenges with the LGPD. Brook. J. Int'l L 44, 859.

Gadoni Canaan, R. (2023). The effects on local innovation arising from replicating the GDPR into the Brazilian General Data Protection Law. Internet Policy Review, 12(1). https://doi.org/10.14763/2023.1.1686

Greenleaf, G. (2021). The "Brussels Effect" of the EU's "AI Act" on Data Privacy Outside Europe. Privacy Laws & Business International Report 117(1), 3-7.

Hacker, P. (2023). AI Regulation in Europe: From the AI Act to Future Regulatory Challenges. arXiv <https://arxiv.org/abs/2310.04072>

Kusche, I. (2024). Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk. Journal of Risk Research, 1–14. https://doi.org/10.1080/13669877.2024.2350720

Novelli, C., Hacker, P., Morley, J., Trondal, J. & Floridi, L. (2024). A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities. Available at SSRN: https://ssrn.com/abstract=4817755 or http://dx.doi.org/10.2139/ssrn.4817755.

Ramiro, A. & Cruz, L. (2023). The grey-zones of public-private surveillance: Policy tendencies of facial recognition for public security in Brazilian cities. Internet Policy Review 12 (1), 1-28.

7. THE RICE GOVERNANCE FRAMEWORK: ENABLING COMPREHENSIVE DATA GOVERNANCE IN AFRICA

CHINASA T. OKOLO, PH.D.

Abstract. New complexities around data production, refinement, and use have impacted African countries, elevating a need for comprehensive data regulation and enforcement measures. While 38/55 African Union Member States have existing data protections, there is a wide disparity in the robustness of these regulations and in the ability of individual countries to enforce these respective protections. This work introduces the RICE (Reformation, Integration, Cooperation, & Enforcement) Data Governance Framework, which aims to operationalize comprehensive data governance in Africa by outlining best measures for data governance policy reform, integrating revamped policies, increasing continental-wide cooperation in AI governance, and improving enforcement actions against data privacy violations.
Keywords. Data privacy, data governance, policy reform, African development, artificial intelligence

INTRODUCTION

The advent of generative artificial intelligence (AI), increasing adoption of AI tools, and the widespread utilization of data workers have changed narratives around data production and use. While data protections exist in 38 out of 55 African Union (AU) Member States, intensifying algorithmization across Africa could impact users through digital platforms used to access education, healthcare, financial, and social services. Given these new complexities and the emerging AI regulatory environment within the continent, African governments must enact comprehensive data protection regulations and reform existing data governance measures to cover aspects such as data quality, privacy, responsible data sharing, transparency, and data worker labor protections. To address these issues, data workers in Kenya have pursued litigation against Facebook regarding subpar working conditions and unfair termination (Musanga, 2023), and data workers across the continent have established organizations such as Techworker Community Africa (TCA)[53], the African Content Moderators Union, the Nigerian Content Moderators and Tech Workers Union (NCMTW)[54], and the Kenyan Content Moderators' Union. Along with general subpar working conditions across the continent in fields such as oil production and garment manufacturing, the concerns imposed by data work underscore requirements for sectoral reform of existing labor protections in areas including agriculture, economics, education, and healthcare. African countries also have context-specific challenges that differ significantly from those within the West, highlighting a need to understand how to develop culturally aligned and feasible governance solutions (Okolo, 2023).

By balancing lessons from the recent ratification of the African Union Convention on Cyber Security and Personal Data Protection, maturing regulatory environments like the EU, and advancing research on regional and country-specific needs, African nations can work towards more robust regulation. This paper analyzes data governance measures in Africa, outlines data privacy violations across the continent, and examines regulatory gaps imposed by a lack of comprehensive data governance to outline the sociopolitical infrastructure required to bolster data governance capacity. Additionally, it proposes the RICE (Reformation, Integration, Cooperation, & Enforcement) Data Governance framework, which African national governments (NGs), Regional Economic Communities (RECs), and

[53] https://techworkercommunityafrica.org/
[54] https://www.linkedin.com/company/nigerian-content-moderators-acmu/

the African Union can leverage to reform and operationalize existing data protection measures. Ultimately, this framework could inform the development and implementation of context-specific AI regulation that centers data privacy rights.

Data Protection Regulation in Africa

The increasing development and adoption of AI have dramatically shifted practices around data, spurring the development of new industries and revealing new forms of exploitation. This has also introduced gaps within existing data protection regulations that could be further exploited as AI development increases throughout the continent. While companies have traditionally leveraged consumer data to improve ad targeting and personalized recommendations, companies are now leveraging existing consumer data to train AI tools, which few existing data protection regulations have sufficient coverage for. These new complexities around data production, refinement, and use elevate a need for comprehensive governance and enforcement measures. Approximately 38 out of 55 African Union Member States have enacted formal data protection regulations. Some of these countries include top economies within the continent, such as Egypt, Nigeria, and South Africa, and emerging players like Benin, Equatorial Guinea, and Zimbabwe. 15 out of 38 data protection laws passed by African countries were enacted in the last five years, and 26 were enacted in the last decade. The first data protection law in Africa was enacted by Cabo Verde in 2001, and data protection laws were recently enacted by Malawi in June 2024 and Ethiopia in July 2024. As of October 2024, Namibia, South Sudan, and The Gambia have drafted data protection laws yet to be enacted. Along with country-specific data governance regulations, regional efforts towards data protection include the African Union Convention on Cyber Security and Personal Data Protection (African Union, 2020), the Economic Community of West African States (ECOWAS) (ECOWAS, 2010), the East African Community (EAC) Legal Framework for Cyberlaws (East African Community, 2008), and the Southern African Development Community (SADC) Model Law on Data Protection (International Telecommunication Union, 2013). At the moment, there have been no regional governance measures proposed or enacted by the Arab Maghreb Union (AMU), the Community of Sahel–Saharan States (CEN–SAD), and the Economic Community of Central African States (ECCAS). While African countries have made significant progress in enacting data protection laws, various factors hinder responsible and sustainable data governance throughout the continent. Additionally, the rising adoption of AI tools introduces new gaps within existing data protection regulations that could be further exploited as AI development increases throughout the continent.

Data Privacy Violations in Africa

Existing data regulatory gaps may also contribute to the growing number of data privacy violations experienced across Africa. In March 2023, the Angolan Agência de Protecção de Dados (APD) issued a fine to Africell, an electronic communications operator, who collected personal consumer data without requesting prior authorization from APD (Agência de Protecção de Dados, 2023). In November 2023, the Telecommunications/ICT Regulatory Authority of Côte d'Ivoire (ARTCI) issued a formal warning to YANGO, a local ridesharing application, for unlawfully recording passenger phone conversations (l'ARTCI, 2023). In July 2023, the South African Information Regulator issued a ZAR 5 million (~USD 273,000) fine against the Department of Justice and Constitutional Development for failure to implement adequate security measures to prevent a ransomware attack in 2021 and noncompliance with required consumer notifications regarding the subsequent data breach (Information Regulator South Africa, 2023). One of the continent's most recent data privacy violations

involves a data breach of Nigeria's National Identity Management Commission of Nigeria (NIMC) system, which has resulted in millions of data points being available for sale on illicit websites for NGN 100 each, which is about USD 6 cents (Paradigm Initiative, 2024). As of October 2024, it is unclear what action the Nigeria Data Protection Commission has taken against the offenders. Kenya Office of the Data Protection Commissioner (ODPC) issued multiple penalties to 4 companies in 2023, totaling over KES 14 million. These fines included noncompliance with a prior enforcement notice on spam calls, harassment from microlending apps, posting minor images, and using customer photos for marketing. ODPC has also made progress in an ongoing investigation regarding violations by Worldcoin, an American cryptocurrency provider that undertook biometric data collection without government notice (Communications Authority Kenya, 2023). While African data protection agencies have increasingly taken actions toward enforcing data protection laws, there is still little understanding of how effective these measures are, given frequent noncompliance with enforcement notices and little information on fine payments by offenders (Lawyers Hub, 2024).

7.1. Operationalizing Data Governance in Africa

In order to ensure that African countries can effectively protect consumers against improper data practices and enforce corrective action against data privacy violations, African governments across every AU Member State must enact comprehensive data regulatory measures. While existing continental-wide efforts, such as the African Union Data Policy Framework, which was published in 2022 to guide AU Member States in designing and reviewing data regulations, and the Malabo Convention on Cyber Security and Personal Data Protection, offer valuable templates for African governments to adopt, these frameworks have unfortunately not seen wide adoption. To help address this lack of adoption and potential challenges from data regulatory gaps, a number of proposals have outlined alternative measures, including regional data governance approaches (Osakwe & Adeniran, 2021; Balogun & Adeniran, 2024), community-centered governance models (Olorunju & Adams, 2024), and data governance reformation (Okolo, 2024). This section introduces the RICE Data Governance Framework to provide a high-level overview of actions African Union Member States can leverage to operationalize data governance effectively.

7.1.1. Reformation, Integration, Cooperation, & Enforcement (RICE) Framework

To begin operationalizing the RICE framework, African governments should pursue regional data governance measures, given the lack of existing coordination with and insufficient protections within existing continental measures such as the Malabo Convention (Yilma, 2022; ALT Advisory, 2022). Efforts to pursue regional data governance would ideally be led by existing RECs such as ECOWAS, EAC, SADC, AMU, CEN-SAD, and ECCAS. Such efforts can then enable the 19 African Union Member States without existing data protections to draft and enact comprehensive data governance measures in a reasonable timeframe. Additionally, enacting regional data governance policies can help address existing capacity constraints for AU Member States unable to individually draft and enact data legislation.

In lieu of functional continental frameworks, countries, regional, and continental bodies should focus on (1) **reforming** existing data regulation and implementing sectoral policy reformation, (2) collaborating with Civil Society Organizations (CSOs) and Academic Research Institutions (ARIs) to improve **integration** of reformed policies, (3) increasing regional and continental **cooperation** in data regulation efforts, and (4) strengthening **enforcement** of reformed data regulation. The RICE Data

Governance Framework recommendations apply at the national, regional, and continental levels, and the core tenets of the framework are defined as follows:

Reformation: To address concerns regarding a lack of comprehensive data governance measures, the AU, RECs, and individual African NGs must reform existing data governance measures and engage in sectoral policy reform. These entities must also establish local expert groups and advisory bodies to enhance policy reform.

The AU, RECs, and NGs should review existing data protection measures, and to meet data governance needs, they should subsequently reform sectoral policies in agriculture, economics, education, healthcare, and other areas.

Integration: To increase awareness and local integration of data protection regulation, RECs and NGs will need to improve outreach to organizations under their jurisdiction. RECs and NGs should also fund outreach and research efforts by CSOs and ARIs to improve public engagement with data protection measures.

ARIs and CSOs should also focus on conducting in-depth research that advances understanding of regional and country-specific needs for data regulation and reduces reliance on standards such as the EU General Data Protection Regulation (GDPR).

Cooperation: To address issues regarding a lack of regional cooperation and inconsistencies in data protection regulation, the AU must lead harmonization efforts across AU Member States. To mitigate issues with prior harmonization efforts (Kenyanito & Chima, 2016), the AU should actively consult RECs and NGs in new harmonization efforts.

The AU should also establish a continental-wide network of National Data Protection Authorities and Offices (NDPAs/NDPOs), as previously recommended in prior work (Data Protection Africa, 2023).

Enforcement: To help address concerns regarding a lack of enforcement of data protection measures, the AU must establish a continental data supervisory body. African governments must also establish and leverage data protection offices to enforce enacted data protection regulations.

The AU should inaugurate a Data Protection Supervisory Authority (DPSA) to increase regional enforcement for data privacy violations and should also help NGs establish NDPAs and NDPOs to mitigate regulatory enforcement gaps.

7.2. CONSIDERATIONS

While this data governance operationalizing framework aims to ease the implementation of comprehensive data regulation within African countries, many considerations exist for the ability of all governments across the continent to leverage this framework. Existing issues with infrastructure, electricity access, education, digital skills literacy, skilled AI talent, climate change, armed conflict, social unrest, national security, and socioeconomic growth may deprioritize and sideline efforts toward

data governance. In light of these existing challenges, however, governments must focus on developing culturally aligned and feasible data governance solutions to ensure that the data rights of African consumers are preserved and that there are adequate outlets for redress of data protection harms.

Regional data governance led by RECs would ideally take precedence over the AU until a formal continental-wide data protection law is passed. However, efforts will be needed to rectify duplicative membership within the RECs and integrate AU Member States without membership in RECs, like the Sahrawi Arab Democratic Republic, which controls the Western Sahara. Prioritizing regional-led data governance before continental reforms are enacted could help address capacity constraints and harmonization issues between AU Member States. Still, there is no guarantee that countries within RECs will reach alignment on data governance measures.

With the growing number of regional and national efforts toward AI regulation throughout the continent, African governments must also understand the fundamental role of data in training ML models, evaluating AI systems, refining predictive models, and improving AI-enabled services (Data Governance Working Group of the Global Partnership on AI, 2020). Given these essential functions, efforts towards enacting effective data governance can also enable more comprehensive AI governance measures. Thus, African governments should consider comprehensive data governance as a viable pathway and complement to AI regulation. To bolster AI-related governance overall, it will also be crucial for African governments to invest in efforts to understand the diverse policy challenges associated with data, including privacy, transparency, labor, interoperability, discrimination, cross-border data flows, and intellectual property.

Conclusion

While the potential of AI is still nascent within Africa, African consumers hold valuable data that is subject to exploitation by both local and international firms alike. Companies are increasingly looking towards African countries to supply them with the necessary data to expand target markets for their AI services. With governments, companies, universities, and other institutions in African countries rapidly adopting AI technologies, there are also concerns that algorithmic harms primarily noted in Western contexts could be exacerbated in ways that disproportionately harm marginalized populations throughout the continent. The limited research examining concrete ethical concerns around data privacy and the lack of extensive efforts toward data protection in Africa is concerning. This work examines data governance measures in Africa, highlighting the regulatory gaps imposed by a lack of comprehensive data governance across Africa that could be further exploited by rising AI adoption. This work presents the RICE Data Governance framework to operationalize comprehensive data governance in African Union Member States to reform and optimize existing data protection measures while bolstering Africa's emerging AI regulatory environment.

References

African Union. (2020). African Union Convention on Cyber Security and Personal Data Protection. https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection

Agência de Protecção de Dados (APD). (2023). APD multa AFRICELL em 150 mil dólares norte americanos por violação da Lei de Protecção de Dados Pessoais (LPDP).

https://www.apd.ao/ao/noticias/apd-multa-africell-em-150-mil-dolares-norte-americanos-por-violacao-da-lei-de-proteccao-de-dados-pessoais-lpdp/

ALT Advisory. (2022). The Malabo Roadmap: Approaches to promote data protection and data governance in Africa. Mozilla. https://dataprotection.africa/wp-content/uploads/malabo_roadmap_Sept_2022.pdf

Balogun, K., & Adeniran, A. (2024). Towards A Sustainable Regional Data Governance Model In Africa. Centre for the Study of African Economies (CSEA).

Communications Authority Kenya. (2023). CA and Data Commissioner Warn Kenyans Over Worldcoin. https://www.ca.go.ke/ca-and-data-commissioner-warn-kenyans-over-worldcoin

Data Governance Working Group of the Global Partnership on AI (GPAI). (2020). The Role of Data in AI. GPAI. https://gpai.ai/projects/data-governance/role-of-data-in-ai.pdf

Data Protection Africa. (2023). Africa: AU's Malabo Convention set to enter force after nine years. ALT Advisory. https://dataprotection.africa/malabo-convention-set-to-enter-force/

East African Community. (2008). Draft EAC Legal Framework for Cyberlaws. http://repository.eac.int/handle/11671/1815

ECOWAS. (2010). Supplementary Act A/SA.1/01/10 on Personal Data Protection Within ECOWAS. https://www.statewatch.org/media/documents/news/2013/mar/ecowas-dp-act.pdf

Information Regulator South Africa. (2023). Media Statement Infringement Notice and R5 Million Administrative Fine Issued to The Department of Justice and Constitutional Development for Contravention of Popia. https://inforegulator.org.za/wp-content/uploads/2020/07/MEDIA-STATEMENT-INFRINGEMENT-NOTICE-ISSUED-TO-THE-DEPARTMENT-OF-JUSTICE-AND-CONSTITUTIONAL.pdf

International Telecommunication Union (ITU). (2013). Data Protection: Southern African Development Community (SADC) Model Law. https://www.itu.int/en/ITU-D/Projects/ITU-EC-ACP/HIPSSA/Documents/FINAL%20DOCUMENTS/FINAL%20DOCS%20ENGLISH/sadc_model_law_data_protection.pdf

Kenyanito, E. P., & Chima, R. J. S. (2016). Room for improvement: Implementing the African Cyber Security and Data Protection Convention in Sub-Saharan Africa. AccessNow.

l'ARTCI. (2023). Communiqué - l'ARTCI. https://www.artci.ci/index.php/33-actualites/informations/629-probables-enregistrements-des-communications-ou-echanges-a-l-interieur-de-vehicules-utilisateurs-de-l-application-denommee-yango-sans-information-prealable-ou-consentement-des-personnes-concernees.html

Lawyers Hub. (2024). Africa Privacy Report 2023/2024. https://www.lawyershub.org/digital

Musanga, M. (2023). Facebook workers in Kenya say Meta hasn't paid them for 6 months amid legal case. openDemocracy. https://www.opendemocracy.net/en/facebook-workers-in-kenya-say-meta-hasnt-paid-them-for-6-months-amid-legal-case/

Okolo, C.T. (2023). AI in the Global South: Opportunities and challenges towards more inclusive governance. The Brookings Institution.

Okolo, C.T. (2024). Reforming data regulation to advance AI governance in Africa. Foresight Africa 2024. The Brookings Institution. https://www.brookings.edu/articles/reforming-data-regulation-to-advance-ai-governance-in-africa/

Olorunju, N., & Adams, R. (2024). African data trusts: new tools towards collective data governance?. Information & Communications Technology Law, 33(1), 85-98.

Osakwe, S., & Adeniran, A. (2021). Strengthening Data Governance in Africa. Centre for the Study of African Economies (CSEA).

Paradigm Initiative. (2024). Major Data Breach: Sensitive Government Data of Nigerian Citizens Available Online for Just 100 Naira. https://paradigmhq.org/major-data-breach-sensitive-government-data-of-nigerian-citizens-available-online-for-just-100-naira/

Yilma, K. (2022). African Union's data policy framework and data protection in Africa. *Journal of Data Protection & Privacy*, 5(3), 209-215.

8. AIED AND STUDENT DATA PRIVACY IN AFRICA: CHALLENGES AND RECOMMENDATIONS FOR LEGISLATORS

ANDREA BAULING

Abstract. Discussions around artificial intelligence in education (AIED) can no longer focus purely on what is technologically possible and pedagogically sound. Advances must be considered within a framework for lawful and responsible learning analytics and data science practices. Lagging efforts to address a widespread lack of AI-specific legislation may be harming millions of students from the majority world. Facilitating the lawful development and implementation of AIED agents that are suited to the needs of African students requires a homegrown approach. Adopting Africa-focussed solutions and legislation could ensure that the great benefit AIED agents may hold for humanity safely includes Africans and others from the global majority.
Keywords. AI in education (AIED), AI regulation, data monetization, data privacy, global majority interests, higher education

INTRODUCTION

The capabilities and social impact of artificial intelligence (AI)[55] agents are expanding at an unprecedented speed. Efforts to regulate AI are lagging, with potentially dire consequences. The challenges faced in the field of AI in education (AIED), which focusses on the use of AI agents to improve educational outcomes and environments, illustrate the need to bolster regulatory efforts. While there are numerous benefits to the use of AIED, many are concerned about the implications for data privacy and data security. Non-existent, ill-suited, and/or unenforced legislation compounds the problem. The inadequacy of the current South African legislative framework demonstrates the potentially corresponding risks threatening many jurisdictions from the majority world. Educators representing the global majority should make their voices heard in spheres where technologies and related policies that affect them are developed. This paper attempts to sketch current and potential future challenges related to data privacy infringements by AIED agents, as well as the legislative steps that could be taken to address these.

8.1. AIED AND THE PROCESSING OF STUDENT DATA

It is necessary to define certain key concepts to facilitate a discussion on the benefits and dangers that AIED may hold for the stakeholders of higher education systems. Educational datamining (EDM) involves the development and application of datamining and machine learning approaches to change raw data collected from education systems and databases into usable information extracted from patterns and connections identified in the data (Maphosa & Maphosa, 2021). The field of learning analytics (LA) concerns developing an understanding of an individual student and their performance in a specific learning environment, often hosted on an online learning management system (LMS), by gathering and analysing personal learning data to ultimately improve learning outcomes and optimise the learning environment (Long & Siemens, 2011; Prinsloo & Slade, 2015). The primary objective of EDM and LA is to support developers, educators, and institutions in their decision-making (Maphosa & Maphosa, 2021).

[55] For the purposes of the paper, AI is understood as defined by Popenici and Kerr (2017). Generative AI falls beyond the ambit of this paper. See Bozkurt et al. (2023) on generative AI and education.

AIHED, a booming subfield of AIED dedicated to higher education, can bolster teaching and learning efficiency. Under the supervision of an educator these systems can facilitate immediate instruction, student supervision, and feedback (Bond et al., 2024; Zawacki-Richter, Marín, Bond & Gouverneur, 2019). Intelligent tutoring systems (ITS) are but one example of an AIED-supported intervention in student learning. ITS can teach content, diagnose strengths and weaknesses in student understanding, curate learning materials, and support peer collaboration (Zawacki-Richter et al., 2019). In some instances, they facilitate a form of computed curriculum that can provide a continuously personalised learning experience in real time, based on a learner's pre-existing knowledge, skills, and rate of progress (Bernhardt, 2023).Clearly, the pedagogical value of such systems is undeniable, but focussing exclusively on their benefits is shortsighted.

8.2. The Interplay Between Personalised Learning and Data Privacy

The complex dichotomy between safeguarding and sharing student, academic, and institutional data, and the development and implementation of AIED agents epitomise "the personalization privacy paradox" (Xu, Luo, Caroll & Rosson, 2011, p.43). AI agents can collect, process, aggregate, and repurpose vast volumes of data housed in institutional silos to generate meaningful insights (Pelletier et al., 2023). But for AI to effectively do so, it must be trained (Bernhardt, 2023). AI agents mainly process personal information in two ways: this data is incorporated in immense datasets employed to train AI machine-learning systems to develop algorithmic models; and once developed, these algorithmic models are applied to other datasets containing personal information to extrapolate predictions about individuals (Bhagattjee, Govuza & Sebanz, 2020). Within an educational context, the individuals in question are students, educators, and administrators.

AIED must be developed by judiciously curating the initial training data used, which is largely based on data generated through EDM and LA activities (Prinsloo & Kaliisa, 2022). Examples of the highly personalised student data processed by LA and EDM systems include learning capabilities and challenges; assessment results and prior academic performance; interaction traces with online content; demographics; funding data; disability status; and health-related indicators (Li, Sun, Schaub & Brooks, 2022; Slade, Prinsloo & Khalil, 2019). From this, AIED agents can deduce students' capabilities, assumed emotional states, mental strategies, and misconceptions (Holmes et al., 2022). Algorithmic models are already capable of diagnosing mental health disorders (Alkahtani, Aldhyani & Alqarni, 2024) and neurodevelopmental disorders such as attention-deficit hyperactivity disorder (Chen et al., 2023). LMSs fully supported by AI are likely to become the new norm (Pelletier et al., 2023). It is not hard to imagine AI-powered diagnostic tools being incorporated into AIED agents and LMSs in the near future, all in the name of pedagogical progress. The implications for data privacy could be astronomical. Many students prefer to keep their highly personal data private and rightfully fear (future) discrimination based thereon, but they mostly have very little (if any) control over what data of theirs is being collected, repurposed, stored, and shared (Li et al., 2022; Zawacki-Richter et al., 2019).

8.2.1. Ownership of and Access to Educational Data

Of great concern is the fact that higher education institutions are (inadvertently) gathering masses of data on their students (Slade et al., 2019). The volume of data collected by LMSs alone is almost

unfathomable.[56] Each student, educator, and administrator's every click is logged and "[t]here are many unanswered questions about who owns this data, who has access to it, [and] how long it will be kept" (Du Boulay, 2023, p.100).[57] At the emergence of LA, most of the data harvested was anonymised, but this is no longer the case (Slade & Prinsloo, 2014). In the pursuit of improved student performance, the prevalence of EDM and LA is increasing, and data is being processed and aggregated in ways that were not initially anticipated or communicated (Willis, Slade & Prinsloo, 2016). Once modern LMSs are implemented "[s]urveillance is insidious and constant" (McGowan et al., 2024, para. 24). The context within which user consent was given, or not,[58] for the collection of (often seemingly harmless) data, becomes further removed from what it may be used for in future, especially as AI algorithmic capabilities progress.[59] The need to protect the personal data of students gathered by higher education institutions, LMSs, and other third-party service providers is evident.

8.3. The South African Legal Position and Potential Regional Interventions

It is essential to consider how we protect the right to data privacy of students and educators from the global majority whose personal data is being collected by LMSs and other for-profit corporations, mainly situated in the developed world. An evaluation of the woefully deficient regulations currently applicable in South Africa provides valuable insights into the legislative challenges faced, which are likely similar to those of various other majority-world jurisdictions.

South African law is enacted, interpreted, and enforced within a constitutionally supreme framework (ss.1(c) & 2 of the Constitution of the Republic of South Africa, 1996). Section 14 of the Constitution protects the right to privacy and the Constitutional Court has confirmed that "the invasion of an individual's privacy infringes the individual's cognate right to dignity" (*AmaBhungane v Minister of Justice* (2021), para.28). To give effect to the right to data privacy, one aspect of the fundamental right to privacy, Parliament enacted the Protection of Personal Information Act (2013) (POPIA). This act currently regulates automated data processing in the jurisdiction, as no other legislation specifically regulating AI has been adopted. As in many other jurisdictions, POPIA is based on its EU counterpart, the General Data Protection Regulation (2016) (GDPR).

Various global data protection laws like the GDPR and POPIA impose data minimisation and purpose limitation principles that restrict what personal data may be collected and how it is processed. These principles are wholly incompatible with the essence of AI-powered data processing and the training of models capable of such activities (Bhagattjee et al., 2020). Unfortunately, POPIA does not prescribe data protection impact assessments or any other accountability requirements as the GDRP does, which diminishes the potency of the Act's regulatory capabilities (Bronstein, 2022). A further point of concern is that POPIA has, to date, not been enforced in earnest (Musoni & Mtuze, 2023). These

---

56 In October 2024, the world's largest LMS, Moodle, boasted hosting more than 2,4 billion enrolments from 239 countries, 427 million active users, and 801 million discussion forum posts (Moodle, 2024). Moodle provides an invaluable, openly available platform that learning institutions can use freely and modify to suit their needs. While this approach is laudable, the effect is that the organisation has access to quadrillions of datapoints on users from across the globe. Worryingly, biometric data is collected for features such as facial recognition, used to proctor online assessments.

57 See McGowan, Paris & Reynolds (2024) on the dangers inherent in procuring AIED systems under "software-as-a-service" (SAAS) agreements.

58 Higher education institutions often grant consent to vendors or external service provider on users' behalf, and without their knowledge (McGowan et al., 2024).

59 See Slade & Prinsloo (2014) on this "context collapse".

challenges have serious implications for achieving the goal with which this law was enacted. Based on these and other concerns, South African legal scholars support the promulgation of AI-specific legislation and argue that this should encapsulate definitive prescripts on the degree of repurposing of personal information by AI agents that would be considered lawful (Bhagattjee et al., 2020; Mahomed, 2018; Musoni & Mtuze, 2023).

Some jurisdictions have moved beyond merely relying on data privacy legislation. In 2023 the European Parliament passed the EU Artificial Intelligence Act (2021). Crucially, the Act classifies the education sector as a high-risk field in which to apply AI systems (arts.6(2), 8 & 9). Because of the high potential for harm to individuals, the Act requires continuous risk assessment and management of AI agents developed for and implemented in educational settings (arts.8 & 9). This special focus on potential risk is sensible.

As in the EU, member states of the Southern African Development Community (SADC) are cooperating in various regional initiatives to coordinate data protection practices (Thaldar & Malekela, 2024). Since a collaborative regional approach to AI regulation would serve SADC citizens and activities (Gwagwa, Kraemer-Mbula, Rizk, Rutenberg & De Beer 2020), engaging these existing working groups could add significant value to discussions on data sharing and data protection, as relevant to AI development projects. While it is paramount that African solutions are adopted to solve African problems, it may be prudent to use the EU AI Act as springboard for a project of this nature (Gwagwa et al., 2020). African AI legislation will need to embrace the inherent dichotomy at play in regulating AI development: promoting technological progress and access to the immense promise of AI, and protecting the interests of the persons these AI agents aim to serve.

8.4. The Necessity of an African Approach to AI and AIED Regulation

Regulators the world over are attempting to circumvent the potential social harm that AI agents may cause by developing global standards for AI (Karanicolas, 2023; 2024). Karanicolas (2023, pp.266-267) argues that "the world would be better served if the standard-setting processes represented … perspectives from the people of the Majority World". One aspect of the potential social harm in question stems from the bias and (often race-based) discrimination inherent in many algorithms and AI models originating in the developed world (Jiao, Afroogh, Xu & Phillips, 2024). The race to prevent or rectify harm of one form by diversifying datasets, may inadvertently cause another: the infringement of the right to data privacy of millions. Campbell-Stephens (2021, p.6) explains that "[t]he term 'global majority' invites social cooperation across groups, existentially to address the mutual interests of the majority on planet earth through collective mobilisation." It is crucial that such collective efforts extend to the sphere of AI and ultimately AIED development and regulation. The case of the open university illustrates the necessity in this regard.

Open distance universities can provide higher education at scale and therefore serve the needs of the majority world well. At a conservative estimate, the 10 largest public open universities in the world[60] service almost 20 million students, almost all from the majority world (Bozkurt, 2019; De Vries, 2019;

[60] Indira Gandhi National Open University (India), Open University of China, Anadolu University (Türkiye), Allama Iqbal Open University (Pakistan), Bangladesh Open University, National Open University of Nigeria, Dr. B.R. Ambedkar Open University (India), Payame Noor University (Iran), and University of South Africa (see Jones, 2018; Quayyum & Zawacki-Richter, 2019). The author contends that this list may be incorrect. Reliable, aggregated, and up to date sources are not readily available.

Quayyum & Zawacki-Richter, 2019; Zhang & Li, 2019). The most cost-effective way for open universities to provide higher education to hundreds of thousands, or millions, is to do so online by means of an LMS. Initial agreements with LMS service providers often entail seemingly innocuous terms and conditions, which upon closer inspections could have significant implications for data privacy through the assetisation of higher education and the commodification of student data (Prinsloo & Kaliisa, 2020).

The higher education sector has come to be regarded as "a site of value and ongoing wealth extraction" (Scott & Gray, 2023, p.606). Higher education institutions have both a fiducial and moral duty to consider the paramountcy of safeguarding the data privacy of their students and staff when contracting with external education platforms (Prinsloo & Kaliisa, 2020). This may be especially true for open universities and African higher education institutions, as they are most vulnerable to "datafication" and exploitation by international corporations (Bozkurt, 2019; Prinsloo & Kaliisa, 2020). This raises legitimate "concerns about Africa being re-colonised and its data exported and capitalised" (Prinsloo & Kaliisa, 2020, p.896). It is therefore crucial that the global majority initiate collaborative efforts to protect their own data privacy interests. Africa should regulate how and when African data may be shared to safely support the interests of her people in AI-related matters.

8.5. Conclusions and Recommendations

In the absence of AI-specific legislation, privacy laws are the only legal safeguards that apply to the development and implementation of AI. Rigid common-law prescripts on privacy and legislation specifically relevant to the right to data privacy stifle innovation in AI, resulting in an untenable and impractical situation. The overarching philosophy that "[p]rivacy promotes safe learning" (Anwar, 2021, p.772) should guide attempts to balance the ostensibly opposing interests inherent in the threats related to the processing of student data by algorithms and AIED agents and facilitating equitable learning experiences as a result thereof. While promulgating comprehensive jurisdiction-specific AI legislation is both critical and urgent, this approach is most likely not a sufficiently judicious regulatory approach to address the complexities of the use of AI data-processing agents at work within higher education systems. AIED-specific legislation and a domestic approach to AI development and regulation are thus crucial, as is set out below.

8.5.1. AIED-Specific Legislation as Ancillary Regulation

International calls for AIED-specific legislation and policies are mounting (Bond et al., 2024). This unique subfield of AI would be best served by a more nuanced approach to regulation. The urgency of this is illustrated by contrasting how we think about consumer and student surveillance. We acknowledge the potential harm that stems from commercial surveillance practices such as the scraping of publicly available information from the internet (Solove & Hartzog, 2024). These practices seem inherently dangerous, as they hold little or no benefit for data subjects. Yet the data collected from users of LMSs through inherent and insidious surveillance practices is of an even more personal, and thus potentially harmful, nature because it involves *inter alia* records of mental and cognitive (dis)abilities, and potentially health information. Global societies mostly regard education as a key endeavour that advances humanity,[61] and rightly so. Sadly, universities' lax procurement practices (Scott & Gray, 2023) and legislatures' failure to act has shown that we are more likely to regard

[61] One example of this perspective is encapsulated in the vision of the University of South Africa: "towards *the* African university shaping futures in the service of humanity" (Unisa, 2024).

infringements on data privacy rights in the name of improved educational outcomes as being for 'the greater good'[62]. Specifically legislating AIED is essential, most importantly because doing so will expose the inherent dangers thereof to all potentially affected persons and institutions, as well as the public. Legislation will convert the moral and ethical obligations to protect users of AIED agents to a legal obligation enforceable by sanctions.

Enacting AIED-specific legislation within a given jurisdiction may take time and such a project could be undertaken as a subsequent, more nuanced phase of AI regulation. Enacting overarching AI-specific legislation at national or federal level is an essential interim measure. Here the EU's approach may provide inspiration, as it highlights the domain of education as one of several in which the implementation of AI agents could potentially engender great harm to individuals.

8.5.2. A Regional Approach to AI(ED) Development and Regulation

There is a growing call for African collaboration in both the development of AI agents and AI-related policies and regulations (AU Specialised Technical Committee, 2019; Musoni & Mtuze, 2023). Modifying thinking around AI to suit local contexts requires local sensitivities: "building trust means taking your people on the journey, so that they can internalise what these ideas mean, bring abstract principles to life in their own language and metaphors, and tell user stories they can inhabit" (Buckingham Shum, 2024). However, a context-specific, multi-disciplinary approach to developing AI(ED) laws for the African region could use the EU AI Act as a point of departure, but not a blueprint. Identifying and then adapting relevant aspects of this act into stipulations that are pertinent to and practically enforceable in Africa could serve as a useful first step. To effectively develop homegrown algorithms and AIED fit for African students and educators, African EDM researchers need access to openly available African data sets (Maphosa & Maphosa, 2020). Collaborative African AI regulation should aim to strike a balance that allows such access, while protecting the interests of data subjects. Such a regulatory project could potentially inspire and influence similar majority-world initiatives.

Maphosa and Maphosa (2020) argue that it is not yet clear how higher education institutions must respond to the legal complexities related to the collection and use of student data. It is, however, clear that these institutions have a fiduciary and moral duty to defend student data privacy (Prinsloo & Slade, 2017).

8.5.3. Final Remarks

Educators, researchers, and actors in the sphere of LA, EDM, and AIED who represent the global majority have a duty to help safeguard students from harm stemming from the infringement of their right to data privacy, the future implications of which cannot be fully known. This is essential, as Africa and other majority-world regions are specifically at risk of gross mass data privacy infringements by profiteers. While pedagogical and technological progress is most certainly desired, the achievement thereof should not steamroller fundamental rights. The weighing up of rights, duties, benefits, and risks should be done with diligent consideration to protect our students from potential harm as best we can. While legislation is never perfect, and the enforcement thereof often complex, legal prescripts are more concrete than ethical or moral guidelines. It is essential to pass general AI-specific, and later AIED-focussed, legislation at national or federal level. Regional cooperation throughout the majority

[62] See Czerniewicz and Cronin (2023) on the role of education "for good" in society.

world could significantly bolster these endeavours, ensuring that education continues to safely support individual and global development.

References

African Union Specialised Technical Committee on Communication and Information Technologies. (2019). *2019 Sharm El Sheikh Declaration*. Retrieved from https://au.int/en/decisions/2019-sharm-el-sheikh-declaration-stc-cict-3

Alkahtani, H., Aldhyani, T. H. H., & Alqarni, A. A. (2024). Artificial intelligence models to predict disability for mental health disorders. *Journal of Disability Research*, *3*(3), 1-12. doi:10.57197/JDR-2024-0022

AmaBhungane Centre for Investigative Journalism NPC v Minister of Justice and Correctional Services [2021] ZACC 3. Retrieved from https://www.saflii.org/cgi-bin/disp.pl?file=za/cases/ZACC/2021/3.html&query=amabhungane%20near%202021%20near%20zacc%20near%203

Anwar, M. (2021). Supporting privacy, trust, and personalization in online learning. *International Journal of Artificial Intelligence in Education*, *31*, 769-783. doi:10.1007/s40593-020-00216-0

Bernhardt, M. (2023). *AI's increasingly important role in L&D*. The Learning Guild, New York, N.Y. Retrieved from https://www.learningguild.com/publications/183/ais-increasingly-important-role-in-ld/

Bhagattjee, P., Govuza, A., & Sebanz, L. (2020). Regulating artificial intelligence from a data protection perspective – lessons from the EU. *Without Prejudice*, December 2020, 9-10.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., … Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*(4), 1-41. doi:10.1186/s41239-023-00436-z

Bozkurt, A. (2019). The historical development and adaptation of open universities in Turkish context: Case of Anadolu University as a giga university. *International Review of Research in Open Distance Learning*, *20*(4), 36-59.

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., … Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, *18*(1), 1-78. Retrieved from http://www.asianjde.com/ojs/index.php/AsianJDE/article/view/709/394

Bronstein, V. (2022). Prioritising command-and-control over collaborative governance: The role of the Information Regulator under the Protection of Personal Information Act. *Potchefstroom Electronic Law Journal*, *25*, 1-41. doi:10.17159/1727-3781/2022/v25i0a11661

Buckingham Shum, S. (2024, January 15). Co-designing AI ethics in education [Blog post]. Retrieved from https://simon.buckinghamshum.net/2024/01/codesigning-ai-ethics-edu/

Campbell-Stephens, R. M. (2021). *Educational leadership and the global majority*. doi:10.1007/978-3-030-88282-2

Chen, T., Tachmazidis, I., Batsakis, S., Adamou, M., Papadakis, E., & Antoniou, G. (2023). Diagnosing attention-deficit hyperactivity disorder (ADHD) using artificial intelligence: A clinical study in the UK. *Front Psychiatry*, *14*:1164433, 1-13. doi:10.3389/fpsyt.2023.1164433

Constitution of the Republic of South Africa, 1996. Retrieved from https://www.saflii.org/content/Constitution-of-the-Republic-of-South-Africa_1996.html

Czerniewicz, L., & Cronin, C. (Eds.). (2023). *Higher education for good: Teaching and learning futures*. doi:10.11647/obp.0363.27

De Vries, I. (2019). Open universities and open educational practices: A content analysis of open university websites. *International Review of Research in Open Distance Learning*, *20*(4), 167-178.

Du Boulay, B. (2023). Artificial intelligence in education and ethics. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of Open Distance and Digital Education* (pp. 93-108). doi:10.1007/978-981-19-2080-6 pp.93-108

Gwagwa, A., Kraemer-Mbula, E., Rizk, N., Rutenberg, I., & De Beer, J. (2020). Artificial intelligence (AI) deployments in Africa: Benefits, challenges and policy dimensions. *African Journal of Information and Communication*, *26*, 1-28.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., … Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *32*, 504-526. doi:10.1007/s40593-021-00239-1

Jiao, J., Afroogh, S., Xu, Y., & Phillips, C. (2024). *Navigating LLM ethics: Advancements, challenges, and future directions*. Retrieved from https://arxiv.org/pdf/2406.18841

Jones, J. (2018). *7 largest universities in the world*. Retrieved from https://largest.org/misc/universities/

Karanicolas, M. (2023). Developing AI standards that serve the majority world. In L. Belli & W. B. Gaspar (Eds.), *The quest for ai sovereignty, transparency and accountability: Official outcome of the UN IGF Data and Artificial Intelligence Governance Coalition* (pp. 265-282). Retrieved from https://direitorio.fgv.br/publicacao/quest-ai-sovereignty-transparency-and-accountability

Karanicolas, M. (2024). Challenging minority rule: Developing AI standards that serve the majority world. *UCLA Law Review DisC*, *71*, 196-213.

Li, W., Sun, K., Schaub, F., & Brooks, C. (2022). Disparities in students' propensity to consent to learning analytics. *International Journal of Artificial Intelligence in Education*, *32*, 564-608. doi:10.1007/s40593-021-00254-2

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, *46*(5), 31-40.

Mahomed, S. (2018). Healthcare, artificial intelligence and the Fourth Industrial Revolution: Ethical, social and legal considerations. *South African Journal of Bioethics and Law*, *11*(2), 93-95. doi:10.7196/SAJBL.2018.v11i2.664

Maphosa, V., & Maphosa, M. (2021). The trajectory of artificial intelligence research in higher education: A bibliometric analysis and visualisation. *2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, *IEEE*, 1-7. doi:10.1109/icabcd51485.2021.9519368

McGowan, C., Paris, B., & Reynolds, R. (2024). Educational technology and the entrenchment of "Business as usual". *Academe*, *110*(1). Retrieved from https://www.aaup.org/article/educational-technology-and-entrenchment-%E2%80%9Cbusiness-usual%E2%80%9D

Moodle. (2024). *Statistics*. Retrieved from https://stats.moodle.org/

Musoni, M. & Mtuze, S. (2023). An Assessment of the Key AI Sovereignty Enablers within the South African Context. In L. Belli & W. B. Gaspar (Eds.), *The Quest for AI Sovereignty, Transparency and Accountability: Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition* (pp. 45-58). Retrieved from https://direitorio.fgv.br/publicacao/quest-ai-sovereignty-transparency-and-accountability

Pelletier, K., Robert, J., Muscanell, N., McCormack, M., Reeves, J., Arbino, N., ... Zimmern, J. (2023). *EDUCAUSE Horizon Report, Teaching and Learning Edition*. Retrieved from https://library.educause.edu/resources/2023/5/2023-educause-horizon-report-teaching-and-learning-edition

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(22), 1-13. doi:10.1186/ s41039- 017- 0062-8

Prinsloo, P., & Kaliisa, R. (2022). Data privacy on the African continent: Opportunities, challenges and implications for learning analytics. *British Journal of Education Technology*, *53*, 894-913. doi:10.1111/bjet.13226

Prinsloo, P., & Slade, S. (2015). Student privacy self-management: implications for learning analytics. *Proceedings of the LAK '15 Fifth International Conference on Learning Analytics and Knowledge, ACM*, 83–92. doi:10.1145/2723576

Prinsloo, P., & Slade, S. (2017). Big Data, Higher Education and Learning Analytics: Beyond Justice, Towards an Ethics of Care. In B.K. Daniels (Ed.), *Big Data and Learning Analytics in Higher Education* (pp. 109-124). Cham, Switzerland: Springer.

Protection of Personal Information Act 4 of 2013. Retrieved from https://www.saflii.org/cgi-bin/disp.pl?file=za/legis/num_act/popia2013380/popia2013380.html&query=protection%20near%20of%20near%20personal%20near%20information%20near%20act

Quayyum, A., & Zawacki-Richter, O. (2019). The state of open and distance education. In O. Zawacki-Richter & A. Qayyum. (Eds.), *Open and distance education in Asia, Africa and the Middle East* (pp. 125-140). doi: 10.1007/978-981-13-5787-9_14.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Pp. 1-88).

Regulation (EU) 2021/0106 of the European Parliament and of the Council 21 April 2021 on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (AI Act) (Pp. 1-108).

Scott, M., & Gray, B. C. (2023). Who cares about procurement? In L. Czerniewicz & C. Cronin (Eds.). *Higher education for good: Teaching and learning futures*. pp. 603-621. doi:10.11647/obp.0363.27

Slade, S., & Prinsloo, P. (2014). Student perspectives on the use of their data: Between intrusion, surveillance and care. In *Challenges for research into open & distance learning: Doing things better – Doing better things* (pp. 291–300). Retrieved from https://oro.open.ac.uk/41229/1/BRPA_Slade_Prinsloo.pdf

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the LAK '19 Ninth International Conference on Learning Analytics and Knowledge*, ACM, 1-10. doi:10.1145/3303772.3303796

Solove, D. J., & Hartzog, W. (2024). The Great Scrape: The clash between scraping and privacy. *California Law Review*, *113* (forthcoming 2025). Advance online publication. doi:10.2139/ssrn.4884485

Thaldar, D., & Malekela, M. (2024). Data protection law: Lessons from Tanzania for South Africa? *South African Journal of Bioethics and Law*, *17*(2), 57-58. Retrieved from https://samajournals.co.za/index.php/sajbl/article/view/2301/1072

Unisa. (2024). *Who we are*. Retrieved from https://www.unisa.ac.za/sites/corporate/default/About/Who-we-are?lang=01

Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: A typology derived from a cross-continental, cross-institutional perspective. *Educational Technology Research and Development*, *64*, 881-901. doi:10.1007/s11423-016-9463-4

Xu, H., Luo, X., Caroll, J. M., & Rosson, M. B. (2011). The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decision Support Systems*, *51*(1), 42-52. doi:10.1016/j.dss.2010.11.017

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(39), 1-28. doi:10.1186/s41239-019-0171-0

Zhang, W., & Li, W. (2019). Transformation from RTVUs to open universities in China: Current state and challenges. *International Review of Research in Open Distance Learning*, *20*(4), 1-20.

9. Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law: a Commentary

Ekaterina Martynova

Abstract. This paper provides a brief commentary on the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, and considers its possible impact on the AI regulation in the third states. It analyses the general characteristics of the Convention: its legal nature, object and purpose, along with specific issues relating to the procedure for the use of remedies, the process of accession, and the mechanisms of implementation of the Convention at the national level. The commentary concludes by highlighting the provisions and approaches of the Convention that could be useful in shaping national AI regulation, as well as common regulatory framework on the BRICS platform. Such provisions include, inter alia, standards of transparency, reliability, risk assessment, accountability and responsibility for negative consequences, as well as remedies for the individuals whose rights have been violated by the use of AI systems.
Keywords. artificial intelligence, human rights, Council of Europe, international treaty

Introduction

The emerging field of international legal regulation of artificial intelligence (hereinafter — the AI) results in the interaction of various sources, such as private agreements (often, market-driven[63]) made by corporations, regulations set by individual states, the body of international law itself, recommendations of international organizations and civil society, which predefines formulation of the normative framework through a range of approaches, from voluntary agreements to formal regulations. Among them, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (hereinafter — the Convention), adopted on 17 May 2024 by the Committee of Ministers of the Council of Europe, is the first international legally binding treaty. Its purpose is to ensure respect for human rights, the rule of law and legal standards of democracy at all stages of the design, development and application of the AI systems.[64] There are three main objectives that the Convention aims to achieve: firstly, to address the problems in interpreting human rights in the context of AI application; secondly, to embed fundamental human rights principles in relation to AI; and thirdly, to establish international human rights norms on the application of AI to promote international trade.[65] This paper provides a brief commentary on the Convention in the light of the stated objectives of its adoption. It first looks at the general characteristics of the Convention and the content of the obligations of signatory states. It then considers the remedies available to individuals whose human rights are allegedly violated in the context of the use of AI systems. The discussion concludes with a consideration of the possible implications of the Convention for third states and the prospects for the BRICS countries to learn from the experience of developing the Convention.

9.1. Discussion

[63] Chinen, M. (2023). *The international governance of artificial intelligence.* Edward Elgar Publishing. P. 72-106.
[64] The Convention, Article 1(1).
[65] Van Kolfschooten, H., & Shachar, C. (2023). The Council of Europe's AI Convention (2023–2024): Promises and pitfalls for health protection. *Health Policy*, *138*, 104935. https://doi.org/10.1016/j.healthpol.2023.104935.

9.1.1. LEGAL NATURE OF THE CONVENTION, ITS OBJECT AND PURPOSE

The development of AI systems is welcomed and encouraged by states because of the vast opportunities that AI offers to improve other technologies, industrial growth and the intensification of trade. However, the use of AI has political, social and economic implications for various social relations, both nationally and internationally, that go beyond the legal regime regulating AI as a technology.[66] The adoption of the Convention as an international treaty is intended to establish a legal framework that will respond to the new challenges that the international community and individual states face in connection with the development of AI, in particular with regard to the functioning of democratic institutions,[67] the protection of rights and freedoms,[68] and the overcoming of social inequalities and discrimination arising from the use of such computer programs.[69]

The term 'AI system' is defined in the Convention as "a machine-based system that for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments" with indication that "[d]ifferent artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment".[70] Thus, the Convention has adopted a so-called 'broad' approach to defining an AI system based on its self-learning and generative abilities, as opposed to a 'narrow' approach to defining AI based on the ability of a system to solve a specific applied task such as translation services or chatbots.[71]

The definition provided in the Convention does not contain an indication of the possible dual-use (military and civilian) nature of AI systems. Herewith, it is important to note that activities related to national defence are excluded from the scope of the Convention.[72] Thus, the obligations of Parties to the Convention to ensure transparency, accountability and responsibility for possible adverse effects do not apply to the activities related to the development or application of AI systems for military purposes. At the same time, the use of AI in defence as an autonomous lethal weapon system has the potential to seriously affect the geopolitical balance between states, creating new international asymmetries.[73]

---

[66] Crawford, K. (2022). Atlas of AI: power, politics, and the planetary costs of artificial intelligence. New Heaven: Yale University Press, 2021. P. 185-186.
[67] Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society a Mathematical Physical and Engineering Sciences*, *376*(2133), 20180089. https://doi.org/10.1098/rsta.2018.0089.
[68] Donahoe, E., & Metzger, M. M. (2019). Artificial Intelligence and Human Rights. *Journal of Democracy*, *30*(2), 115–126. https://doi.org/10.1353/jod.2019.0029.
[69] Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Retrieved from https://openlibrary.org/books/OL26681102M/Automating_Inequality.
[70] The Convention, Article 2.
[71] Despite the ubiquitous nature of AI discussions lately, there is no consistent 'official' definition of AI. In some cases, the technical descriptions offered by computer scientists are not suitable for legal analysis, for example when AI is defined in terms of an 'algorithm', which in turn requires a separate definition and understanding of the social meaning and legal content. For the review of different approaches to define AI for the purposes of legal studies, refer, e.g. to Lee, J. (2022). *Artificial intelligence and international law*. Singapore: Springer. P. 6-8. On the 'broad' and 'narrow' approach to defining AI see, e.g. Meltzer, J. P. (2018, December 13). The impact of artificial intelligence on international trade. *Brookings*. Retrieved from https://www.brookings.edu/research/the-impact-of-artificial-intelligence-on-international-trade/#footnote-1.
[72] The Convention, Article 3(4).
[73] Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. *Defense and Security Analysis*, *35*(2), 147–169. https://doi.org/10.1080/14751798.2019.1600800.

With regard to the scope and application of the Convention, states Parties to the Convention are expected to take the legislative, administrative or other measures necessary to ensure compliance with the provisions of the Convention by both public authorities and private actors acting on their behalf.[74] The Convention provides an alternative means of regulating private actors not acting on behalf of a state: Parties may extend the principles and obligations set out in the Convention to the private sector (thereby putting it on an equal regulatory footing with the public sector) or take other appropriate measures to manage the risks of the use of AI systems by private actors in a manner consistent with the object and purpose of the Convention.[75] The chosen method of fulfilling the obligation to regulate the private sector shall be communicated at the time of signing or depositing the instrument of ratification, acceptance, approval or accession to the Convention (the chosen method can be subsequently changed). There can be no derogation from or limitation on the application by a Party of its international obligations relating to human rights, democracy and the rule of law.[76] This 'fork in the road' in the methods of regulating the private sector does not seem to be entirely appropriate, as it creates an imbalance in the scope of the obligations of the Parties to the Convention, depending on the option chosen.

9.1.2. The main obligations of the Parties to the Convention

As a general comment, it should be noted that, although the Convention enumerates the obligations of states Parties, certain 'saving clauses' anticipate its framework nature. Thus, as a general principle of regulation, it is stipulated that each Party shall fulfil its obligations under the Convention "in a manner appropriate to its domestic legal system".[77] In the text of the Convention, states' obligations are formulated as 'soft' goals and obligations of conduct rather than obligations of result. On the one hand, this approach ensures flexibility in the application of the Convention and is likely to increase the number of the Parties, but on the other hand, it 'blurs' the content of the states' obligations and leaves significant room for interpretation. At the same time, according to Martti Koskenniemi, a possible shift in the balance between normativity and certainty towards normativity will inevitably lead to inconsistency in practice and thus to the politicisation of the relevant regulation.[78] In particular, the obligation of the Parties to take measures aimed at protecting democratic processes in the lifecycle of AI systems, including ensuring the "ability to freely form opinions", as set out in the Convention,[79] may be implemented in significantly different ways by states, depending on the chosen approach to regulating social networks.

The Convention does not establish strict requirements for the adoption of specific measures, hence the provisions of this international treaty are largely non-self-executing. This distinguishes the Convention from the European Union AI Act[80] which applies directly on the territory of all EU Member States and creates very specific positive obligations of the Member States with certain deadlines —

[74] The Convention, Articles 1(2) and 3(1)(a).
[75] The Convention, Article 3(1)(b).
[76] Ibid.
[77] The Convention, Article 6.
[78] Koskenniemi, M. (2006). *From Apology to Utopia: The Structure of International Legal Argument*. Cambridge University Press.
[79] The Convention, Article 5(2).
[80] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828. PE/24/2024/REV/1 // *OJ L, 2024/1689, 12.7.2024.*

e.g., to establish rules for penalties and enforcement measures, including warnings and non-monetary actions, that can be applied to operators who violate the Act's regulations;[81] to introduce laws, regulations or administrative provisions, more favourable to workers in terms of protecting their rights in respect of the use of AI systems by employers;[82] and to introduce, in accordance with EU law, restrictive laws on the use of postremote biometric identification systems.[83]

The Convention establishes an obligation for the Parties to implement "adequate transparency and oversight requirements" for the lifecycle activities of AI systems, including the identification of content generated by such systems, taking into account specific contexts and risks.[84] This commitment is linked with the requirement to ensure accountability and responsibility for possible adverse impacts of AI systems on human rights, democracy and the rule of law.[85] Precise scope of relevant standards is defined by the state Parties themselves.

Parties to the Convention shall also take measures to ensure that AI systems respect equality, including gender equality, as well as the prohibition of discrimination and do not violate privacy rights of individuals.[86] At the same time, the relevant articles of the Convention include reservations that such obligations shall be implemented by states taking into account international and national law. It appears that the actual content of these obligations in states of different legal systems may vary significantly, in particular with regard to the gender equality and approaches to the grounds for permissible restrictions on the right to privacy. In general, it can be assumed that the choice and 'calibration' of the instruments laid down in the Convention and their implementation in the national regulation of AI will be determined by the specificities of the political regime of the state Party to the Convention: in particular, the degree of involvement of stakeholders in the process of normative regulation, the effectiveness of institutions that determine the rules of behaviour of participants in the life cycle of AI systems, as well as the role of civil society and organisations for the protection of human rights and freedoms.[87]

9.1.3. REMEDIES

Parties to the Convention are obliged to ensure that remedies are available to persons whose human rights have been violated in connection with the use of AI systems (again, with the proviso that such measures are taken to the extent that they comply with the requirements of the domestic legal system).[88] As basic procedural safeguards for the protection of human rights when interacting with AI systems, the Convention provides for notification to any persons of their interaction with AI systems, documentation of information about the use of AI systems that potentially violates human rights, and the possibility for interested persons to access such information and lodge a complaint with the competent public authority.[89] However, the precise approach to be taken to answer the question of whether the rights of applicants have been violated is left outside the scope of this international treaty,

[81] Ibid, Recital 168 / 179 and Article 99, 113.
[82] Ibid, Recital 23 and Article 2(11).
[83] Ibid, Recital 96 and Article 27(10).
[84] The Convention, Article 8.
[85] The Convention, Article 9.
[86] The Convention, Articles 10, 11.
[87] For an overview of national AI policy regimes and their typology by political regime, see: Filgueiras, F. (2022). Artificial Intelligence Policy Regimes: Comparing Politics and Policy to National Strategies for Artificial Intelligence. *Global Perspectives*, *3*(1). https://doi.org/10.1525/gp.2022.32362.
[88] The Convention, Article 14(1).
[89] The Convention, Article 14(2).

leaving room for a variety of models. Thus, the Convention merely establishes a general procedural vector by guaranteeing the availability of a remedy.

9.1.4. EXEMPTIONS FOR NATIONAL SECURITY AND SCIENTIFIC RESEARCH PURPOSES

Parties to the Convention are exempt from compliance with their obligations when carrying out activities related to the defence of national security interests, provided that such activities are carried out without violating international law, including international human rights obligations, and with respect for democratic institutions and processes.[90] The motivation for this exception is obvious, but its danger is that states may apply it broadly, without providing an explanation of the reasons for applying the exception, on the grounds that the mere explanation of the reasons poses a threat to national security. The Convention also does not restrict the Parties from conducting research and development activities regarding AI systems, provided that such activities do not involve risks to human rights, democracy or the rule of law.[91]

9.1.5. OVERSIGHT MECHANISMS

In order to ensure the effective implementation of the provisions of the Convention, a monitoring mechanism is established in the form of a Conference of the Parties with advisory powers.[92] In addition, Parties to the Convention undertake to establish their own independent mechanism to oversee compliance with the Convention and assess risks of human rights violations, to take measures to raise public awareness, encourage informed public debate and consultation with all stakeholders on the use of AI systems,[93] as well as to send periodic reports on progress in the implementation of the Convention for consideration by the Conference of the Parties.

9.1.6. SIGNATURE PROCEDURE, POSSIBILITY OF RESERVATIONS

The Convention is open for signature by the Member States of the Council of Europe, the European Union and the states that participated in its elaboration (the states whose representatives participated in the work of the Committee on Artificial Intelligence: Argentina, Australia, Canada, Costa Rica, the Holy See, Israel, Japan, Mexico, Peru, the Unites States and Uruguay). The signing took place in Vilnius, Lithuania, on 5 September 2024 during the Conference of Ministers of Justice. Andorra, Georgia, Iceland, Norway, the Republic of Moldova, San Marino, the United Kingdom, Israel, the United States of America as well as the European Union have signed the Convention.[94] Once the Convention enters into force (on the first day of the month following the expiry of a period of three months after the date of ratification of the Convention by five signatories, including at least three member states of the Council of Europe), states that did not participate in its drafting will be able to accede to it, provided that such accession is approved by a decision adopted by a majority in accordance with Article 20.d of the Statute of the Council of Europe and by a unanimous vote of the representatives of the parties to the Convention entitled to sit on the Committee of Ministers.[95] This strict procedure for accession to the Convention is not unique to Council of Europe treaties: most of them, including the so-called 'open'

[90] The Convention, Article 3(2).
[91] The Convention, Article 3(3).
[92] The Convention, Article 23.
[93] The Convention, Articles 16, 19, 20.
[94] Council of Europe. (2024, September 13). Council of Europe opens first ever global treaty on AI for signature. *Portal*. Retrieved from https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature.
[95] The Convention, Article 31(1).

treaties, i.e. allowing accession by non-Council of Europe member states, require the unanimous consent of the parties.[96]

9.1.7. Possible consequences of the adoption of the Convention for the third parties

The Convention does not impose any obligations on states not parties to it. Even though the Convention contains a declaratory provision on the Parties' endeavour to encourage non-parties to act consistently with its principles,[97] this provision does not create any normative obligations.

To date, none of the BRICS countries have signed the Convention. The accession of the Russian Federation, which ceased to be a member of the Council of Europe in September 2022, to the Convention is unlikely to be an issue in the near future, including due to the unanimous vote required of the representatives of the parties to the Convention entitled to sit on the Committee of Ministers for accession by a non-member state of the Council of Europe. At the same time, the two-year experience of the Committee on Artificial Intelligence in drafting the text of the Convention and some of its provisions may be useful in the formation of national and international normative regulation of activities using AI systems, particularly at the BRICS level. Specifically, the principles of transparency, reliability, risk assessment, accountability and responsibility for negative consequences seem to be the most important foundations for the activities of public authorities and private actors within the lifecycle of AI systems. Guarantees of information to citizens, as provided for in the Convention, may also be perceived by other legal systems — for example, in the form of labelling of the content generated by an AI system and information that an interaction with an AI system is taking place (for example, when a consumer receives services via a telephone call). In addition, an analysis of the practice of states Parties to the Convention in providing remedies to citizens whose rights are violated by the use of AI systems may be useful for improving other states' national legislation, especially in areas that are sensitive from the perspective of protecting citizens' rights, such as the use of facial recognition systems.

9.2. Conclusion

As noted above, the framework nature of the Convention has influenced the formulation of the obligations assumed by states. The Convention does not lay down strict requirements for the adoption of specific measures. As a result, the provisions of this international treaty are largely non-self-executing, and the nature of the obligations set forth in the Convention gives states wide discretion in their implementation. Moreover, the broad discretion of states in implementing this treaty is reflected in the right of states Parties to determine the extent to which the Convention applies to the development and use of AI systems in the private sector. The Convention thus embodies a 'soft' model of international legal regulation in the field of AI. At the same time, the adoption of the rules enshrined in the Convention will certainly be a positive incentive for the development of domestic legislation regulating the development and use of AI systems. Moreover, this soft regulatory approach seems to be a possible first step towards developing a common regulatory framework at the international level among states with less integrated legal systems compared to, for example, the European Union. In this

[96] Participation of Non-member States. (2023, October 7). Retrieved from https://www.coe.int/ru/web/conventions/participation-of-non-member-states.
[97] The Convention, Article 25(1).

sense, this model could be considered for the development of similar international legal instruments, in particular on the BRICS platform.

References

Chinen, M. (2023). The international governance of artificial intelligence. Edward Elgar Publishing.

Crawford, K. (2022). Atlas of AI: power, politics, and the planetary costs of artificial intelligence. New Heaven: Yale University Press, 2021.

Donahoe, E., & Metzger, M. M. (2019). Artificial Intelligence and Human Rights. Journal of Democracy, 30(2), 115–126. https://doi.org/10.1353/jod.2019.0029.

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. Retrieved from https://openlibrary.org/books/OL26681102M/Automating_Inequality.

Filgueiras, F. (2022). Artificial Intelligence Policy Regimes: Comparing Politics and Policy to National Strategies for Artificial Intelligence. Global Perspectives, 3(1). https://doi.org/10.1525/gp.2022.32362.

Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. Defense and Security Analysis, 35(2), 147–169. https://doi.org/10.1080/14751798.2019.1600800.

Koskenniemi, M. (2006). From Apology to Utopia: The Structure of International Legal Argument. Cambridge University Press.

Lee, J. (2022). Artificial intelligence and international law. Singapore: Springer.

Meltzer, J. P. (2018, December 13). The impact of artificial intelligence on international trade. Brookings. Retrieved from https://www.brookings.edu/research/the-impact-of-artificial-intelligence-on-international-trade/#footnote-1.

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. Philosophical Transactions of the Royal Society a Mathematical Physical and Engineering Sciences, 376(2133), 20180089. https://doi.org/10.1098/rsta.2018.0089.

Van Kolfschooten, H., & Shachar, C. (2023). The Council of Europe's AI Convention (2023–2024): Promises and pitfalls for health protection. Health Policy, 138, 104935. https://doi.org/10.1016/j.healthpol.2023.104935.

10. HUMAN CAPACITY (ABILITY)-CENTRED AI POLICY: EURASIAN AND TRANSATLANTIC SAFETY DIALOGUE

YONAH WELKER

Abstract. The Bletchley Declaration was signed by 28 countries that agreed on a risk-based approach to frontier AI models, including areas of social protection, health, education, labor. It involved African nations, such as Nigeria, Kenya and Rwanda, countries from the Middle East, including Saudi Arabia and the United Arab Emirates; and major Western economies, such as Canada and the US. Emerging AI policies and frameworks make an attempt to categorize AI systems based on risks, related compliance frameworks and explanations. Such mechanisms are aimed at both regulating and facilitating a human-centered approach to AI systems development,connecting stakeholders and broader society. However, existing approaches to understanding highand unacceptable-risk systems still miss disability-specific vocabulary, scenarios and associated risks, categorization of impairments, spectrums, actions and non-actions, and complex understanding of intersectionality behind it. It includes not only the areas of law enforcement, police, biometrical and public security systems, but less covered areas of silos, misuse or manipulation presented by autonomous systems.
Keywords. accessibility, disability, AI, safety, policy, ethics

INTRODUCTION

There is an estimated 1 billion people — 15% of the world — live with disabilities[98], according the World Health Organization (WHO). And 80% of those people live in developing countries. Historically, individuals with disabilities were excluded from the workplace, educational system, and sufficient medical support. For instance, around 50-80% of the population with disabilities1 are not employed full time, 50% of children with disabilities in low- and middle-income countries are still not enrolled in school, public spaces meet only 41.28% to 95% [99]of the expectations of people with disabilities, and only 10% of the population have access to assistive technologies. For cognitive disabilities, the level of discrimination is even higher. The unemployment rate among those with autism may reach [100]85%, dependent on the country; while among people with severe mental health disorders, it can be between[101] 68%-83%, and for those with Down's syndrome, 43%.

Along with exclusion, individuals with disabilities are disproportionally affected by unjust law enforcement, violence and brutality. Persons with disabilities were victims of 26% of all nonfatal violent crime. 30-50% of individuals subject to the use of force or killed by police have a disability. People with intellectual disabilities are seven times more likely to be sexually assaulted than members of the general population. About one-third of young children and teenagers with disabilities faced emotional and physical abuse.

As for conflicts and crises, people with disabilities are also recognized as among the most marginalized and at-risk population. An estimated 9.7 million people with disabilities are forcibly[102] displaced as a

[98] https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities/disability-and-employment.html
[99] https://www.researchgate.net/publication/341493122_Disability-friendly_public_space_performance
[100] https://link.springer.com/article/10.1007/s40489-014-0041-6
[101] https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-020-00399-0
[102] https://www.hrw.org/news/2018/12/03/un-wars-impact-people-disabilities

result of conflict and persecution and are victims of human rights violations and conflict-related violence. As a result, these groups are also more affected by posttraumatic disorders and conditions.

Finally, there is a strong component of intersectionality behind disabilities that may amplify this exclusion and discrimination, including aspects of demography, co-occurring conditions and socioeconomic factors. For instance, individuals with learning disabilities also experience mental health problems, with estimates suggesting that between 25 and 40% [103]fall into this category. Girls are often diagnosed at a much lower rate than boys, with a ratio of 4:1, and may also be misdiagnosed due to different manifestations. Certain ethnic and social groups [104]have been historically excluded from research data and resources.

10.1. AI Systems and Disability Support

It's important to highlight that ethically developed and implemented assistive technologies can eliminate particular social barriers and create more accessible workplaces, hiring and learning experiences, and accommodation practices.

For instance, in order to support physical impairments, AI algorithms can be used to augment smart wheelchairs[105], walking sticks[106], geolocation and city tools, bionic and rehabilitation technologies. In the case of sensory impairments, it includes facial and sign recognition for sign language identification and support of deaf individuals[107], and computer vision algorithms that can interpret images and videos and then translate that information into braille or audio output to help individuals with visual impairments.

In the area of cognitive impairments, it includes social robotics and algorithms for emotional training for students with autism[108], wearables and devices that improve emotion recognition[109], and adaptive platforms that support dyslexia and attention deficit and hyperactivity disorders. Such technologies can serve to support the general population as well, including further advancement of healthcare, education, labor and city systems, and support of elders, neurodisabled groups and individuals with psycho-emotional disorders.

10.2. Data, Models and Errors of autonomous systems

Algorithms do not create biases themselves but perpetuate societal inequities and distortions. The reasons behind it include lack of access to data for target populations, the models trained to demonstrate efficiency for broader objectives, but lacking accuracy for specific groups or conditions, historical exclusion from research and statistics, simplification and generalization of the target group's parameters (proxies), subjectiveness introduced to labelled data or models' objectives.

For instance, AI systems are known to be less accurate towards individuals with facial differences or asymmetry, different gestures, gesticulation, speech impairment, or different communication patterns.

---

[103] https://www.learningdisabilities.org.uk/learning-disabilities/help-information/learning-disability-statistics-/187699

[104] https://journals.sagepub.com/doi/10.1177/1362361317722030

[105] https://www.redalyc.org/journal/474/47471676003/html/

[106] https://pubmed.ncbi.nlm.nih.gov/34308631/

[107] https://www.sciencedirect.com/science/article/pii/S2667305321000454

[108] https://pmc.ncbi.nlm.nih.gov/articles/PMC9216612/

[109] https://spectrum.ieee.org/upgraded-google-glass-helps-autistic-kids-see-emotions

It especially affects groups with physical disabilities[110], cognitive and sensory impairments, and autism spectrum disorders. There are examples of direct life-threatening scenarios when police and autonomous security systems[111], or military AI may falsely recognize assistive devices as a weapon or dangerous objects, or misidentify facial or speech patterns. These concerns were raised by UN Special Rapporteur [112]on the Rights of Persons with Disabilities, disability organizations such as EU Disability Forum.

There are a variety of physical, cognitive and social parameters that may lead to errors or inaccuracies towards individuals with disabilities. These errors can be grouped into several categories, including recognition, identification and cues, aids, semantic errors:

- Assistive tools and devices - individuals with disabilities may use a wheelchair, walking stick, rehabilitation or assistive devices, bionic hands or legs, or other tools and devices of different shapes, forms and patterns that may not be properly recognized by autonomous systems;
- Assistance and users - solutions, addressing individuals with disabilities frequently involves not only one end-user but an "ecosystem" of users, such as family members, and caregivers. For instance, specialized solutions for autism frequently involve two interfaces - one for the parent, and one - for the child. Public and city systems may not take it into consideration;
- Physical impairments. A person with a disability may lack particular limbs, or have different body shape, posture, and movement pattern, making it more difficult for proper recognition;
- Visual impairments. Blind persons and those with a visual impairment may not not properly understand visual cues given by automated systems;
- Hearing impairments. Individuals with hearing impairments may not hear and comply with audible commands or warnings, making it especially cautions for police and law- enforcement systems;
- Speech impairments. Neurological conditions may affect speech and the ability to communicate, thus not meeting "typical" speech patterns;
- Cognitive impairments. Individuals with cognitive disabilities may communicate differently, lack emotional recognition or social skills;
- Behavioral and psychomotor patterns - individuals with disabilities may exhibit a different pattern of user behavior related to attention span, activities and cognitive parameters;
- Facial recognition that may not identify persons with eye deviation or facial neuropathy;
- Tactile recognition that is built on the assumption that everyone has hands, fingers, and fingerprints and has similar tactile parameters excludes many individuals with disabilities
- Semantic, intersectional, age and other biases - systems may add negative connotations to disability keywords for individuals of particular ethnicities. Besides, algorithms may perpetuate existing ageism[113].

Each parameter alone or in combination with others may lead to greater inaccuracies presented by autonomous systems.

[110] https://disabilitystudies.nyu.edu/disability-bias-and-ai-report/
[111] https://international-review.icrc.org/articles/the-risks-of-autonomous-weapons-analysis-centred-on-rights-of-persons-with-disabilities-922
[112] https://documents-dds-ny.un.org/doc/UNDOC/GEN/N22/433/14/PDF/N2243314.pdf?OpenElement
[113] https://link.springer.com/article/10.1007/s00146-022-01553-5

These risks might be also affected by parameters of computing or physical chains. In particular, *supervised learning*[114], a category of machine learning that uses labeled datasets to train algorithms to predict outcomes and recognize patterns, is known for human-induced errors during the selection, labeling or existing in pretrained models (smart glasses and computer vision, visual objects), *unsupervised learning* – statistical lack of input, representation, raw data can reinforce social disparities and dismiss particular populations (e.g. DNA data clustering for medical solutions), *reinforcement learning* – environment driven errors, "problem of initial experience", experiment's limitations (e.g. learning based on a "reward system", social robotics and assistants). Data points may not exist for certain groups, identities or communities. People who collect or label data may introduce subjectiveness (reporting, selection, systemic or group attribution errors), lack evidence or access to target population. Errors can be also driven by model objectives and constraints.

As for physical chain, risks can be affected by physical human-robot Interaction, issues in balance and stability, durability and robustness, dexterity and haptic manipulation, motion and sensing components safety (servos and kinematics components related to the robots physical reliability and agility; touch, feedback, visual or voice sensors, 3D/depth cameras, LiDARs to collect data for mobility and task processing analysis, spatial intelligence), power components and environmental safety, quality of production and training cycle –planning and control, testing and simulation, sensing and perception.

10.3. GENERATIVE AI AND LANGUAGE MODELS - OPPORTUNITIES AND RISKS

Generative AI and language-based models further expand this impact and the R&D behind it. In particular, such systems may fuel existing assistive ecosystems, health, work, learning and accommodation solutions, requiring communication and interaction with the patient or student, social and emotional intelligence and feedback. Such solutions are frequently used in areas involving cognitive impairments, mental health, autism, dyslexia, attention deficit disorder and emotion recognition impairment, which largely rely on language models and interaction.

With the growing importance of web and workplace accessibility, Generative AI-based approaches can be used to create digital accessibility solutions, associated with speech-to-text or image-to-speech conversion. It may also fuel accessible design and interfaces involving adaptive texts, fonts and colors benefiting reading, visual or cognitive impairments. Similar algorithms can be used to create libraries, knowledge and education platforms that may serve the purpose of assistive accommodation, social protection and micro-learning, equality training and policing. Finally, approaches explored through building such accessible and assistive ecosystems may help to fuel the assistive pretext - when technologies created for groups with disabilities can be later adapted for a broader population, including fueling new forms of interaction, learning and creativity, involving biofeedback, languages and different forms of media.

When compared to existing AI systems, however, language-based platforms require even more attention and ethical guidance. In particular, they can imitate human behavior and interaction, involve more autonomy and pose challenges in delegating decision-making. They also rely on significant volumes of data, a combination of machine-learning techniques and the social and technical literacy behind it.

[114] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3724556

There are different ways, in which generative AI-associated systems[115] may pose risks for individuals with disabilities. In particular:

- They may fuel bias in existing systems, such as automated screening and interviews, public services involving different types of physical and digital recognition and contextual and sentiment bias.
- They may lead to manipulative scenarios, cognitive silos and echo chambers. For instance, algorithms were used to spread misinformation among patients during the COVID-19 pandemic.
- Language-based systems [116]may add a negative connotation to disability-related keywords and phrases or provide wrong outcomes due to a public data set containing statistical distortions or wrong entries.
- Privacy - in some countries, governmental agencies were accused of using data from social media without consent to confirm patients' disability status for pension programmes.

10.4. HUMAN-CAPACITY CENTERED AI POLICY AND REGIONAL CONTEXTS

Addressing the AI policy towards groups with disabilities requires complex oversight and assessment. In particular, disability-centered deployment is *multimodal and multisensory* – it involves visual, hearing, cognitive parameters, necessity of accuracy for different modalities, *It's modular* - may involve interconnected devices and interfaces, It's *multistakehold*er – it may involve families, caregivers. It also requires *Identifying misuse*, actions and non-actions (omissions), manipulation, addictive design, specific attention to data, models and systems oversight, privacy and consent.

For instance, AI-driven dashboards for children with cognitive disabilities may have 2 interfaces – one for the child and one for the parent, solutions can be *data-interconnected* (dashboards and interfaces for analytics and tracking, compact wearable trackers, smart glasses helping with recognition and learning, social assistants and companions). users can have tactile impairment, differences in the accuracy of color memory and search, sound and sight sensitivity)

As for the assistive systems regulation, some facial recognition systems [117]used ear shape or the presence of ear canal to determine whether or not an image included a human face. However, this system didn't learn from sufficient patterns to recognize people who lack these parts or suffer craniofacial syndromes. Medical assessment and analytics systems are known to be created based on "normalized attributes" demographic and health groups. However, it may predominantly exclude some conditions or parameters for younger patients, attributing it only to older groups.

As for the medical data, in some countries immigrants with disabilities tend [118] to *avoid medical examinations* and tests in fear to being deported or face high medical costs which lead to misrepresentation in available medical data sets. People with disabilities may have additional conditions and impairments which do not exist in data sets (e.g. allergies, digestive system disorders). Particular social groups *more likely report concerns* related to cognitive disabilities due to the better medical and educational access. Conditions affecting general population are presented with more

[115] https://www.sciencedirect.com/science/article/abs/pii/S0003999324011912
[116] https://arxiv.org/pdf/2402.01732
[117] https://www.sciencedirect.com/science/article/abs/pii/S0925231223002825
[118] https://pmc.ncbi.nlm.nih.gov/articles/PMC4634824/

sufficient evidence and statistics than rare genetic disorders. Infrastructure and urban datasets used for city planning are known to be "*gender-blind*", affecting accuracy of solutions for women patients.

This complex nature can be addressed through the combination of legal and policy frameworks. For instance, in the European Union disability cases and safety considerations are potentially affected in AI Act, Digital Services Act, data regulation and specific frameworks such as Accessibility Act.

- *Classifications and taxonomies* – Accessibility Act [119]and Standardization directives (e.g. Regulation - 1025/2012)
- *Data profiling, manipulation, addictive design* – AI Act, DSA[120], GDPR
- *Identifying "high-risks" for systems* related to certain critical infrastructures, medical devices, systems to determine access to educational, institutions or for recruiting people, law enforcement – AI Act
- *"Specific transparency risk"*. AI systems such as chatbots or assistive companions should notify users that they are interacting with a machine – AI Act
- *Prohibiting particular use* of affective computing and emotion recognition for publicly accessible spaces - workplaces and educational institutions, law enforcement and migration – AI Act
- *Ensuring code of conduct* for minimal-risk systems, including accessibility ones which meet its requirements.

However, these complex efforts face several challenges at regional level.

- *Local AI solutions.* It's known that even the leading AI models (with 100, 400B
- fail in accuracy for non-English languages [121]or specific environments – indigenous populations, R&D, health, educational environments
- *Accessibility* –The World Health Organization (WHO) estimates that only 1 in 10 people [122]have access to the assistive technology they need
- *Necessity of "Guardian" models* - specialized models addressing fairness and transparency-related features which may complement / track existing ones
- *Area specific literacy and frameworks*. Current efforts include Unesco – AI ethics frameworks and literacy in education[123], WHO – AI in health[124], OECD – disability, AI and labor markets[125], accidents repositories, UNDP's Digital Inclusion in a dynamic world).
- *Controlling vendors influence* - when the same companies invest in data centers and hyperscallers across regions[126], creating data and market silos, limiting competition, and access
- *Other challenges* include digital and physical infrastructure (such as energy and water scarcity), limited cases and taxonomies, not reflecting the uniqueness of historical and social patterns for health and public solutions.

10.5. WAY FORWARD. DISABILITY-CENTERED POLICY, RISKS AND IMPACT ASSESSMENT

[119] https://ec.europa.eu/social/main.jsp?catId=1202
[120] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en
[121] https://www.nature.com/articles/d41586-024-02579-z
[122] https://www.who.int/news-room/fact-sheets/detail/assistive-technology
[123] https://www.unesco.org/en/articles/what-you-need-know-about-unescos-new-ai-competency-frameworks-students-and-teachers
[124] https://www.who.int/publications/i/item/9789240029200
[125] https://www.oecd.org/en/publications/using-ai-to-support-people-with-disability-in-the-labour-market_008b32b7-en.html
[126] https://www.cio.com/article/648048/hyperscalers-in-crosshairs-for-anti-competitive-pricing-and-lock-in.html

Disability is not a monolith, but a spectrum, affected by underlying conditions, demographic, socio-economic and historical criteria. This complexity poses an important reminder that disability exclusion is a social issue first and only then - algorithmic. Existing AI policies and acts attempt to categorize and describe systems through primarily generalized visions of technologies, scenarios and posed risks. These categories do not address specific groups, physical or cognitive differences, unequal access to medical support or education, or economic status.

With more risks of emerging data silos and monopolization of AI development posed by corporate agents, there is an urgent need for collective action to address disability representation in policy development. It includes Introduction of *AI safety institutes, regulatory sandboxes* and testbeds (which involve units of regulation and compliance, accessible engineering and policy coordination), *risk-based categories* (unacceptable, high, low, minimum), *scenarios* (workplaces, education, law-enforcement, immigration), *specific systems considerations* (affective computing, biometrics), *systems taxonomies, frameworks and accidents repositories,* accessible digital and physical infrastructure, specialized policies and Minors Protection (e.g. 193 countries signed commitment to effectively implement children's digital safety - UN's General Assembly's Third Committee[127]).

[127] https://press.un.org/en/2023/gashc4377.doc.htm