BRITISH UNIVERSITY IN EGYPT FACULTY OF LAW
CENTRE OF LAW AND EMERGING TECHNOLOGIES

White Paper on Responsible AI Usage by Social Media Platforms from African Human Rights Law Perspective

Report to funder: The Centre for Intellectual Property and Information Technology Law (CIPIT), Strathmore University

Report authored by:
Mostafa Elkadi, Ibrahim Sabra, Abdelrahman Gamal
Supervised by: Prof. Dr. Hassan Abdelhameed

Strathmore University
Centre for Intellectual Property and Information Technology Law

Table of Contents

Introduction

Today, artificial intelligence (AI) plays a paramount role in a range of sectors, including health, work, communication, education, and law enforcement. AI-driven technology already has a tremendous impact on our daily lives, starting with social media monitoring, smart assistants and facial recognition to self-driving vehicles and autonomous weapons. Technology shapes the way people access, share, and interact with information and is also capable of extracting different particularities about people through their data.

Both public and private sectors are increasingly employing Artificial Intelligence (AI) for varied purposes ranging from tracking and identifying people, evaluating individuals' personality traits or skills, and performing decision making functions.[1] For instance, numerous corporations such as Google, Microsoft, Amazon, and recently Facebook are investing heavily in AI in an attempt to dominate the field.[2]

Due to the widespread usage of AI applications, several countries have adopted national and regional AI strategies.[3] Also, private companies including IBM, Microsoft and Google have

---

[1] Lindsey Andersen, 'Human Rights in the Age of Artificial Intelligence' (Access Now, 2018) <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf> accessed 07 September 2023; David Kaye, 'Report of The Special Rapporteur to The General Assembly on Artificial Intelligence Technologies and Implications for the Information Environment' (UNHRC, 2018) <https://www.un.org/ga/search/view_doc.asp?symbol=A/73/348> accessed 07 September 2023 para 5.

[2] Tom Taulli, 'Facebook AI (Artificial Intelligence): Will M&A Help?' (*Forbes*, 2020) <https://www.forbes.com/sites/tomtaulli/2020/02/15/facebook-ai-artificial-intelligence-will-ma-help/> accessed 07 September 2023; Springboard India, 'Google Vs. Amazon Vs. Microsoft Vs. Facebook — Who Is Leading the AI Race?' (*Medium*, 2019) <https://medium.com/@springboard_ind/google-vs-amazon-vs-microsoft-vs-facebook-who-is-leading-the-ai-race-9e9cefbc5c45> accessed 07 September 2023; Tech Advisor Staff, 'How Tech Giants Are Investing in Artificial Intelligence' (*Tech Advisor*, 2019) <https://www.techadvisor.co.uk/feature/small-business/tech-giants-investing-in-artificial-intelligence-3788534/> accessed 07 September 2023.

[3] Vincent Van Roy, 'AI Watch - National Strategies On Artificial Intelligence: A European Perspective In 2019' (Publications Office of the European Union, 2020) < https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119974/national_strategies_on_artificial_intelligence_final_1.pdf> accessed 07 September 2023; Research and Markets, 'National AI Strategies' (*Research and Markets*, 2019) <https://bit.ly/2ZkztCp> accessed 07 September 2023; Future of Life Institute, 'National And International AI Strategies' (*Future of Life Institute*) <https://bit.ly/389Bu8m> accessed 07 September 2023.

formulated AI principles.[4] Additionally, a number of academics, civil society organisations, and international organisations have managed to set out AI guidelines.[5]

Pacing with this march, the African Commission on Human Rights (ACHPR) announced that there is an urgent "need to undertake a Study on human and peoples' rights and AI, robotics and other new and emerging technologies in Africa".[6] This call on the responsible usage of AI under African Human Rights Law is, in effect, of additional urgency since the ACHPR has been also calling on the African Union (AU) to finalise its policy on Business and Human Rights.[7]

This paper tries to assess whether the African human rights framework is sufficient for business accountability, and more specifically Social Media Platforms (SMPs). This assessment is in relation to the SMPs' usage of AI. Although AI, in this case, may offer benefits in some areas, like spam filters, it, in fact, still presents detrimental implications for a plethora of human rights, particularly the right to privacy, the right to freedom of expression, and the right to health. This paper will discuss all these issues.

All sections follow a solution-oriented approach as they entail a three-phase approach: 1) determining the problem; 2) Finding the solution through analysing international human rights law (IHRL) with African human rights law; and 3) Presenting the conclusions and the relevant recommendations.

---

[4] 'Artificial Intelligence At Google: Our Principles' (*Google LLC*) <https://ai.google/principles> accessed 07 September 2023; 'Microsoft AI Principles' (*Microsoft Corporation*) <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6> accessed 07 September 2023; 'Everyday Ethics For Artificial Intelligence' (*IBM*) <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf> accessed 07 September 2023.

[5] See e.g., UNESCO, 'Recommendation on the Ethics of Artificial Intelligence, SHS/BIO/PI/2021/1' (UNESCO 2021) <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence> accessed 07 September 2023; OECD, 'Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449' (*OECD* 2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 07 September 2023; Amnesty International and Access Now, 'The Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems' (*Access Now* 2018) <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/> accessed 07 September 2023.

[6] 'Resolution On The Need To Undertake A Study On Human And Peoples' Rights And Artificial Intelligence (AI), Robotics And Other New And Emerging Technologies In Africa - ACHPR/Res. 473 (EXT.OS/ XXXI) 2021' (achpr.au.int, 2021) <https://achpr.au.int/en/adopted-resolutions/473-resolution-need-undertake-study-human-and-peoples-rights-and-art > accessed 07 September 2023.

[7] 'Resolution on Business and Human Rights in Africa - ACHPR/Res.550 (LXXIV) 2023' (achpr.au.int, 2023) <https://achpr.au.int/en/adopted-resolutions/550-resolution-business-and-human-rights-africa-achprres550-lxxiv-2023> accessed 10 September 2023.

Section one focuses on the right to privacy and AI technologies as well as the techniques that potentially infringe upon the right to privacy, which encompass, inter alia, profiling, data surveillance, and data sharing. This is especially crucial, given that the Banjul Charter lacks an Article that directly addresses the right to privacy. Such spurs a debate on the recognisability of the right to privacy under African human rights law, especially in light of recent African human rights instruments.

Afterwards, section two of this paper sheds light on some of the AI implications for the right to freedom of expression, in the context of the digital realm. Social media companies, for example, use AI technology to moderate content on their platforms and take down any content that goes against their rules. They also deploy AI tools on numerous occasions to analyse and curate user-generated content to influence what users see on their newsfeed and when they see it.[8] These practices pose serious threats, particularly due to the absence of transparency, accountability, and safeguards policies to mitigate such threats.[9] This section then explores both the international and African human rights jurisprudences, concluding that they provide well-established rules on how to safeguard people's human rights in the era of AI. Accordingly, international and African human rights systems are well-developed to sit as a solid basis for a comprehensive African AI regulatory framework to ensure that human rights lie at the core of the conceptualising, designing, developing, and deploying of AI systems within Africa.

Finally, section three grapples with the right to health in the digital realm, which is an area that is unregulated by both international and African law, yet is still inextricably connected to AI content moderation. The section focuses on the impact of AI systems that operate on SMPs, on the individuals' mental health, and more specifically neurological health. After establishing the

[8] David Kaye, 'Report of The Special Rapporteur to The General Assembly on Artificial Intelligence Technologies and Implications for the Information Environment' (UNHRC, 2018) <https://www.un.org/ga/search/view_doc.asp?symbol=A/73/348> accessed 07 September 2023 para 5.

[9] Dunja Mijatović, 'In the Era of Artificial Intelligence: Safeguarding Human Rights' (*openDemocracy*, 2018) <https://www.opendemocracy.net/en/digitaliberties/in-era-of-artificial-intelligence-safeguarding-human-rights/> accessed 07 September 2023.

link between the technology and health,[10] this section tackles the legal principles that can regulate such matters.[11]

Section 1: AI in Business, and the Right to Privacy

The right to privacy occupies a distinct place in the discussion of the impact of AI in business on human rights. That is because the conversation of privacy under the Banjul Charter starts from the point of debate on whether it exists at all, and then proceeds to explore the limits of this uncharted right. Perceived as an enabler of other human rights, the discussion of the right to privacy in this section commences with a brief survey of the forms of threats emerging from

[10] Kalpathy Ramaiyer Subramanian, 'Product Promotion in an Era of Shrinking Attention Span' (2017) 7 International Journal of Engineering and Management Research; Tatiana de Campos Aranovich and Rita Matulionyte, 'Ensuring AI Explainability in Healthcare: Problems and Possible Policy Solutions' (2022) 32 Information Communications Technology Law 2; See, eg, Ziad Obermeyer and others, 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations' (2019) 366 Science 447; Karen Hao, 'This Is How AI Bias Really Happens-and Why It's so Hard to Fix' (MIT Technology Review, 4 February 2019) <https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/ > accessed 7 September 2023; Jenna Wiens and others, 'Diagnosing Bias in DataDriven Algorithms for Healthcare' (2020) 26 Nature Medicine 25; see also Rich Caruana and others, 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission' [2015] Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Hüseyin Bilal MACİT, Gamze MACİT, and Orhan GÜNGÖR, 'A Research on Social Media Addiction and Dopamine Driven Feedback' (2018) 5 Journal of Mehmet Akif Ersoy University Economics and Administrative Sciences Faculty; Maartje Boer and others, 'Attention Deficit Hyperactivity Disorder-symptoms, Social Media Use Intensity, and Social Media Use Problems in Adolescents: Investigating Directionality' (2019) 91 Child Development e854; Aryn C Karpinski and others, 'An Exploration of Social Networking Site Use, Multitasking, and Academic Performance among United States and European University Students' (2013) 29 Computers in Human Behavior; L.D. Rosen and others, 'The Media and Technology Usage and Attitudes Scale: An Empirical Investigation' (2013) 29 Computers in Human Behavior.

[11] See e.g., Tatiana de Campos Aranovich and Rita Matulionyte, 'Ensuring AI Explainability in Healthcare: Problems and Possible Policy Solutions' (2022) 32 Information Communications Technology Law 2; R Matulionyte, 'Reconciling Trade Secrets and AI Explainability: Face Recognition Technologies as a Case Study' (2022) 44(1) European Intellectual Property Review 3; UN Committee on Economic, Social and Cultural Rights (CESCR), General Comment No. 14: The Right to the Highest Attainable Standard of Health (Art. 12 of the Covenant), 11 August 2000, E/C.12/2000/4 <https://www.refworld.org/docid/4538838d0.html> accessed 10 September 2023; HRC 'Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' (2011) <https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf> accessed 10 September 2023; African Commission on Humans' and Peoples' Rights, General Comment No. 7 on Article 14(1)(d) and (e) of the African Charter on Human and Peoples' Rights: The Right to Participate in Government (2022) <https://achpr.au.int/sites/default/files/files/2022-10/general-comment-7-english.pdf> accessed 10 September 2023; ACHPR Guidelines on the Right to Water in Africa (2022) <https://achpr.au.int/sites/default/files/files/2022-08/eng-achprguidelinesontherighttowaterinafrica.pdf> accessed 10 September 2023; African Commission on Human and Peoples' Rights 'Principles and Guidelines on the Implementation of Economic, Social and Cultural Rights in the African Charter on Human and Peoples' Rights' (2011); Communication 155/96, Social and Economic Rights Action Center (SERAC) and Center for Economic and Social Rights (CESR) v Nigeria, 27 October 2001, para 46; African Commission on Humans' and Peoples' Rights, General Comment No. 7 on Article 14(1)(d) and (e) of the African Charter on Human and Peoples' Rights: The Right to Participate in Government (2022) < https://achpr.au.int/sites/default/files/files/2022-10/general-comment-7-english.pdf> accessed 10 September 2023.

current digital practices, with glimpses on the condition in Africa. Then, the discussion shifts to the legal conundrum of privacy under the Banjul Charter.

I. Redefining Privacy: Privacy Implications of AI in Business

The relationship between modern technologies, particularly AI-based technologies, and personal data has grown beyond any doubt or utopian scepticism. Data is actually the petrol of our modern age and a paramount economic driver.[12] The perspectives from which this robust correlation between AI and ML, and data could be looked at are various. One is certainly the technical necessity of data for the growth of AI. It is a rhetoric now that the quality, amount, diversity of data largely determines the quality of the AI product,[13] which prompts more demand thereof.

Another compelling perspective to see through is the economic value of data. The economisation of data is virtual in the fact that data has become a commodity per se.[14] Economic paradigms are growing data centric. Trading information is not the business of intelligence only anymore, Shoshanna Zubbof has captured the idea of the shift of the role of information for businesses, where information about consumers is curated, traded and used as a food for behavioural analysis and control, calling it 'surveillance capitalism.'[15] Similarly, Sarah West posited the idea of 'data surveillance', arguing that our digital traces are used as a tool for power redistribution based on a capitalist logic that is geared towards those who are able to extract information from this data.[16]

AI is omnipresent, and the channels through which AI can reach and impact personal data are growing. IoT for instance is offering ease and enhanced forms of services, but its ubiquitousness is entrapping.[17] Announced recently, an autonomous home assistant, a robot, will provide interactive services, including connecting and controlling almost all home appliances and

---

[12] Grzegorz Mazurek and Karolina Małagocka, 'Perception of privacy and data protection in the context of the development of artificial intelligence')2019(6 (4) Journal of Management Analytics 344.

[13] Council of Europe, *Artificial Intelligence and Data Protection*, (2019), available at <https://edoc.coe.int/en/artificial-intelligence/8254-artificial-intelligence-and-data-protection.html> accessed 12 August 2023.

[14] Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review 528.

[15] Shoshanna Zuboff, 'Big other: Surveillance capitalism and the prospects of an information civilization' (2015) 30(1), Journal of Information Technology, 75-89.

[16] Sarah Myers West, 'Data Capitalism: Redefining the Logics of Surveillance and Privacy' (2019) 58(1) Business and Society 23.

[17] Jahanzeb Shahid, Rizwan Ahmad, Adnan K. Kiani, Tahir Ahmad, Saqib Saeed and Abdullah M. Almuhaideb, 'Data Protection and Privacy of the Internet of Healthcare Things (IoHTs)' (2022) 12(1) Applied Sciences Review.

devices,[18] collecting tons of data along the way. The modern consumer would find eyes and ears in every corner of the *smart life*. Between the smart homes with all the networked appliances, smart cars, which along with navigation applications save your habitual destinations, and even smart gadgets attached to one's very body and collecting his vital signals, if misused can draw a very elaborate chart of a consumer's habits and tendencies,[19] and could be transferred to a third party, such as SMPs without users' knowledge.

Another technology is the large language models (LLMs), which are integrated in numerous services such as customer support, education and virtual assistants, among many others.[20] Application of such advanced interactive abilities, have they saved personal data, or at least chat logs of their users, shall be able to draw accurate conclusions about them.[21] ChatGPT for example has achieved rocket rise in usage in different mediums and for different ends. It is expected to hold a promise for Africa that it could be used in healthcare, agriculture and language preservation,[22] yet its handling of data is concerning.[23]

In Africa, as well as anywhere else, social media platforms are deemed to be the leviathan of data attractability and vulnerability, together. SMPs have grown into a large business, capitalising on the personal data of its users, presenting marketeers, statisticians, and politicians a great competitive advantage. SMPs are placed perfectly on the top of the digital ecosystem pyramid, commanding an unrivalled position in data collection, user engagement, and targeted advertising. Such could be attributed to two features that exclusively combine in the SMPs.

The first feature is SMPs ability to keep users attached to their services for extended periods of time. Studies suggest that an average working-age internet user spends 2.5 hours a day on SMPs

---

[18] Samsung, 'A Day in the Life with Ballie: An AI Companion Robot for the Home' (8 January 2024) <https://news.samsung.com/us/samsung-ballie-ai-companion-robot-home-video-ces-2024/> accessed 10 January 2024.
[19] Andrew Laughlin, 'The Smart Device Brands Harvesting Your Data' *Which?* (7 September 2023) <https://www.which.co.uk/news/article/the-smart-device-brands-harvesting-your-data-al4vp6Z3ePDf> accessed 10 December 2023.
[20] Walid Hariri 'Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing' arXiv preprint arXiv:2304.02017 (2023).
[21] Maanak Gupta, Charan Kumar Akiri, Kshitiz Aryal, Eli Parker and Lopamudra Praharaj, 'From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy' (2023) 11 IEEE doi:10.1109/ACCESS.2023.3300381.
[22] Tshilidzi Marwala, 'ChatGPT In Africa: Opportunities And Challenges' (2023) Forbes Africa <https://www.forbesafrica.com/opinion/op-ed/2023/06/28/chatgpt-in-africa-opportunities-and-challenges/> accessed 30 December 2023.
[23] Shiona McCallum, 'ChatGPT banned in Italy over privacy concerns' *BBC*, (1 April 2023) <https://www.bbc.com/news/technology-65139406> accessed 15 June 2023.

in 2023.[24] By holding users' interest, SMPs become an irresistible beacon for advertisers' dollars. The second feature is the wealth of data SMPs get to gather about users' preference, habits and personalities, which enables them to serve users' interests to third party businesses on a golden plate. Being the principal public sphere of the era, SMPs offer their users a versatile forum.[25] They enable users to communicate privately and publicly, beside hosting contents that act as a trigger of user reactions, such as trends, news, and events, in addition to a whole lot of other options that differ by platform. All of which heighten the accuracy of character depiction the SMPs is capable of. Such undoubtedly gives a competitive advantage to SMPs. Estimations of SMPs revenues show that social media advertisement global spending stood at approximately $270 billion in 2022 and is expected to rise to $300 billion in 2023.[26] SMPs use the data in their disposition to make direct monetary benefits primarily through two means: the sale of data to third parties; or the targeted advertising.[27]

Barry Brown has noted a contradiction in the behaviour of internet users between their announced positions and actual conduct towards their privacy, calling this discrepancy a 'privacy paradox.'[28] Internet users show willingness to trade their privacy for other considerations.[29] Personalisation is one of the considerations that reveal an inconsistent user behaviour, as one would prefer personalised advertisements and products, which itself necessitates large amounts of data, while wanting to protect their privacy.[30] Personalisation is 'the ability to proactively tailor products and product purchasing experiences to tastes of individual consumers based upon their personal and preference information.'[31] SMPs are very

---

[24] Simon Kemp, 'DIGITAL 2023 DEEP-DIVE: HOW MUCH TIME DO WE SPEND ON SOCIAL MEDIA?' (26 January 2023) *Datareportal* <https://datareportal.com/reports/digital-2023-deep-dive-time-spent-on-social-media#:~:text=The%20company's%20latest%20data%20reveals,at%20the%20start%20of%202022.> accessed 30 December 2023.

[25] Erlis Çela, 'Social Media as a New Form of Public Sphere' (2015) 2(3) European Journal of Social Sciences Education and Research 127.

[26] See 'Social media advertising and marketing worldwide - statistics & facts' *(Statistia, 18 December 2023)* <https://www.statista.com/topics/1538/social-media-marketing/#topicOverview> accessed 8 November 2023.

[27] Sabrina Karwatzki, Olga Dytynko, Manuel Trenz and Daniel Veit, 'Beyond the Personalization–Privacy Paradox: Privacy Valuation, Transparency Features, and Service Personalization' (2017) 34(2) Journal of Management Information Systems 370.

[28] Barry Brown, 'Studying the Internet Experience' (HP Laboratories Bristol, HPL-2001-49, 2001), <https://www.hpl.hp.com/techreports/2001/HPL-2001-49.pdf> accessed 17 August 2023.

[29] Juan Pablo Carrascal, Christopher Riederer, Vijay Erramilli, Mauro Cherubini and Rodrigo de Oliveira, 'Your browsing behavior for a Big Mac: economics of personal information online' (22nd international conference on World Wide Web, May 13–17, 2013) 189-200.

[30] Elizabeth Aguirre, Anne L. Roggeveen, Dhruv Grewal and Martin Wetzels, 'The personalization-privacy paradox: implications for new media' (2016) 32(2) Journal of Consumer Marketing 99.

[31] *Ibid.*

adept in this regard to the extent of choosing not to overly customise their advertisements, lest they appear intrusive.[32]

Personalisation of content is normally associated with other practices that can equally threaten privacy. Algorithmic profiling is of high relevance for example. Profiling is the 'systematic and purposeful recording and classification of data related to individuals.'[33] Relatedly, microtargeting is the black horse of modern time for both marketing and politics. Microtargeting is "a form of online targeted advertising that analyses personal data to identify the interests of a specific audience or individual in order to influence their actions".[34] c

Generally, AI tools provide precise audience-targeting models which advertisers use for commercial and political marketing.[35] These techniques raise not only concerns about data proportionality, but also about the collection in the first place, as it can be conducted covertly, using means such as the cookies.[36]

II. Giving Path to the Right to Privacy in Africa

Seeking guidance on the right to privacy, amidst this technological fuss, would normally require a legal provision, i.e., text of law, for normative reference. Discerning the practical contours and the actual standard of privacy in Africa would be a result of an accumulative jurisprudence of case law, indirect sources, and other soft law. Unlike freedom of expression or the right to health, the discussion of business usage of AI in Africa in relation to privacy starts from a rather earlier point where the recognition of the right to privacy altogether is brought to question. Surely, domestic law, comparative law, and primarily international human rights law (IHRL) could provide a rich reference for the African efforts, but not before deciding first whether the protection of African law extends, or should extend, to the right to privacy or not.

---

[32] John T. Gironda, P. Korgaonkar 'iSpy? Tailored versus Invasive Ads and Consumers' Perceptions of Personalized Advertising' (2018) 29 Electronic Commerce Research and Applications 64-77; Baek, Tae and Morimoto, Mariko 'Stay away from me' (2012) 41(1) Journal of Advertising 59-76.

[33] Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, Shruthi Velidi and Salome Viljoen, 'The chilling effects of algorithmic profiling: Mapping the issues'(2020) 36 Computer Law and Security Review 3.

[34] UK Information Commissioner's Office, 'Social media privacy settings: Microtargeting' (ICO) <https://ico.org.uk/your-data-matters/be-data-aware/social-media-privacy-settings/microtargeting/>.

[35] David Kaye, 'Report of The Special Rapporteur to The General Assembly on Artificial Intelligence Technologies and Implications for the Information Environment, A/73/348' (UNHRC, 2018) <https://www.un.org/ga/search/view_doc.asp?symbol=A/73/348> para 17.

[36] Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' (2020) 35(2) Berkeley Technology Law Journal 370.

The deliberate omission of the right to privacy from the African Charter[37] that was adopted in relatively modern time, June 1981, has triggered a debate as to whether this is just a reflection of the absence of the right to privacy in Africa. Some commentators take the position that the right to privacy is indeed non-recognizable under African law. This is due to collectivist African culture that confers little weight to individual rights in favour of the values of the community.[38] As such, the Banjul Charter was considered as a *lex imperfecta*, which can only witness progress through amendment to include an express provision on privacy.[39]

Amending the Banjul Charter to incorporate the right to privacy would be the loudest statement of recognition. Nonetheless, the present African landscape accommodates the right to privacy, ergo it is recognisable. Culturally, a shift in the view of privacy could be noted, reflecting a rise in individualistic tendencies. This shift is largely induced by the growing use of SMPs, and broadly digital devices associated with private use, mainly mobile phones.[40] SMPs have created a digital space for self-expression and communication that is easy and not limited by the restraints of the family, the clan or the community,[41] which have significantly boosted this shift.

On the official level, this dynamism has stimulated domestic and regional reaction in an attempt to regulate data protection. By 2023, 35 of the 55 AU States have enacted data protection laws. Regionally, the ECOWAS Supplementary Act adopted in 2010 has been an early step with influence on West African laws.[42] On a larger scale, the AU Convention on Cyber Security and Personal Data Protection (Malabo Convention), adopted in 2014, is expected to have a significant impact on the data protection landscape in Africa now that it has finally come into effect on 8 June 2023. Furthermore, in 2019, the Declaration of Principles of Freedom of Expression and Access to Information in Africa has put forward the right to privacy, addressing

---

[37] The first draft of the African Charter on Human and Peoples' Rights contained an article, numbered 24, on the right to privacy, which was presented in the Meeting of Experts for the Preparation of the Draft African Charter of Human and Peoples' Rights in Dakar, Senegal from 28 November to 8 December 1979, CAB/LEG/67/1.

[38] Lee A. Bygrave, 'Privacy and Data Protection in an International Perspective' (2010) 56(8) Scandinavian studies in law 175; Lee A. Bygrave, *Data Privacy Law: An International Perspective* (Oxford University Press, 2014) 80; Alex Makulilo, 'A person is a person through other persons: critical analysis of privacy and culture in Africa' (2016) 7(1) Beijing Law Review 194.

[39] Yohannes Eneyew Ayalew, 'Untrodden paths towards the right to privacy in the digital era under African human rights law' (2022) 12(1) International Data Privacy Law 18.

[40] Jake Reichel, Fleming Peck, Mikako Inaba, Bisrat Moges, and Brahmnoor Singh Chawla, 'I have too much respect for my elders': Understanding South African Mobile Users' Perceptions of Privacy and Current Behaviors on Facebook and WhatsApp' (29th USENIX Security Symposium. August 12–14, 2020) 1951.

[41] John Suler, 'The Online Disinhibition Effect' (2004) *7*(3) Cyber Psychology and Behavior, 321–326. https://doi.org/10.1089/1094931041291295.

[42] Graham Greenleaf, Bertil Cottier, 'International and Regional Commitments in African Data Privacy Laws: A Comparative Analysis' (2022) 44 Computer Law and Security Review 5.

primarily digital privacy. These instruments, on part of official regional and sub-regional African entities, indicate that the recognition of the right to privacy under African law is indeed going at a pace faster than that depicted in some of the literature. It would even indicate that collectivism, as an inhibitor of privacy, is not too dominant in African culture as often portrayed.[43] At least not to the extent that makes any identification of individual rights unachievable.

The way forward for the right to privacy could be paved through other means of interpretation as well. Calls have been made to put the right to privacy to action, although not stipulated in the Banjul Charter.[44] Appeal to IHRL could give effect to the right to privacy, thanks to Article 60 of the Banjul Charter. Article 60 binds the ACHPR to 'draw inspiration' from IHRL, with special reference to UN jurisprudence.[45] As such, this inspiration that the ACHPR *shall* seek in comparative law, would include sources such as Universal Declaration of Human Rights (UDHR) and the International Covenant of Civil and Political Rights (ICCPR), in addition to their explanatory notes e.g., General Comment 16.[46] Article 60 also directs the ACHPR to seek guidance in African national laws, which opens the door for a sensitive and receptive reactiveness from the ACHPR towards the domestic efforts, if employed on the right to privacy.

This understanding could be further backed by international law. The rules of interpretation of the VCLT require that the process of treaty interpretation to be contextual, taking into account the whole body of the treaty including the other provisions other than the one in question, the preamble, the annexes, and any related agreement or instrument.[47] Even more, Article 31 imposes a consideration of other factors such as "*any subsequent practice in the application of the treaty which establishes the agreement of the parties regarding its interpretation; any*

---

[43] Nkonko M. Kamwangamalu, 'Ubuntu in South Africa: a sociolinguistic perspective to a pan-African concept' (1999) 13(2) *Critical Arts* 27; Liberty Eaton and Johann Louw, 'Culture and Self in South Africa: Individualism-Collectivism Predictions' (2000) 140(2) The Journal of social psychology 210.

[44] Avani Singh and Michael Power 'The Privacy Awakening: The Urgent Need to Harmonise the Right to Privacy in Africa' (2019) 3 African Human Rights Yearbook 202, 211

[45] Article 60 reads "*The Commission shall draw inspiration from international law on human and peoples' rights, particularly from the provisions of various African instruments on Human and Peoples' Rights, the Charter of the United Nations, the Charter of the Organisation of African Unity, the Universal Declaration of Human Rights, other instruments adopted by the United Nations and by African countries in the field of Human and Peoples' Rights, as well as from the provisions of various instruments adopted within the Specialised Agencies of the United Nations of which the Parties to the present Charter are members."*

[46] Human Rights Committee, 'General Comment No. 16: Article 17 (The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation)' (8 April 1988) UN Doc CCPR/C/GC/16.

[47] Article 31(1,2) Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1155 UNTS 331.

relevant rules of international law applicable in the relations between the parties."[48] Working Article 31, in this sense, would give extra relevance to the African instruments mentioning the right to privacy, such as the ACRWC, which was adopted in 1990, and explicitly recognises the right to privacy for children.[49]

Most importantly, the interpretation of the Banjul Charter, as a human rights instrument, the primary in Africa, shall be contextual and purposive.[50] Purposive interpretation of the Banjul Charter requires that the right to privacy, although has no express mention, is present as an underlying value, or rather a prerequisite for other textual rights to realise their full application, such as the right to integrity, dignity, liberty and security, and the right to health.

This purposive approach is not new to the ACHPR. As a matter of fact, it has previously recognised the right to housing or shelter and the right to food into the Banjul Charter, despite not being explicit therein. In the Ogoni case, the ACHPR ruled in favour of the Ogoni identifying the Nigerian government's violation of Ogoni people rights' including the said right. The ACHPR has insightfully reasoned that:

Although the right to housing or shelter is not explicitly provided for under the African Charter, the corollary of the combination of the provisions protecting the right to enjoy the best attainable state of mental and physical health, cited under article 16 above, the right to property, and the protection accorded to the family forbids the wanton destruction of shelter because when housing is destroyed, property, health and family life are adversely affected. It is thus noted that the combined effect of articles 14, 16 and 18(1) reads into the Charter a right to shelter or housing which the Nigerian government has apparently violated.[51]

What is particular about this reasoning is that it did not just form a silent recognition. The ACHPR has particularly tackled the note that the right to housing or shelter are not explicitly provided for under the African Charter. The Commission advanced what it called "the combined

[48] Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1155 UNTS 331.
[49] Art 10 of the African Charter on the Rights and Welfare of the Child (adopted 11 July 1990, entered into force 29 November 1999) CAB/LEG/ 24.9/49.
[50] Christine M. Chinkin, 'Human rights', in Michael J. Bowman and Dino Kritsiotis, *Conceptual and Contextual Perspectives on the Modern Law of Treaties* (Cambridge University Press, 2018) 520.
[51] Social and Economic Rights Action Centre and another v Nigeria (2001) Communication 155/96, ACmHPR 2001 para 43.

effect" of the other relevant rights, namely the rights to property[52], health[53] and family life[54] whose satisfaction cannot be without reading the right to privacy into the Charter.[55]

To this end, the right to privacy could be said to be conceptually recognisable under the Banjul Charter, and already recognised under the African law in general. However, the operationality of such a right, namely how it would apply, remains since it was not practically tested under the African conditions. With the plethora of privacy threats business use of AI brings about, this challenge is exacerbated. To this moment, two factors contribute to an indecisive understanding of the limits of the right to privacy. First, the lack of a primary textual reference e.g., an express provision in the Banjul Charter or another parallel treaty, which is largely the main source of normative power of the right. Second, the lack of jurisprudence on the right to privacy by the ACHPR and the African Court on Human and Peoples' Rights, which could have supplied an analysis of the African understanding of the right to privacy.

Nevertheless, the African landscape is not void of any assisting sources in this regard. We can seek help in two directions. Regarding the broader framework of privacy, resort can be made to the three-part test as an established framework for limitations under the UDHR and ICCPR, where an interference against the right to privacy shall be prescribed by law, pursuing a legitimate aim, necessary in a democratic society. Such a test is endorsed and recognised by rich European jurisprudence.

Additionally, the few, but useful to this end, African sources tackling the right to privacy, can provide guidance on its gist and boundaries in Africa. The ACRWC for instance prohibits *arbitrary or unlawful* interference with a child's privacy, family home or correspondence.[56] Also, Principle 9 of the 2019 African Declaration recognises the three-part test as well, requiring that *the limitation:*

> *a. is prescribed by law;*
>
> *b. serves a legitimate aim; and*
>
> *c. is a necessary and proportionate means to achieve the stated aim*
>
> *in a democratic society.*

[52] Article 14 of the Banjul Charter.
[53] Article 16 of the Banjul Charter.
[54] Article 18(1) of the Banjul Charter.
[55] *Ogoni,* para 60.
[56] Art 10 of the African Charter on the Rights and Welfare of the Child (adopted 11 July 1990, entered into force 29 November 1999) CAB/LEG/ 24.9/49.

As for the second direction, African data protection law is significantly enriched by the adoption and coming into force of the Malabo Convention that has put in place a whole framework of data protection. And given that the practice has shown that data protection law *per se* plays a paramount role as bulwark vis-à-vis the practices targeting data, being the grandiose threat of SMPs and internet generally.

As will be discussed, the Malabo Convention does indeed provide a substantive protection to the right to privacy, specifically data protection. But it does further emphasise the conceptual perception of the right to privacy in Africa. This is especially notable in the fact that the Malabo Convention is the first large-scale primary source that explicitly affirms the right to privacy as an objective and consideration of African law. In fact, the drafters of the Malabo Convention have made several mentions of the right to privacy, in different points and different phrasings. This, however, is not particularly the habit of its parallel data protection conventions of regional standing, which rarely, if at all, even mention the word 'privacy.'

This could be noted in several instances in the Convention. The preamble, for example, indicates in a resounding statement that *the goal* of this convention is "to establish in each State party a mechanism capable of combating violations of privacy that may be generated by personal data collection, processing, transmission, storage and use".

It is also worth noting that the Convention placed additional emphasis on the right to privacy by mentions such as "*each State Party shall ensure that the measures so adopted will not infringe on… other basic rights such as freedom of expression, the right to privacy and the right to a fair hearing, among others."*[57] Also, that transborder transfer can only be done to a State with an adequate level of protection of privacy, freedoms and fundamental rights.[58] This persistent emphasis of the right to privacy in the context of data protection could be an attempt from the African lawmaker to compensate for the lack of regulation of privacy, and particularly the view that it has no room in the African culture, being social or legal, altogether.

The Malabo Convention provides broad-ranging protection of personal data. It follows a structure not dissimilar to the GDPR. It identifies six principles of data protection that serve as a safety net for the other data protection rules.[59] Yet, the principles of the Malabo Convection are expected to play a much bigger role in the jurisprudence of the African Court and Commission

[57] Article 25.3.
[58] Article 15.
[59] These are consent and legitimacy, lawful and fair processing, purpose, relevance and retention of data, accuracy of data over its lifespan, transparency of processing, confidentiality and security of personal data, see *Ibid* Article 13.

given that the Convention is brief on rules. It further recognises a set of data subject rights i.e., the right to information,[60] the right of access,[61] the right to object,[62] and the right of rectification or erasure[63]

Overall, one of the perks of the Malabo Convention to business, and the privacy landscape in Africa in general, is that it can reconcile the different African views of underlying the existing data protection laws, setting a minimum standard. Also, it can provide guidance to future legislative efforts.[64]

III. Recommendations

- The ACHPR should act promptly and decisively to formally recognize the right to privacy and set a clear, continent-wide standard for privacy protection in Africa.
- On a priority-basis, launching campaigns on the national level sponsored by the AU, to raise awareness and educate SMPs users, and internet users in general, about the right to privacy.
- Encourage SMPs to implement efficient self-regulation by upholding an equal standard of privacy to the standard.
- Urge SMPs to adopt effective self-regulation practices, ensuring their privacy standards align closely with established standards in other continents.
- In the interim, before establishing a unique African legal perspective on privacy, stakeholders are advised to integrate standards derived from IHRL and comparative law.
- Facilitate access to privacy justice by ensuring individuals can readily approach courts or independent bodies for redress.
- Urge research bodies and civil society groups to actively participate in shaping Africa's approach to privacy by undertaking detailed studies and publishing their findings on privacy status in Africa.

[60] Article 16.
[61] Article 17.
[62] Article 18.
[63] Article 19.
[64] African Union, 'AU Data Protection Policy' (2022); Mohamed Aly Bouke, Sameer Hamoud Alshatebi, Azizol Abdullah, Korhan Cengiz, Hayate El Atigh, 'African Union Convention on Cyber Security and Personal Data Protection: Challenges and Future Directions' arXiv preprint arXiv:2307.01966 (2023) 4.

Section 2: The Impact of SMPs AI-Powered Content Moderation on Freedom of Expression

According to the United Nations (UN) and other regional organisations, human rights that people enjoy offline are protected online as well. The extensive deployment of AI tools across the Internet, particularly by social media platforms, may play an important role within the complex online ecosystem; however, it might have problematic consequences that require attention.[65] This section first focuses on possible ramifications of AI-powered content moderation, the main AI application, used by social media platforms, on the right to freedom of expression. Second, it delves into the African human rights system, which is built around the principles of fundamental rights, democracy, and rule of law, to assess whether it could stand as a coherent and robust legal framework that can capture the ramifications of AI tools deployed by these social media giants on freedom of expression and emanating rights.

I. Social Media's AI-Powered Content Moderation Poses Serious Implications on Freedom of Expression

Content moderation usually functions according to enigmatic rules that lack reference to human rights standards and are set by private actors. Content moderation is also subject to pressure that governments impose on social media platforms in order to monitor and remove user-generated content, using rapid automatic filters to attain more effective results in the fight against alleged illegal content on the Internet.[66] Advertisers also influence online content moderation since they pay millions of dollars per year for online ads.[67] These practices, coupled with employing AI

[65] David Kaye, 'Report of The Special Rapporteur to The General Assembly on Artificial Intelligence Technologies and Implications for the Information Environment, A/73/348' (UNHRC 2018) <https://www.un.org/ga/search/view_doc.asp?symbol=A/73/348> accessed 30 December2023 para 9.

[66] European Commission, 'Recommendation On Measures To Effectively Tackle Illegal Content Online C(2018) 1177' (European Commission 2018) <https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online> accessed 30 December 2023; Javier Pallero, 'Honduras: New Bill Threatens To Curb Online Speech - Access Now' (*Access Now* 2018) <https://www.accessnow.org/honduras-new-bill-threatens-curb-online-speech/> accessed 30 December 2023; Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act) 2017; National Cohesion and Integration Commission and the Communications Authority of Kenya, 'Guidelines for Prevention of Dissemination of Undesirable Bulk Political SMS and Social Media Content via Electronic Communications Networks' (The Communications Authority of Kenya 2017) <https://ca.go.ke/wp-content/uploads/2018/02/Guidelines-on-Prevention-of-Dissemination-of-Undesirable-Bulk-and-Premium-Rate-Political-Messages-and-Political-Social-Media-Content-Via-Electronic-Networks-1.pdf> accessed 30 December 2023; Cybersecurity Law of the People's Republic of China 2016.

[67] James Clayton, 'X ad boycott gathers pace amid antisemitism storm' (*BBC* 2023) <https://www.bbc.com/news/world-us-canada-67460386> accessed 30 December 2023; Tiffany Hsu and Gillian Friedman, 'CVS, Dunkin', Lego: The Brands Pulling Ads from Facebook Over Hate Speech' (*The New York Times* 2020)

extensively in content moderation by social media platforms, can undermine people's right to freedom of expression and other interdependent rights on different levels;[68] by, for example, erroneously removing lawful content in order to avoid liability.[69]

In this regard, this section agrees with both Llansó's argument that, beyond the technical challenges of AI-driven moderation, proactive moderation can act as a prior restraint on speech[70] and Gorwa, Binns, and Katzenbach's view that scepticism regarding "decisional transparency", "justice/fairness" and "de-politicisation" represents a substantial critique of AI-powered automated moderation.[71]

i. AI-Driven Content Moderation Faces Several Technical Challenges

Several studies have shown that AI-driven moderation systems are incapable of assessing, for example, the context, parody, language variations, criticism, and cultural particularities.[72] Research has also demonstrated that AI systems may adopt different architectural approaches, leading to some divergences in their decisions during moderation. This is mainly because despite being trained on the same dataset, neural networks of these systems "may learn to infer the function between inputs and outputs during supervised learning differently. These factors can result in inconsistent detection of harmful content across content types and platforms".[73]

Furthermore, users continuingly attempt to evade automated moderation systems by slightly editing the content they publish using texts, symbols, or images. More importantly, AI can be used to manipulate AI-based moderation systems. For instance, an AI image classifier called ResNet50 faced an AI adversarial attack that made it recognise a police van image as a

---

<https://www.nytimes.com/2020/06/26/business/media/Facebook-advertising-boycott.html> accessed 30 December 2023.

[68] David Kaye, 'Report of The Special Rapporteur on The Promotion and Protection of The Right to Freedom of Opinion and Expression, A/HRC/38/35' (UNHRC 2018) <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement> accessed 30 December 2023 para 1.

[69] Ibid paras 15 and 16.

[70] Emma J Llansó, 'No Amount Of "AI" in Content Moderation Will Solve Filtering's Prior-Restraint Problem' (2020) 7 Big Data and Society.

[71] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in The Automation of Platform Governance' (2020) 7 Big Data and Society.

[72] David Kaye, 'Report of The Special Rapporteur on The Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/38/35' (UNHRC, 2018) <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement> accessed 30 December 2023 para 56.

[73] Cambridge Consultants on Behalf of Ofcom, 'Use of AI in Online Content Moderation' (Ofcom, 2019) <https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf> accessed 30 December 2023 42.

typewriter with 98% confidence, but, before the attack, it correctly identified the image with 85% confidence.[74] These technical limitations show that automated moderation tools are not reliable to work at scale and, therefore, reliance on automated tools should not increase as "automation is not a sufficient replacement for having a human in the loop".[75]

ii. AI-Powered Content Moderation May Act as a Prior Restraint

Prior restraints refer to situations where approval from a third party is required to publish content.[76] Under international human rights law, a well-established presumption against the legality of prior restraints holds that people should be free to express their opinions and face consequences only if they violate rules. This applies to the online atmosphere, since published content endures expanded scrutiny by pre-moderation AI systems that deem any content a potential violation of the rules.[77] Some researchers argue that AI-powered content-filtering tools represent an unjustifiable form of collateral censorship[78] which, according to Jack Balkin, happens when authorities pressure intermediaries to regulate the content of their users. Balkin emphasises that collateral censorship, which he perceives as a distinct technique of speech policing in the digital era, has "affinities ...to systems of prior restraint".[79]

Moreover, David Kaye, the former UN Special Rapporteur on freedom of opinion and expression, has reiterated that stakeholders should stop pushing for proactive moderation or content filtration as both are inconsistent with the privacy and right to freedom of expression and amount to prior censorship.[80] Also, AI-assisted pre-moderation processes of all published content to detect harmful content can have a severe chilling effect on freedom of expression and violate the principle of "no monitoring obligation for intermediaries" which is embodied in EU law and Council of Europe policy, and reiterated in several UN reports on freedom of opinion

---

[74] Ibid 40; News Media Service '9 Sneaky Ways People Bypass Auto Moderation' (*News Media Service*, 2020) <https://newmediaservices.com.au/9-ways-to-bypass-auto-moderation/> accessed 30 December 2023.

[75] Jillian C. York and Corynne McSherry, 'Automated Moderation Must be Temporary, Transparent and Easily Appealable' (*Electronic Frontier Foundation*, 2020) <https://www.eff.org/deeplinks/2020/04/automated-moderation-must-be-temporary-transparent-and-easily-appealable> accessed 30 December 2023.

[76] Emma J Llansó, 'No Amount Of "AI" In Content Moderation Will Solve Filtering's Prior-Restraint Problem' (2020) 7 Big Data and Society 3.

[77] Ibid.

[78] Amélie Pia Heldt, 'Upload-Filters: Bypassing Classical Concepts of Censorship?' (2019) 10 JIPITEC 57-65.

[79] Jack M. Balkin, 'Old-School/New-School Speech Regulation' (2013) 127 Harv. L. Rev. 2309.

[80] David Kaye, 'Report of the Special Rapporteur on the Promotion and Protection of The Right to Freedom of Opinion and Expression, A/HRC/38/35' (UNHRC, 2018) <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement> accessed 30 December 2023 para 67.

and expression.[81] This makes it clear, with no room for confusion, that proactive moderation mechanisms should be avoided. This section agrees with these views that the current mass AI-driven automated filtration of content often leads to a disproportionate amount of online content deletion before being published, albeit the content might not violate any rules or laws.[82]

iii. AI-Assisted Content Moderation Lacks Transparency

Online content moderation has been, for a long time, opaque, specifically the automated decision process which remains secretive due to the surrounding intellectual property and corporate opacity claims by AI systems owners.[83] Without transparency, it is hard to understand both the dynamics of automated systems and the criteria that constitute the basis of AI-driven moderation decisions. This arguably leads to inconsistent moderation and makes it difficult to identify potential human rights abuses.[84] From a user perspective, the way automated systems function remains unknown as the current models are either unexplained or ambiguously explained. Companies usually use the black-box nature of these automated systems as a veil to hide behind.[85]

Llansó warns against what appears to be "scope creep" around automated moderation systems where companies keep trying new functionalities with no oversight or transparency.[86] For example, in the aftermath of the Christchurch terrorist attack in New Zealand, the big tech companies (Facebook, Microsoft, Twitter and YouTube) through the Global Internet Forum to Counter Terrorism announced that they will share a hash database for alleged terrorist content. Such strategies have raised concern among dozens of civil society organisations since the content of the hash database remains closed off to almost everyone, including trusted third

---

[81] Committee of Experts on Internet Intermediaries (MSI-NET), 'Draft Recommendation of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries' (Council of Europe, 2017) <https://rm.coe.int/draft-recommendation-on-internet-intermediaries-version-4/1680759e67> accessed 30 December 2023 7; Council of Europe Declaration on freedom of communication on the Internet [2003] Principle 6; Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L 178/1 Article 15.

[82] Amélie Pia Heldt, 'Upload-Filters: Bypassing Classical Concepts of Censorship?' (2019) 10 JIPITEC 63.

[83] Jenna Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data and Society 3.

[84] Cambridge Consultants on Behalf of Ofcom, 'Use of AI in Online Content Moderation' (Ofcom, 2019) <https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf> accessed 30 December 2023 41.

[85] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in The Automation of Platform Governance' (2020) 7 Big Data and Society 11.

[86] Emma Llansó, 'Platforms Want Centralized Censorship. That Should Scare You' (*Wired*, 2019) <https://www.wired.com/story/platforms-centralized-censorship/> accessed 30 December 2023.

parties and researchers. That is why experts are calling for clear standards of decisional transparency so that users and third parties can understand the role of automated systems in online moderation.[87]

iv. AI-Based Content Moderation Can Be Unfair, Biased, and Discriminatory

Recently, there have been several discussions about the unfair and discriminatory impact of AI-driven decision-making systems on minority groups. Despite focus on this issue within the fields of social welfare, distribution of outcomes, and criminal justice, there are parallels in the domain of automated online content moderation.[88] Research has shown that AI content moderation systems can have a disadvantageous impact on vulnerable groups due to bias, intentional or unintentional, and discrimination which can result from the under-representative or biased datasets used to train the AI algorithms.[89] Studies also demonstrate that introducing back-propagation – which is an algorithm used to "optimise a wide range of deep learning algorithms"[90] – during training can lead to or enhance bias.[91] One example of this is the Instagram DeepText algorithm which treated the word Mexican as an insult because the training dataset associated Mexican with the term illegal and this was eventually fed to the algorithm.[92]

It is inevitable that even the most accurate automated system will be more or less favourable to online content associated with, for example, a specific gender, race, or ethnicity and thus, privilege certain groups over others.[93] Think of Amazon's AI tool for recommending best CVs,

---

[87] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation Of Platform Governance' (2020) 7 Big Data and Society 11; Ángel Díaz, 'Global Internet Forum to Counter Terrorism Transparency Report Raises More Questions Than Answers' (*Brennan Center for Justice*, 2019) <https://www.brennancenter.org/our-work/analysis-opinion/global-internet-forum-counter-terrorism-transparency-report-raises-more> accessed 30 December 2023.

[88] Ibid (Gorwa); Solon Barocas and Andrew Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671-672.

[89] David Kaye, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/38/35' (UNHRC 2018) <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement> accessed 30 December 2023 para 56; Aylin Caliskan, Joanna J. Bryson and Arvind Narayanan, 'Semantics Derived Automatically From Language Corpora Contain Human-Like Biases' (2017) 356 Science 2 and 12.

[90] Cambridge Consultants on Behalf of Ofcom, 'Use Of AI In Online Content Moderation' (Ofcom, 2019) <https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf> accessed 30 December 2023 26.

[91] Ibid 41.

[92] Nicholas Thompson, 'Instagram's Kevin Systrom Wants to Clean Up The &#%@! Internet' (*WIRED*, 2017) <https://www.wired.com/2017/08/instagram-kevin-systrom-wants-to-clean-up-the-internet/> accessed 30 December 2023.

[93] Michael D. Ekstrand and others, 'Exploring Author Gender in Book Rating and Recommendation', *Proceedings of the 12th ACM Conference on Recommender Systems* (ACM Digital Library, Vancouver, British Columbia, Canada, 2018) 249;

which was trained on a dataset that included a disproportionately higher number of CVs from men compared to women. As a result, the algorithm became biased towards male applicants.[94] Consequently, using AI in content moderation can unfairly impact people's freedom of expression, particularly minority groups that are poorly represented in the training datasets.[95] That is why training AI - which is employed in online content moderation - should use more inclusive datasets and be rigorously audited.

v. AI-Facilitated Content Moderation May Be Politicised

Another key concern is the politicisation of automated moderation since automated tools may amplify the obscurity of the decision-making process with regard to political and economic issues such as terrorism, hate speech and copyright.[96] For years, governments have been working on policing online content by either putting pressure on social media companies or entering into public-private partnerships with them. This has allowed governments to regulate online speech through private, overbroad automated monitoring systems, using rules that would normally fail to pass legal scrutiny.[97]

For example, Facebook and Israel have agreed on a partnership to monitor and combat inciting content online. This coordination has resulted in a massive wave of content removal against Palestinians with reports claiming that Facebook has deployed an algorithm which detects and removes posts including names of Palestinian factions, such as Hamas, Jihad, Qassam, or Shahid.[98] Also, Facebook was criticised for deleting, under pressure from the Indian

---

Meike Zehlike and others, 'Fa* Ir: A Fair Top-K Ranking Algorithm', *Proceedings of the 2017 ACM Conference on Information and Knowledge Management* (ACM Digital Library, Singapore, Singapore, 2017) 1.

[94] Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women' (*Reuters*, 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> accessed 30 December 2023.

[95] Cambridge Consultants on Behalf of Ofcom, 'Use Of AI In Online Content Moderation' (Ofcom, 2019) <https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf> accessed 30 December 2023 43; Reuben Binns, 'Fairness in Machine Learning: Lessons From Political Philosophy', *Proceedings of Machine Learning Research* (Conference on Fairness, Accountability, and Transparency, New York, 2018) 1; David Kaye, 'Report of The Special Rapporteur to The General Assembly on Artificial Intelligence Technologies and Implications for the Information Environment, A/73/348' (UNHRC, 2018) <https://www.un.org/ga/search/view_doc.asp?symbol=A/73/348> accessed 30 December 2023 para 15.

[96] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data and Society 12.

[97] Milton Mueller, *Networks and States: The Global Politics of Internet Governance* (MIT Press 2010) 213.

[98] IMEMC News, 'Activists Protest Facebook Policy Against Palestinian Content' (*International Middle East Media Center*, 2019) <https://imemc.org/article/activists-protest-facebook-policy-against-palestinian-content/> accessed 30 December 2023; 7amleh - Arab Center for Social Media Advancement, 'Facebook And Palestinians: Biased Or Neutral Content Moderation Policies?' (7amleh - Arab Center for Social Media Advancement, 2018) <https://7amleh.org/wp-content/uploads/2018/10/booklet-final2-1.pdf> accessed 30 December 2023 11-12.

government, dozens of videos, photos and accounts of journalists and academics who used the platform to report on violent clashes between protestors and the police in Kashmir after the death of a Kashmiri separatist leader.[99] Moreover, several statements issued by Facebook suggest that its automated moderation focuses almost solely on terrorist content related to ISIS and Al-Qaeda, failing to adequately address proliferating far-right terrorist propaganda.[100] Such examples raise questions about both the reliability of using a political and loose term such as terrorism to police online content as well as the unrepresentative datasets used to train AI algorithms.

Taking into account that AI-powered moderation lacks transparency and takes place on a massive scale, it is arguable that these implications can adversely impact freedom of expression and other human rights and stifle the public debate through the over-blocking and removal of legitimate content. Also troublesome is the trying to detect unlawful content based on vague and slippery notions such as hate speech, extremism, or terrorism.[101]

II. International and African Human Rights Jurisprudences Can Rein In Social Media Giants' AI-Powered Content Moderation

Unlike the numerous AI governance attempts through lax ethical guidelines and business-driven national strategies, it is argued that international human rights law (IHRL) represents a timeless and agile internationally recognised framework that is legally binding at national, regional, and international levels and is proven to appropriately apply to contemporary challenges such as the

---

[99] Vidhi Doshi, 'Facebook under fire for 'censoring' Kashmir-related posts and accounts' (*The Guardian,* 2016) <https://www.theguardian.com/technology/2016/jul/19/facebook-under-fire-censoring-kashmir-posts-accounts> accessed 30 December 2023; Rama Lakshmi, 'Facebook is censoring some posts on Indian Kashmir' (*The Washington Post,* 2016) <https://www.washingtonpost.com/news/worldviews/wp/2016/07/27/facebook-is-censoring-posts-on-indian-kashmir-some-say/> accessed 30 December 2023.

[100] Ángel Díaz, 'Global Internet Forum To Counter Terrorism Transparency Report Raises More Questions Than Answers' (*Brennan Center for Justice*, 2019) <https://www.brennancenter.org/our-work/analysis-opinion/global-internet-forum-counter-terrorism-transparency-report-raises-more> accessed 30 December 2023; Counter Extremism Project, 'The Far Right On Facebook' (*CEP,* 2019) <https://www.counterextremism.com/sites/default/themes/bricktheme/pdfs/The_Far_Right_on_Facebook.pdf> accessed 30 December 2023 1-2; Salvador Rodriguez, 'Facebook says it's gotten a lot better at removing material about ISIS, al-Qaeda and similar groups' (*CNBC,* 2018) <https://www.cnbc.com/2018/11/08/facebook-details-progress-vs-isis-al-qaeda-affiliates.html> accessed 30 December 2023.

[101] Committee of experts on Internet Intermediaries (MSI-NET), 'Algorithms and Human Rights' (Council of Europe, 2018) <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> accessed 30 December 2023 21.

recognition of digital rights, which were afforded equal protection as human rights offline.[102] Given the "universal, indivisible and interdependent and interrelated"[103] nature of human rights and their enshrinement in a series of international treaties (known as the International Bill of Human Rights), IHRL can serve as the forefront legal foundation for a binding and actionable AI regulatory framework.[104]

In line with the ICCPR, the 1981 Banjul Charter, which is the primary source of African human rights jurisprudence, recognises freedom of expression as a fundamental right that represents the bedrock for other rights and the cornerstone of democracy. Article 9 of the Charter stipulates that: "1. Every individual shall have the right to receive information. 2. Every individual shall have the right to express and disseminate his opinions within the law."[105] The African human rights jurisprudence also recognises, in line with international human rights law, that human rights, including freedom of expression, enjoy equal protection both offline and online.

Back in 2002, the ACHPR issued the "Declaration of Principles on Freedom of Expression in Africa", which noted that the free flow of information and ideas, availability of heterogeneous information, diversity of communication channels, and media plurality are paramount for the enjoyment of the right to freedom of expression.[106] To achieve this, the Declaration emphasised that the media with a capacity to reach a wide audience plays a key role in "promoting the free flow of information and ideas, in assisting people to make informed decisions and in facilitating and strengthening democracy".[107] It further stressed that new forms of communication technologies will contribute to the realisation of the right to freedom of expression.[108]

---

[102] Kate Saslow and Philippe Lorenz, 'Artificial Intelligence Needs Human Rights: How The Focus On Ethical AI Fails To Address Privacy, Discrimination And Other Concerns' (SNV Think Tank, 2019) 15; United Nations Human Rights Council, 'The promotion, protection and enjoyment of human rights on the Internet, Resolution 20/8' (HRC, 2012).

[103] Amnesty International and Access Now, 'The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems' (Access Now, 2018) <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/> accessed 30 December 2023.

[104] Aparajitha Narayanan, 'A Human Rights Framework Is Necessary To Govern Artificial Intelligence' (*Human Rights Pulse*, 2020) <https://www.humanrightspulse.com/mastercontentblog/a-human-rights-framework-is-necessary-to-govern-artificial-intelligence> accessed 30 December 2023; ARTICLE 19, 'Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence' (Article 19, 2019) 16-17.

[105] African Charter on Human and Peoples' Rights [Banjul Charter] (Nairobi, Kenya, 27 June 1981), 21 I.L.M. 59 (1981), entered into force 21 Oct. 1986.

[106] Declaration of Principles on Freedom of Expression in Africa (Banjul, 23 Oct. 2002) ACHPR (2002).

[107] Ibid.

[108] Ibid.

A closer look at the African jurisprudence shows that Africans were well aware of the importance of having free, independent, and pluralistic press and media as a tool for the promotion of freedom of expression, "the development and maintenance of democracy in a nation, and for economic development".[109] Therefore, African journalists issued the Windhoek Declaration - which was endorsed by the United Nations Educational, Scientific and Cultural Organisation (UNESCO) – restating the aforementioned, setting out definitions for what independent and pluralistic press represent, and recognising censorship as an intolerable human rights violation.[110]

And with the rise of the Internet and social media platforms, people found an ample space to exercise their right to freedom of expression and receive information, amongst other rights, away from the traditional defined borders. Social media platforms have also played a significant role in changing the news environment by increasingly becoming the primary source of news, particularly among younger generations.[111] This is well evidenced in Africa, for example, where mobile Internet penetration and Internet access have been steadily growing, which resulted in Africans spending a vast amount of time on social media platforms.[112] This tangible shift from traditional media to digital media, specifically social media platforms, does not negate that the rules set forth for independent and pluralistic media would further apply to social media platforms and their AI-powered content moderation practices.

For instance, employing AI-assisted content moderation by social media platforms is unlikely to align with the abovementioned "independence" and "pluralism" criteria. This is simply because they function in a manner that throttles the free and diverse flow of information for profit purposes coupled with their susceptibility to political influence. Accordingly, the content posted on these platforms undergo a ranking and curation process – a form of control over materials dissemination – based on which some content receives more exposure than other content. This can also result in certain content getting demoted if it does not align with social media platforms' monetisation policies or if it causes political sensitivity. With such practices in place,

---

[109] Declaration of Windhoek (Windhoek, 3 May 1991) UNESCO (1991).

[110] Ibid. The Windhoek Declaration defines independent and pluralistic press as: *"... 2. By an independent press, we mean a press independent from governmental, political or economic control or from control of materials and infrastructure essential for the production and dissemination of newspapers, magazines and periodicals. 3. By a pluralistic press, we mean the end of monopolies of any kind and the existence of the greatest possible number of newspapers, magazines and periodicals reflecting the widest possible range of opinion within the community...".*

[111] Nic Newman and Others, 'Digital News Report' (Reuters Institute for the Study of Journalism 2021).

[112] Aaron Olaniyi Salau, 'Social media and the prohibition of 'false news': can the free speech jurisprudence of the African Commission on Human and Peoples' Rights provide a litmus test?' (2020) 4 African Human Rights Yearbook 231-254, 233.

it is hard to consider social media platforms as vehicles that foster "the widest possible range of opinion within the community" as indicated in the Windhoek Declaration.[113]

As a response to the emerging digital sphere as the new reality, governments attempted to provide regulatory approaches as a response. Consequently, and in order to ensure that any regulatory approach, public or private, will align with human rights standards, the ACHPR Special Rapporteur on Freedom of Expression and Access to Information, Faith Pansy Tlakula, signed a joint declaration on the protection of freedom of expression on the Internet in 2011, which involved the Special Rapporteurs for Freedom of Expression of the Americas, Europe, Africa, and the United Nations.[114] The Joint Declaration emphasised that information management practices employed by intermediaries must be non-discriminatory, publicly accessible, and transparent to maintain the Internet as a space that enhances independence, diversity, and pluralism, which are core elements of the right to freedom of expression.[115] The Joint Declaration also stated that "[c]ontent filtering systems which are imposed by … commercial service providers and which are not end-user controlled are a form of prior censorship and are not justifiable as a restriction on freedom of expression".[116] This was followed by the issuance of Resolution number 362 by the ACHPR, guided by the steps of the UN Human Rights Council, affirming that "the same rights that people have offline must be protected online".[117]

In another attempt to address freedom of expression in the digital realm, a consortium of human rights and digital rights organisations agreed in 2014 on the African Declaration on Internet Rights and Freedoms.[118] The 2014 Declaration drew upon the previous human rights instruments and soft law, adding that the Internet must remain "a vehicle for free, open, equal and non-discriminatory exchange of information, communication and culture"[119] and that there must not be special treatment for content based on economic, social, cultural or political grounds.[120] The Declaration additionally outlined that "blocking, filtering, removal and other technical or legal limits on access to content … is an extreme measure – analogous to banning a

[113] Declaration of Windhoek (Windhoek, 3 May 1991) UNESCO (1991).
[114] Joint Declaration on Freedom of Expression and the Internet (Washington DC, 1 Jun. 2011) OAS (2011).
[115] Ibid.
[116] Ibid.
[117] Resolution on the Right to Freedom of Information and Expression on the Internet in Africa (Banjul, 4 Nov. 2016) ACHPR/Res. 362(LIX) (2016).
[118] African Declaration on Internet Rights and Freedoms (Istanbul, 4 Sep. 2014) IGF (2014).
[119] Ibid.
[120] Ibid.

newspaper or broadcaster – which can only be justified in accordance with international standards…".[121]

This was further consolidated by the adoption of the 2019 Declaration of Principles on Freedom of Expression and Access to Information in Africa, which built on the work of the 2002 Declaration by recognising the right to access information and tackling the interplay between the right to freedom of expression and the Internet.[122] In the 2019 Declaration, the ACHPR noted that the Internet and emerging technologies are "central to the enjoyment of [the right to freedom of expression and] other rights and essential to bridging the digital divide".[123] The Declaration also adopted a number of principles, such as Principle 39, which discusses in detail the role of Internet intermediaries and social media platforms in the realisation of the right to freedom of expression. In this regard, it aligns with international standards by calling for equal Internet access without discrimination, discouraging proactive content monitoring, and mandating the integration of human rights safeguards in content moderation processes, including transparency and effective remedies for rights violations.[124]

More importantly, the ACHPR recognised, in the 2019 Declaration, that deploying artificial intelligence by Internet intermediaries and social media platforms could entail serious repercussions on freedom of expression, and therefore stressed that "States shall ensure that the development, use and application of artificial intelligence, algorithms and other similar technologies by Internet intermediaries are compatible with international human rights law and standards, and do not infringe on the rights to freedom of expression, access to information and other human rights."[125]

---

[121] Ibid.

[122] Declaration of Principles on Freedom of Expression and Access to Information in Africa (Banjul, 10 Nov. 2019) ACHPR (2019).

[123] Ibid.

[124] Ibid Principle 39 stipulates that *"1. States shall require that Internet intermediaries enable access to all Internet traffic equally without discrimination on the basis of the type or origin of content or the means used to transmit content, and that Internet intermediaries shall not interfere with the free flow of information by blocking or giving preference to particular Internet traffic. 2. States shall not require Internet intermediaries to proactively monitor content which they have not authored or otherwise modified. 3. States shall require Internet intermediaries to ensure that in moderating or filtering online content, they mainstream human rights safeguards into their processes, adopt mitigation strategies to address all restrictions on freedom of expression and access to information online, ensure transparency on all requests for removal of content, incorporate appeal mechanisms, and offer effective remedies where rights violations occur.".*

[125] Ibid.

In 2021, the ACHPR reiterated, in Resolution number 473, the serious challenges artificial intelligence and similar technologies employed by tech giants pose to human rights.[126] More specifically, the Commission noted that AI-powered algorithmic moderation systems can be discriminatory, and adversely impact the enjoyment of freedom of expression and other human rights, for instance by assisting in the proliferation of "disinformation, … and materials for radicalization and incitement of violence,".[127] The Resolution also underscored that the development of new technologies must be human rights-centred and be free from biased narratives that would influence the automated decision-making process, especially since "computerised algorithms … may lack reflective capacity or effective independent oversight".[128]

Now, delving into court decisions in Africa shows that African courts recognised the primacy of international human rights law and interacted with legal precedents established by the United Nations Human Rights Committee,[129] the European Court of Human Rights,[130] and the Inter-American Court of Human Rights.[131] In the case of *Article 19 v. Eritrea*[132], for example, the ACHPR, setting as a quasi-judicial body, recognised that international and African human rights standards must be afforded a higher hierarchy than domestic legal frameworks because "allowing national laws to restrict the right to freedom of expression without setting boundaries would render the right an illusion".[133] By doing so, African courts managed to devise an African jurisprudence on freedom of expression drawn from the Banjul Charter, ACHPR's soft law instruments, and international law.

On numerous occasions, the ACHPR ruled that freedom of expression holds fundamental significance in the development of societies, promoting democracy, and realising all human

---

[126] Resolution on the need to undertake a Study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa (Virtual, 10 Mar. 2021) ACHPR/Res. 473 (EXT.OS/ XXXI) (2021).
[127] Ibid.
[128] Ibid.
[129] The African Court in Lohé Issa Konaté v. Burkina Faso (Application 004/2013) [2014] AfCHPR (5 December 2014), relied on the UNHRC decision in Keun-Tae Kim v. The Republic of Korea [UNHRC] Comm No 574/1994 (1998) to define the term "prescribed by law".
[130] In Umuhoza v. Rwanda (Application No. 003/2014) [2018] AfCHPR 73 (7 December 2018), the African Court cited the ECtHR case Handyside v. United Kingdom App no 5493/72 (ECtHR, 7 December 1976), to underline that the right to freedom of expression extends to speech that intends to offend, shock, or disturb.
[131] The ACHPR in Agnes Uwimana-Nkusi v. Rwanda (Communication 426/12) [2021] ACHPR 526 (16 April 2021), referred to the IACtHR case of Herrera Ulloa v. Costa Rica I/A Court H.R., Merits. Judgement of July 2, 2004. Series C No 197 to emphasise that political speech must be accorded a higher degree of tolerance.
[132] ACHPR, Comm No 275/03 (2007).
[133] Jennifer Veloz, 'Special Collection on the Case Law on Freedom of Expression: African System of Human and Peoples' Rights' (Columbia Global Freedom of Expression, 2022) <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2022/11/Special-Collection-on-the-Case-Law-on-Freedom-of-Expression-African-System-of-Human-and-Peoples%C2%B4-Rights.pdf> accessed 30 December 2023 9.

rights, and that any limitation on freedom of expression has to be narrowly couched and comply with the rules of legality, legitimacy, and proportionality.

For instance, in *Law Offices of Ghazi Suleiman v. Sudan*[134], the ACHPR stated that "by denying the Applicant's right to express his opinion on the human rights issues in Sudan, the Sudanese community was also prevented from accessing valuable information concerning their human prerogatives, resulting in a violation of Article 9 of the Charter".[135] In the same vein, the High Court of South West Africa ruled in the *Free Press of Namibia (Pty) Ltd. v. Cabinet of the Interim Government of South West Africa*[136] case that a key element, which contributes to the significance of freedom of expression and correlated rights (e.g. freedom of press), is the ability to reach "a large number of people and rallying their support [so] that these freedoms can be utilised for the benefit of society".[137] The East African Court of Justice in *Burundian Journalists' Union v. Attorney General*[138] and *Mseto v. Attorney General*[139] further held that "a government should not determine what ideas or information should be placed in the marketplace...".[140]

While such conclusions, reached by the ACHPR in these cases, concerned incidents that took place within the offline sphere, they remain analogous to incidents taking place in the digital sphere, where measures, such as content moderation, encroach upon users' right to freely express their opinions and the wider audience's right to receive and engage with heterogeneous information. This is also coupled with the fact that these obligations do not only apply to States but also extend to businesses, including social media platforms, which will be tackled later on in detail.

On another note, the case of *Media Council of Tanzania v. Attorney General*[141] demonstrated the views of the East African Court of Justice on the issue of legality, stressing that, in order to pass the legality test, limitations on freedom of expression have to be sufficiently precise and narrowly construed. Therefore, a provision that incorporates undefined terms, such as "hate

[134] ACHPR, Comm No 228/99 (2003).
[135] Jennifer Veloz, 'Special Collection on the Case Law on Freedom of Expression: African System of Human and Peoples' Rights' (Columbia Global Freedom of Expression, 2022) <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2022/11/Special-Collection-on-the-Case-Law-on-Freedom-of-Expression-African-System-of-Human-and-Peoples%C2%B4-Rights.pdf> accessed 30 December 2023 5.
[136] [1987] 4 SA (SWA).
[137] Ibid 72.
[138] (Reference 7 of 2013) [2015] EACJ 91 (15 May 2015).
[139] (Reference 7 of 2016) [2018] EACJ 44 (21 June 2018).
[140] Ibid para 67.
[141] (Reference 2 of 2017) [2019] EACJ 2 (28 March 2019).

speech," renders the provision vague and too broad.[142] The Supreme Court of Zimbabwe also touched upon the legality issue in *Chavunduka v. Minister of Home Affairs*.[143] The Court stated that the "use of the word 'false' is wide enough to embrace a statement, rumour or report which is merely incorrect or inaccurate, as well as a blatant lie"[144], and such vagueness could instil fear into people causing a "chilling effect".[145] Moreover, in the case, *Andama v. Director of Public Prosecutions*[146], the Nairobi High Court in Kenya held that the prohibition of publishing obscene material in electronic form without defining "obscene" infringed upon the freedom of expression, since the term is overly broad, leaving unfettered discretion to subjective interpretation by decision-makers.[147]

Building on this, African courts can use the legality requirement to address the serious, entrenched drawbacks of content moderation policies put in place by social media platforms, which have been highlighted earlier in this section. These content moderation policies have been noted on several occasions in the literature as being "largely opaque and often deployed by platforms in self-serving ways that can conceal the harmful effects of their policies and practices ...",[148] providing social media platforms with colossal discretion on what would be violating content and how to identify and sanction it.[149]

This was also asserted by the Meta's Oversight Board (OB) in its decision on *Claimed COVID-19 cure*[150], where the OB criticised the scattered status of Facebook's COVID-19 policies, which were added as patches to different pages of Facebook's website, making it challenging for users to find out on what basis content is restricted. The OB further stressed that "Facebook's misinformation and imminent harm rule ... to be inappropriately vague and inconsistent with international human rights standards."[151]

---

[142] Ibid para 66.
[143] [2000] JOL 6540 (ZS).
[144] Ibid 15.
[145] Ibid.
[146] Petition No. 214 of 2018 (2019) eKLR.
[147] Ibid para 64.
[148] Ángel Díaz and Laura Hecht-Felella, 'Double Standards in Social Media Content Moderation' (Brennan Center for Justice 2021) <https://www.brennancenter.org/sites/default/files/2021-08/Double_Standards_Content_Moderation.pdf> accessed 30 December 2023 3.
[149] Ibid 10.
[150] 2020-006-FB-FBR >https://www.oversightboard.com/decision/FB-XWJQBU9A/ <accessed 30 December 2023.
[151] Ibid.

On the issue of "prior restraints", the Constitutional Court of South Africa also emphasised in *Print Media South Africa v. Minister of Home Affairs*[152] that prior classification of publications by the government "entails a transfer of control from the right-bearer [the publisher], seeking to exercise the right to freedom of expression, to an administrative body"[153], which stifles the publisher's liberty to exercise freely his/her right to freedom of expression.[154] The Court ruled accordingly that the prior classification of publications "amounts to a form of prior restraint, which is an inhibition on expression before it is disseminated,"[155] and thus represents an unjustifiable interference with the right to freedom of expression.[156]

Likewise, through African jurisprudence on prior restraints, courts could actively address the challenges posed by AI-powered content moderation where social media platforms pre-screen user-generated content before publication to ensure compliance with their policies. The same can be said for AI-driven content ranking and recommendation. These tools could be seen as a form of prior restraint, as social media platforms utilise them to pre-determine which content should be published, prioritised, or deprioritised, ultimately shaping the content that users consume.

Equally important comes the issue of false news, which was tackled by several African courts. The Constitutional Court of South Africa acknowledged, in the case of *Democratic Alliance v. African National Congress*[157], that "the publication of false statements with the intention of influencing the outcome of elections infringes the right to free and fair elections, and thus, not protected under the right to freedom of expression."[158] Nevertheless, the Court emphasised that this must not apply to an interpretation of content, honest opinions, or the expression of comments, which remain protected under freedom of expression.[159]

Concurrently, the High Court of Zambia in *Chipenzi v. The People*[160] and the Supreme Court of Zimbabwe in *Chavunduka v. Minister of Home Affairs*[161] concluded that "statements, opinions and beliefs regarded by the majority as being false and wrong plainly fell within the protections

[152] CCT 113/11 [2012] ZACC 22.
[153] Ibid para 58.
[154] Ibid.
[155] Ibid para 16.
[156] Ibid.
[157] CCT 76/14 [2015] ZACC.
[158] Ibid para 50.
[159] Ibid para 84.
[160] (2014) HPR/03/2014 [HC].
[161] [2000] JOL 6540 (ZS).

afforded by the right to freedom of expression".[162] Hence, a restriction on statements that do not necessarily come under the definition of truth set by the majority infringes upon the right to freedom of expression as it goes beyond false statements to honest opinions.[163] The ECOWAS simultaneously stressed in *Federation of African Journalists (FAJ) and others v. The Gambia*[164] that the "erroneous statement is inevitable in free debate, and that it must be protected if the freedoms of expression are to have the 'breathing space' that they need … to survive".[165]

In this respect, these legal precedents could capture the serious distortion of the digital realm resulting from the employment of AI by social media platforms in moderating user-generated content. On the one hand, due to the profit-oriented nature of social media platforms, the AI-driven systems employed for content curation contribute to the wide dissemination of false information and the proliferation of malicious AI bots associated with high engagement rates. This consequently gives rise to political polarisation, extremism, and electoral manipulation. On the other hand, social media platforms attempt to tackle the issue of false news by once again implementing AI-powered moderation systems. However, these systems often apply loose and ambiguous rules, causing inconsistent application and encroachment on protected speech and honest opinion.

As for court decisions in Africa that dealt directly with freedom of expression in the digital realm, the Community Court of Justice of the Economic Community of West African States (ECOWAS) entertained this issue on two occasions. The first time was in 2020 in the case of *Amnesty International Togo and Ors v. The Togolese Republic.*[166] In this case, the Court opined that although "[a]ccess to Internet is not *stricto senso* a fundamental human right but since Internet service provides a platform to enhance the exercise of freedom of expression, it then becomes a derivative right that it is a component to the exercise of the right to freedom of expression",[167] hence should be jointly treated as an element of human right that requires protection of the law.

The second was in 2022 when the ECOWAS held in *SERAP v. Federal Republic of Nigeria*[168] that the right to freedom of expression could be exercised through whatever medium of choice

[162] Ibid 8; (2014) HPR/03/2014 [HC] 5.
[163] Ibid.
[164] (ECW/CCJ/APP/ 36 of 2015) [2018] ECOWASCJ 4 (13 February 2018).
[165] Ibid 43.
[166] (ECW/CCJ/APP/ 61 of 2018) [2020] ECOWASCJ 9 (25 June 2020).
[167] Ibid para 38.
[168] (ECW/CCJ/JUD/40/22) [2022] ECOWASCJ 7 (19 July 2022).

and any media channel. The Court acknowledged that freedom of expression is not absolute but noted that any limitation must have a legal justification stipulated by the law.[169] The ECOWAS further underlined the pivotal role social media platforms play in enabling "the exchanges of ideas, views and opinions," and how they are "of much relevance in the attainment of the intended objectives of Articles 9 of the ACHPR and 19 of the ICCPR and in like manner, relevant in the enjoyment of the exercise of the right to freedom of expression".[170] The Court therefore recognised that, similar to print media, electronic media (including social media platforms) are considered media channels through which people can freely express their opinions and receive, disseminate, and impart information.[171] The Court, particularly, highlighted that access to social media platforms to "is one such derivative right that is complementary to the enjoyment of the right to freedom of expression" enshrined in the Banjul Charter and the ICCPR.[172]

In short, despite the modest attempts by African courts to address the issue of freedom of expression in the digital realm, African human rights jurisprudence can still effectively entertain the human rights implications stemming from automated content moderation, content personalisation, and microtargeting.

Moving to the applicability of these rules, it would be erroneous to assume that what African human rights jurisprudence offers here solely addresses States; It indeed extends to businesses as well, especially that Africa has been witnessing massive corporate human rights violations since the colonial era. This matter was therefore attended by the ACHPR when it rendered its decision in *SERAC and CESR v Nigeria*, followed by another landmark decision in *IHRDA and Others v DRC*, where the ACHPR asserted that States have a positive obligation to "investigate, prosecute and redress corporate human right abuses".[173]

The 2019 Declaration on Freedom of Expression and Access to Information stipulated that States are obliged to afford adequate protection to those exercising their right to freedom of expression "against acts or omissions of non-State actors that curtail the enjoyment of freedom of expression".[174] On top that, the ACHPR utilised the niche perspective on human rights

[169] Ibid 22.
[170] Ibid 24.
[171] Ibid 23.
[172] Ibid.
[173] Wubeshet Tiruneh, 'Holding corporations liable for human rights abuses committed in Africa: the need for strengthening domestic remedies' (2022) 6 African Human Rights Yearbook 227-246, 227, 231.
[174] Declaration of Principles on Freedom of Expression and Access to Information in Africa (Banjul, 10 Nov. 2019) ACHPR (2019).

obligations of the Banjul Charter – which, unlike other human rights instruments, imposes human rights obligations over individuals as well – and held that since these obligations apply to individuals, "there is an even stronger moral and legal basis for attributing these obligations to corporations and companies".[175]

Furthermore, a 2023 resolution issued by the ACHPR addressing business and human rights in Africa recognises that "[t]he activities of big businesses often have adverse impacts on the rights of peoples and communities"[176] and clearly stating that "[r]espect for human rights norms and principles by business enterprises in the countries of operation is a prerequisite for the sustainable development envisaged in AU's Agenda 2063".[177]

In a few words, the African human rights system has a promising body of hard and soft law instruments accompanied by well-established case law to address the challenges arising from AI deployment by social media platforms, but this still needs to be consolidated through African courts, which should actively seek to address and decide upon these unorthodox human rights issues emanating from the current digital era we live in.

III. Recommendations

- Enforce a "traceability and evidence requirement" to ensure AI developers prove compliance with human rights through documentation and audits.
- Mandate human rights impact assessments at all AI life cycle phases, including conception, design, and development, to identify and address potential issues.
- Establish independent oversight bodies with legal and technical expertise to monitor AI impact assessments, address risks, and ensure compliance with human rights obligations.
- Facilitate access to effective remedy channels, including independent redressal bodies, to address human rights violations resulting from AI content moderation practices.
- Enrich African human rights jurisprudence by empowering African courts to actively engage with freedom of expression issues arising from AI in content moderation.
- Advocate for balanced AI regulations to avoid overly restrictive or permissive rules, learning from examples such as the German hate speech law 'NetzDG'.

[175] Advisory note to the African group in Geneva on the legally binding instrument to regulate in international human rights law, the activities of transnational corporations and other business enterprises (legally binding instrument) (Nov. 2019) ACHPR/WGEI (2019).
[176] Resolution on Business and Human Rights in Africa (Virtual, 7 Mar. 2023) ACHPR/Res.550 (LXXIV) (2023).
[177] Ibid.

- Address the imbalance in holding businesses accountable for human rights violations by enhancing the capacity of African States to investigate, prosecute, and provide redressal channels.
- Ensure transparency in AI decision-making, disclose automated decisions, and establish human review systems for AI-related remedy requests.
- Social media platforms should be transparent with their users by explicitly notifying them whenever AI is used on their platforms, by informing them of decisions that are solely made by or involve automated systems and how these decisions are made and by making data and statistics on automated content moderation practices publicly accessible.
- Social media platforms should stop proactive automated moderation or at least use it in extremely limited circumstances. It is recommended that proactive automated moderation should only alert human reviewers to potentially harmful content.

Section 3: AI Content Moderation and Health

Artificially intelligent content moderation tools have been developed, as demonstrated in the sections preceding this one. This section will focus on drawing the link between how SMPs operate with AI algorithms, and how such operation has a negative impact on the users' health.[178]

When addressing AI content moderation technologies, it is imperative to note that there are several technologies that are deployed by SMPs, however for the sake of clarity, this section will only be addressing "filter bubbles". A filter bubble is an expression that is used to describe how a user is placed inside a bubble with content that the user finds stimulating – in other words, the algorithms show the user tailored content to users' interests. This is often followed by creating an "echo chamber", where the user clicks on the content that they find most interesting, and therefore resulting in the prolonged usage of the SMP.[179]

---

[178] Kalpathy Ramaiyer Subramanian, 'Product Promotion in an Era of Shrinking Attention Span' (2017) 7 International Journal of Engineering and Management Research 85 and 80.

[179] Seth Flaxman, Sharad Goel, and Justin M. Rao , "Filter Bubbles, Echo Chambers, and Online News Consumption" (2016) 80 Public Opinion Quarterly 298; Ana Sofia Cardenal and others, "Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape News Media Exposure" (2019) 24 The International Journal of Press/Politics 465; Elizabeth Dubois and Grant Blank, "The Echo Chamber Is Overstated: the Moderating Effect of Political Interest and Diverse Media" (2018) 21 Information, Communication and Society 729; Aniko Hannak and others, "Measuring Personalization of Web Search" [2013] Proceedings of the 22nd international conference on World Wide Web - WWW 13.

Such AI generated filter bubbles are allegedly responsible for negatively impacting users' attention span due to its impact on the dopamine loop of the users' brain.[180] This can be the result of internet addiction and can result in neurological implications -even without addiction- such as attention deficit hyperactivity disorder ("ADHD") as well as other neurologically related issues.[181]

After drawing the link between AI content moderation and the users' health, there will be specific focus on the framework that could be followed from an IHRL as well as from an African human right law perspective. This will mostly follow a comparative approach and will be followed by several recommendations with regards to what can be done about this negative impact on users' right to health.

I. The Unregulated Problem: Health and AI Content Moderation

This section will mainly describe the problem and focus on drawing the link between how SMPs and AI generated filter bubbles can impact the users' health. AI generated filter bubbles have demonstrated that they have a negative impact on the neurological health of SMPs users.[182] This claim has recently been acknowledged by several US States, as there are over 41 States that are suing meta for its algorithmic harms on mental health.[183]

A good starting point for addressing attention span is internet addiction, which is an addiction that is comparable to alcohol addiction, as it happens that the criteria for determining whether a person has internet addiction or not, is based on similar criteria, being the criteria for determining alcohol addiction.[184]

---

[180] Hüseyin Bilal MACİT, Gamze MACİT, and Orhan GÜNGÖR, 'A Research on Social Media Addiction and Dopamine Driven Feedback' (2018) 5 Journal of Mehmet Akif Ersoy University Economics and Administrative Sciences Faculty 883 and 890.

[181] Jing Wu and others, 'Role of Dopamine Receptors in ADHD: A Systematic Meta-Analysis' (2012) 45 Molecular Neurobiology 606.

[182] Hüseyin Bilal MACİT, Gamze MACİT, and Orhan GÜNGÖR, 'A Research on Social Media Addiction and Dopamine Driven Feedback' (2018) 5 Journal of Mehmet Akif Ersoy University Economics and Administrative Sciences Faculty 883 and 890; Jing Wu and others, 'Role of Dopamine Receptors in ADHD: A Systematic Meta-Analysis' (2012) 45 Molecular Neurobiology 606; Kalpathy Ramaiyer Subramanian, 'Product Promotion in an Era of Shrinking Attention Span' (2017) 7 International Journal of Engineering and Management Research 85 and 89.

[183] Cristiano Lima and Naomi Nix, '41 States Sue Meta, Claiming Instagram, Facebook Are Addictive, Harm Kids' (The Washington Post, 24 October 2023) <https://www.washingtonpost.com/technology/2023/10/24/meta-lawsuit-facebook-instagram-children-mental-health/> accessed 26 October 2023.

[184] Hüseyin Bilal MACİT, Gamze MACİT, and Orhan GÜNGÖR, 'A Research on Social Media Addiction and Dopamine Driven Feedback' (2018) 5 Journal of Mehmet Akif Ersoy University Economics and Administrative Sciences Faculty 883 and 890.

Internet addiction happens as a result of the excessive release of dopamine in the users' brain. Dopamine is the neurochemical that is ultimately responsible and necessary for several functions that the brain carries out. Such functions extend to "thinking, carrying, sleeping, mood, attention, motivation, seeking and rewarding". Additionally, it is responsible for feeling pleasure, as it is known as the hormone of happiness, and when it comes to SMPs, "Social media dopamine loop is explained which is a similar version of drug addicted dopamine loop." Therefore, it can be understood that the way SMPs are designed is evidentially addictive and the intensity of using SMPs can have negative effects on the users' mental health.[185]

AI generated filter bubbles and echo chambers, as displayed above, are designed in a manner that keeps the user engaged by tailoring content to their interests.[186] Every time the user clicks on relevant content, dopamine gets released in the user's brain, thus they get addicted to the act that releases the dopamine, which is surfing the internet, thus resulting in the term internet addiction.[187]

The criteria for internet addiction is five pronged, whereas only fulfilling three parts of the five can lead to diagnosis of addiction. The five prongs are 1) Emotional change; 2) Draw attention; 3) Deprivation; 4) Tolerance; and 5) Recurrence.[188]

Emotional change can be summed up in the change of emotions that occur when an individual receives an addictive substance. The drawing attention part occurs when the substance draws the individual's attention to the point where it becomes the most important in their mind and life.[189] Both prongs can easily apply to SMPs, whereas the dopamine released when interacting with SMPs results in the emotional change, and at the same time, if an individual is at the point where social media is the most important component in their mind and life, then their attention has been drawn to SMPs, and it will be difficult to stop controlling going online.

---

[185] Ibid 883-894.

[186] Seth Flaxman, Sharad Goel, and Justin M. Rao , "Filter Bubbles, Echo Chambers, and Online News Consumption" (2016) 80 Public Opinion Quarterly 298; Ana Sofia Cardenal and others, "Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape News Media Exposure" (2019) 24 The International Journal of Press/Politics 465; Elizabeth Dubois and Grant Blank, "The Echo Chamber Is Overstated: the Moderating Effect of Political Interest and Diverse Media" (2018) 21 Information, Communication and Society 729; Aniko Hannak and others, "Measuring Personalization of Web Search" [2013] Proceedings of the 22nd international conference on World Wide Web - WWW 13.

[187] Hüseyin Bilal MACİT, Gamze MACİT, and Orhan GÜNGÖR, 'A Research on Social Media Addiction and Dopamine Driven Feedback' (2018) 5 Journal of Mehmet Akif Ersoy University Economics and Administrative Sciences Faculty 883-890.

[188] Ibid 890.

[189] Ibid.

The prong of deprivation is relevant to drawing attention, as when a person is deprived from the addictive substance, it results in the "unpleasant sensation". Further, tolerance occurs when the individual's tolerance for staying away from the substance decreases, whereas they cannot continue their day without it for example.[190] Both prongs are relevant to the previously discussed prongs, and can be apparent if the SMP user keeps increasing their online time, and get unpleasant sensations if their access is cut off to SMPs, which can include sensations of restlessness, trembling, nervousness, etc.…

Finally, recurrence occurs, when the individual quits using the addictive substance and then resumes using it while having the same feeling towards it as before quitting it.[191] This can be demonstrated through an SMP user that decided to quit SMPs then resumes using them with the same intensity as before quitting with the same strong release of dopamine.

Thus, the main problem with AI content moderation from a health perspective, is its addictive function since it releases the hormone of happiness "the individual is conditioned to perform the activity more often in order to obtain a similar satisfaction with his first experience". It is also crucial to note that SMPs are designed as a reward-based system, the more reactions a user gets from other users, the more reactions the user will seek. Therefore, it is the ideal environment for dopamine release in the brain.[192]

Such constant release can result in the "dysfunction of the dopaminergic system in the brain has been implicated in a lot of neuropsychological diseases, such as (…), ADHD, addiction, and schizophrenia".[193] Thus, it is clear that it will impact the attention of the user if it results in ADHD for example.[194]

Furthermore, there has been a specific study concerning SMPs usage and its relation to ADHD on adolescents aged between 11 and 15. The conclusion reached was that when adolescents develop SMP "problems" it is more likely that they will develop ADHD symptoms.[195] "Problems" were defined as "addiction-like behaviours, such as the displacement of other activities for Social Media usage, or having conflicts with others due to their Social Media

---

[190] Ibid.
[191] Ibid.
[192] Ibid 891-894.
[193] Jing Wu and others, 'Role of Dopamine Receptors in ADHD: A Systematic Meta-Analysis' (2012) 45 Molecular Neurobiology 606.
[194] See e.g., Lydia Furman, 'What Is Attention-Deficit Hyperactivity Disorder (ADHD)?' (2005) 20 Journal of Child Neurology.
[195] Ibid.

usage[196] (…) such as constant urge to go online or the inability to control Social Media Usage", such definition was drawn from the substance dependency criteria.[197]

SMPs using reels or "Stories" such as Instagram or other SMPs can contribute further to ADHD symptoms, as there is more intensity to SMPs usage when the content remains online for less time. Thus, resulting in more usage time, and putting the user at more risk of having SMPs problems, and increasing their ADHD symptoms.[198]

II. The Solution to Regulating AI Content Moderation from a Right to Health Perspective

i. Ethics

AI regulation has been a problematic area when it comes to any right. A few principles were developed in order to ensure compliance in the health sector[199] and with human rights, more specifically there have been two principles that seemed to repeat themselves amongst all other principles, those being 1) Explainability; and 2) Transparency. In sum, they require that AI systems, when deciding, should be transparent and explainable to the concerned stakeholders, thus ensuring compliance to human rights.[200]

---

[196] Maartje Boer and others, 'Attention Deficit Hyperactivity Disorder-symptoms, Social Media Use Intensity, and Social Media Use Problems in Adolescents: Investigating Directionality' (2019) 91 Child Development e853; Regina J. J. M van den Eijnden, Jeroen S. Lemmens, and Patti M. Valkenburg, 'The Social Media Disorder Scale' (2016) 61 Computers in Human Behavior; Mark D. Griffiths, Daria J. Kuss and Zsolt Demetrovics, 'Chapter 6 - Social Networking Addiction: An Overview of Preliminary Findings' in Daria J Kuss (ed), Behavioral Addictions: Criteria, Evidence, and Treatment (Academic Press 2014).

[197] Maartje Boer and others, 'Attention Deficit Hyperactivity Disorder-symptoms, Social Media Use Intensity, and Social Media Use Problems in Adolescents: Investigating Directionality' (2019) 91 Child Development e862

[198] Ibid e862-e863.

[199] See, eg, International Medical Device Regulators Forum, 'Software as a Medical Device (SaMD): Clinical Evaluation' (2016), <https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf> accessed 10 September 2023; WHO, 'Ethics and governance of artificial intelligence for health: WHO guidance' (2021) <https://www.who.int/publications/i/item/9789240029200> accessed 10 September 2023; Tatiana de Campos Aranovich and Rita Matulionyte, 'Ensuring AI Explainability in Healthcare: Problems and Possible Policy Solutions' (2022) 32 Information Communications Technology Law 2.

[200] Tatiana de Campos Aranovich and Rita Matulionyte, 'Ensuring AI Explainability in Healthcare: Problems and Possible Policy Solutions' (2022) 32 Information Communications Technology Law 2; See also R Matulionyte, 'Reconciling Trade Secrets and AI Explainability: Face Recognition Technologies as a Case Study' (2022) 44(1) European Intellectual Property Review 3; See e.g., Maja BRKAN and Grégory BONNET, 'Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas' (2020) 11 European Journal of Risk Regulation 18; Alyssa M. Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms' (2017) 103(303) Iowa Law Review 303; W. Nicholson Price II and Arti K. Rai, 'Clearing Opacity through Machine Learning' (2021) 106 (2) Iowa Law Review 775.

Such principles imply SMPs' responsibility for not harming their users' health, thus respecting the users' right to health. For example, YouTube adopted a practice, where it limits the number of hours of work for content moderators for the sake of their mental health.[201] Additionally, they notify users when using YouTube for a long period of time.[202] Though this practice was not adopted to abide with AI principles nor to preserve the users' health,[203] yet it remains effective, as it pauses the dopamine release, what would make it more favourable is to put a disclaimer about the risks of using SMPs.

ii. International and African Legal Framework

The right to health originates from Article 12 of the International Covenant on Economic, Social and Cultural Rights (ICESCR), Article 16 of the Banjul Charter, and Article 14 of the African Charter on the Rights and Welfare of the Child (ACRWC),[204] where physical and mental health are guaranteed.[205] The Articles are further elaborated on by general comment number 14 of the committee on ESCR and the general comment number 7 of the ACHR. They stipulate that the right to health has four components that should be guaranteed in every case, those being a) Availability; b) Accessibility; c) Acceptability; and d) Quality.[206]

In sum, the element of availability is satisfied when health care facilities are well equipped with necessary services and medicines. Including the standards, such drinkable water, clinics,

---

[201] Nick Statt, 'YouTube Limits Moderators to Viewing Four Hours of Disturbing Content per Day' (The Verge, 13 March 2018) <https://www.theverge.com/2018/3/13/17117554/youtube-content-moderators-limit-four-hours-sxsw> accessed 10 September 2023; see also, Sebastian Smart and Alberto Coddou McManus, 'Closing the Gap between UNGPs and Content Regulation/Moderation Practices' (2022) 19 Braz J Int'l L 271.

[202] 'How Do I Stop YouTube from Pausing a Video Every 1 Hour in Background Play? It's Kind of Annoying Asking Me to Confirm If I Still Want To...' (Quora) <https://www.quora.com/How-do-I-stop-YouTube-from-pausing-a-video-every-1-hour-in-background-play-It-s-kind-of-annoying-asking-me-to-confirm-if-I-still-want-to-watch> accessed 10 September 2023; 'Why YouTube Asks, "Continue Watching?" - Youtube Help' (Google) <https://support.google.com/youtube/answer/12819304?hl=en#:~:text=Continue%20watching%3F%E2%80%9D%20message%20will%20surface,you%20can%20select%20%E2%80%9CYes%E2%80%9D> accessed 10 September 2023.

[203] 'Why YouTube Asks, "Continue Watching?" - Youtube Help' (Google) <https://support.google.com/youtube/answer/12819304?hl=en#:~:text=Continue%20watching%3F%E2%80%9D%20message%20will%20surface,you%20can%20select%20%E2%80%9CYes%E2%80%9D> accessed 10 September 2023.

[204] It is worth noting that the African treaties refer to mental health in a more direct manner than the ICESCR.

[205] Ebenezer Durojaye 'An analysis of the contribution of the African human rights system to the understanding of the right to health' (2021) 21 African Human Rights Law Journal 441-468.

[206] UN Committee on Economic, Social and Cultural Rights (CESCR), General Comment No. 14: The Right to the Highest Attainable Standard of Health (Art. 12 of the Covenant), 11 August 2000, E/C.12/2000/4 <https://www.refworld.org/docid/4538838d0.html> accessed 10 September 2023 Para. 12; African Commission on Humans' and Peoples' Rights, General Comment No. 7 on Article 14(1)(d) and (e) of the African Charter on Human and Peoples' Rights: The Right to Participate in Government (2022) <https://achpr.au.int/sites/default/files/files/2022-10/general-comment-7-english.pdf> accessed10 September 2023 17.

etc.…[207] This element should be taken together with the second element of accessibility, as it requires the accessibility of necessary medical information to the individuals that could have their health impacted.[208] When taken together with the Ruggie principles,[209] one may argue that it is the duty of SMPs to guarantee the availability of information regarding the risks of mental and neurological health for users when they interact with addictive AI algorithms.[210] This could be done in a manner similar to cigarette packaging in an informative manner regarding the risks of AI over SMPs.[211]

Further, the element of acceptability can be summarised in having methods and facilities that are compatible and accommodating of different sexes, cultures and communities.[212] This is not essentially applicable to SMPs, yet if SMPs take further steps than the dissemination of information regarding the health risks, and provides counselling for impacted users, or even directs the users to counselling services in their vicinity. Then they should be mindful of the acceptability criterion. This would be further in conformity with the Ruggie principles since the corporation would be promoting the right to health and thus human rights.[213]

---

[207] UN Committee on Economic, Social and Cultural Rights (CESCR), General Comment No. 14: The Right to the Highest Attainable Standard of Health (Art. 12 of the Covenant), 11 August 2000, E/C.12/2000/4 <https://www.refworld.org/docid/4538838d0.html> accessed 10 September 2023 Para 12.

[208] Ibid.

[209] HRC 'Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' (2011) <https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf> accessed 10 September 2023.

[210] Communication 155/96, Social and Economic Rights Action Center (SERAC) and Center for Economic and Social Rights (CESR) v Nigeria, 27 October 2001, para 53-57; See e.g., Opal Masocha Sibanda, 'Towards a More Effective and Coordinated Response by the African Union on Children's Privacy Online in Africa', African Human Rights Yearbook, vol 6 (Pretoria University Law Press 2022) 160; HRC 'Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' (2011) <https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf> accessed 10 September 2023 3-8, 14-18 and 27-28.

[211] Nicholas J. Diamond, 'The Final Say on Australia's Plain Packaging Law at the WTO' (O'Neill, 6 August 2020) <https://oneill.law.georgetown.edu/the-final-say-on-australias-plain-packaging-law-at-the-wto/> accessed 10 September 2023.

[212] UN Committee on Economic, Social and Cultural Rights (CESCR), General Comment No. 14: The Right to the Highest Attainable Standard of Health (Art. 12 of the Covenant), 11 August 2000, E/C.12/2000/4, <https://www.refworld.org/docid/4538838d0.html> accessed 10 September 2023 Para 12.

[213] HRC 'Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' (2011) <https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf> accessed 10 September 2023 13-16.

The final element is quality, where it requires the existence of scientific and medical quality for any medically related service. [214] This would be applicable to SMPs' AI systems in the sense of having to scientifically explain how such AI generated filter bubbles impact the neurological and mental health of the users, which would be further in compliance with the aforementioned AI principles.[215] This should be further accompanied by scientific monitoring from the respective governments.[216]

In addition to such 4 principles, the obligation to prevent against interference from 3rd parties is existent when it comes to the right to health, this extends to corporations, and could occur through legislations to compel SMPs to stop their addictive functionalities[217] by becoming neutral and not creating a filter bubble, this is evident through accepting the argument that previous regulations need to be amended. [218] This should occur due to exposing the users to this harm in the first place.[219]

---

[214] UN Committee on Economic, Social and Cultural Rights (CESCR), General Comment No. 14: The Right to the Highest Attainable Standard of Health (Art. 12 of the Covenant), 11 August 2000, E/C.12/2000/4 <https://www.refworld.org/docid/4538838d0.html> accessed 10 September 2023 Para 12.

[215] Tatiana de Campos Aranovich and Rita Matulionyte, 'Ensuring AI Explainability in Healthcare: Problems and Possible Policy Solutions' (2022) 32 Information Communications Technology Law 2; See also R Matulionyte, 'Reconciling Trade Secrets and AI Explainability: Face Recognition Technologies as a Case Study' (2022) 44(1) European Intellectual Property Review 3.

[216] Communication 155/96, Social and Economic Rights Action Center (SERAC) and Center for Economic and Social Rights (CESR) v Nigeria, 27 October 2001, para 53-57.

[217] UN Committee on Economic, Social and Cultural Rights (CESCR), General Comment No. 14: The Right to the Highest Attainable Standard of Health (Art. 12 of the Covenant), 11 August 2000, E/C.12/2000/4 <https://www.refworld.org/docid/4538838d0.html> accessed 10 September 2023 Para. 33, 37, 50-52; African Commission on Humans' and Peoples' Rights, General Comment No. 7 on Article 14(1)(d) and (e) of the African Charter on Human and Peoples' Rights: The Right to Participate in Government (2022) <https://achpr.au.int/sites/default/files/files/2022-10/general-comment-7-english.pdf> accessed 10 September 2023 Para. 43; see also ACHPR Guidelines on the Right to Water in Africa (2022) <https://achpr.au.int/sites/default/files/files/2022-08/eng-achprguidelinesontherighttowaterinafrica.pdf> accessed 10 September 2023 para 32.5; African Commission on Human and Peoples' Rights 'Principles and Guidelines on the Implementation of Economic, Social and Cultural Rights in the African Charter on Human and Peoples' Rights' (2011), para 7; Communication 155/96, Social and Economic Rights Action Center (SERAC) and Center for Economic and Social Rights (CESR) v Nigeria, 27 October 2001, para 46 and 52; African Commission on Humans' and Peoples' Rights, General Comment No. 7 on Article 14(1)(d) and (e) of the African Charter on Human and Peoples' Rights: The Right to Participate in Government (2022) <https://achpr.au.int/sites/default/files/files/2022-10/general-comment-7-english.pdf> accessed 10 September 2023 Para 43.

[218] Communication No. 241/2001, Purohit and Moore v The Gambia, 29 May 2003, para 3.

[219] Communication 279/03-296/05, Sudan Human Rights Organisation and Centre on Human Rights and Evictions v Sudan, 27 May 2009, para 208-210.

Another method of compelling SMPs to stop such practices, is through States' constitutions, specifically through the horizontal application of human rights, where States compel individuals to maintain other individuals' rights, regardless of being juristic of natural persons.[220]

III. Recommendations

- States must compel SMPs to stop their addictive functions by having a neutral internet without filter bubbles. Either through legislation, international and/or African law, or even horizontal application of human rights.
- Compelling SMPs to provide information on their addictive practices, and to provide information for recourse to nearest and most suitable counselling services for its impacted users.
- The creation of guidelines for safe SMPs usage while focusing on the right to health. This should involve SMPs' expertise when being drafted.
- The academic community must further explore measures to regulate AI while focusing on the right to health.

Conclusion

From exploring the nuances of privacy rights in the African context to the complex interplay of AI with freedom of expression, and finally, examining the impacts of social media platforms (SMPs) on mental health, this paper underscores the multifaceted challenges and opportunities presented in the digital age. Here, we synthesise these findings and extend actionable recommendations.

In reference to section one concerning privacy, as earlier established, the right to privacy is an enabler for the realisation of other rights. If we say that the right to privacy is conceptually recognisable under the Banjul Charter, there would still remain a need for an official action to confer legitimacy over such a right, which will in turn encourage individuals to approach the African Court of Justice and Human Rights (ACJHR) with their privacy complaints. It is not unrealistic to assume that the hesitation of the African population to submit claims on the right

[220] Aoife Nolan 'Holding non-state actors to account for constitutional economic and social rights violations: Experiences and lessons from South Africa and Ireland' (2014) 12 International Journal of Constitutional Law, Oxford University Press 87; Danwood Mzikenge Chirwa 'The Horizontal Application of Constitutional Rights in a Comparative Perspective' (2006) 10 Journal of UWC Faculty Of Law; See e.g., Danwood Mzikenge Chirwa 'The Horizontal Application of Constitutional Rights in a Comparative Perspective' (2006) 10 Journal of UWC Faculty Of Law; 'Ghana Constitution 1996' (Constitutionproject.org) <https://www.constituteproject.org/constitution/Ghana_1996> Accessed 10 September 2023.

to privacy to the ACJHR is the lack of an explicit legal basis. Had the African population overcome the other efficiency doubts and could overcome the practical challenges of litigating before the ACJHR, uncertainty about the formal recognition and the lack of a comprehensive standard of privacy would suffice to dissuade them.

Besides the value of formal recognition, building up a jurisprudence on the right to privacy in the African realm shall largely benefit from the normative force this recognition would allow. It will serve as an anchor for judicial and scholarly effort in regard to the right to privacy. The Malabo Conventions is undoubtedly a great step forward as it lays the groundwork for protection of personal data. However, the way ahead for privacy is still long, and requires activism on part of the ACHPR, and the AU as a whole. On the flipside, the lack of a comprehensive framework carries with it the advantage of a joint vision towards privacy. Being in the phase of laying the foundation of this right and its practices offers a great opportunity to collaborate on a joint vision and collaborative effort, allowing diverse stakeholders to shape a privacy paradigm that is culturally nuanced, technologically adaptable, and universally respectful of human rights.

The African landscape is rather more mature as to the freedom of expression, yet the preparedness of the African regulatory framework for digital constraints is examinable. Essentially, the swift shift towards a data-driven society has allowed AI to immensely and rapidly proliferate in numerous social domains and influence our daily lives in various ways. While AI can be helpful through, for example, enhancing healthcare access, helping the visually impaired and improving agriculture, there are growing, legitimate concerns relating to the potential adverse human rights impact of deploying AI in other areas such as social media platforms. The paper examined one of the contentious AI-based applications employed by online platforms to shape our online experience, namely content moderation, identifying a range of human rights ramifications this tool may generate. First, AI-powered content moderation suffers from different technical limitations and could amount to a prior restraint and a monitoring mechanism. Moreover, it lacks decisional transparency, could be politicised to police legitimate content and is likely to exacerbate bias and discrimination.

Finally, concerning the right to health, after having demonstrated the manner in which SMPs operate, and the damage that they could cause to the neurological and mental health of its users, it is imperative to address the issue in a manner that lies in conformity with the recommendations.

Once more, the lack of a complete framework on one of the rights seems to hold a promise for the future African efforts. Given the opportunity to benefit from the foundation of the international framework, and further build upon its principles, the African landscape seemingly forges ahead with a slight edge. However, this does not negate the fact that there is still room for improvement, especially that both frameworks need to compel SMPs to stop using AI in an addictive manner, and to try to the furthest extent possible to reconcile the issues it caused to its users.

Indeed, a human rights framework is necessary, demonstrating that African human rights jurisprudence is a convenient framework for AI even if it may suffer some limitations and require future development. The paper explained that it incorporates a set of well-established, binding human rights norms, imposes responsibilities on both State actors and businesses, introduces rigorous accountability and oversight mechanisms and facilitates access to justice. Thus, establishing a human rights framework proves advantageous, considering the multitude of actors involved across the AI life cycle. This framework would enable addressing each party's responsibilities and mandates the adoption of accountability mechanisms to safeguard human rights against potential risks. It would also offer guidance on the necessary measures to be implemented.

This study was made possible by a grant provided by the International Development Research Center (IDRC). We thank the organisation for their continued support.