MODEL REGULATORY FRAMEWORK ON AI & MACHINE LEARNING IN SOUTHERN AFRICA

PROMOTING FREE EXPRESSION IN SOUTHERN AFRICA

PREAMBLE

States should be committed to proportionate regulation of the risks of digital technologies. They can do so by exploring a range of mechanisms, including improving the policymaking process. These regulatory measures include norms, self-regulation, statutory codes of conduct, and rules in primary legislation.

Given how digital technologies can quickly outpace law-making, States should also adopt non-regulatory tools to complement or provide alternatives to 'traditional' regulation, including industry-led technical standards.

As digital technologies accumulate vast amount of data, they demand a distinct regulatory approach. They are a complex system whose creation and use relies on critical elements such as algorithms, access to data, and sufficient computing resources.

To respond to these distinct challenges, States should take holistic, flexible, adaptable, transparent, and objective regulatory approaches that bring clarity by balancing the competing multi-stakeholders interests, including industry and civil society, as well as international commitments and obligations.

This guidance focuses on the right to freedom of expression enshrined in Article 9 of the African Charter on Human and Peoples Rights, Article 19 of the Universal Declaration of Human Rights, and Article 19 of the International Covenant on Civil and Political Rights (ICCPR) and the right to privacy in Article 12 of the Universal Declaration of Human Rights and Article 17 of the ICCPR.

1. ETHICS AND LAW TO ENSURE THAT AI SERVES THE PUBLIC GOOD.

States should firmly centre regulation on universal ethical principles and human rights as ultimate benchmarks for assessing the social acceptability of AI systems developed and deployed in and out of the African Union.

2. APPLICATION OF INTERNATIONAL HUMAN RIGHTS LAW.

States have a three-fold obligation to respect, protect, and fulfil human rights and fundamental freedoms.

3. FREEDOM OF EXPRESSION

The states shall take into account the preceding guidelines in adhering to the following principles in international law:

a. Respect and protect everyone's right to "seek, receive and impart information and ideas of all kinds, regardless of frontiers".

b. Whenever free speech is restricted through content moderation or otherwise, it must be provided by law and necessary for respecting the rights or reputation of others and protecting national security, public order, public health, or morals, for example, incitement and hate speech.

c. In categorising hate speech and incitement speeches, intermediaries and states may find the following helpful categorisation: At the top is incitement to genocide, incitement to violence, and possibly to terrorism, and incitement to discrimination. This is followed by other forms of hate speech, including vilification, glorification, promotion, and justification, which are part of the pyramid below incitement. At the bot-

tom of the pyramid is hate speech that disseminates, propagates, or spreads hostility (and these are deemed to be different from those that vilify or glorify because they really can be just mere avenues of dissemination as opposed to an intent to condemn or to celebrate acts or persons.

To avoid doubt, the States shall ensure that they do not use criminal law procedures to request intermediaries to take down defamatory statements or public criticism directed at politicians. Although intermediaries should remove defamatory content through civil procedural routes, defamation should not be criminal.

Further, because they rely on digital technologies to inform the public or to form opinions, journalists and bloggers are to be protected against abuse or intimidation. Journalists and bloggers should not be regularly prosecuted, jailed, or fined for libel.

4. THE RIGHT TO PRIVACY.

States must respect, promote, and protect the right to privacy of their citizens and can only use existing criminal and national security laws to limit the request in response to legitimate national security considerations and the necessities of law enforcement — in well-defined cases and under specific circumstances.

They can only justify the breaches of the right to privacy when necessary to achieve a legitimate aim prescribed by the law and proportionate to the purpose pursued.

When considering whether interference is justified, states shall assess the interference not only in the privacy of family, home, and correspondence but also, in certain circumstances, citizens' honour and reputation.

In meeting the above obligations, States shall: —
Rely on the existing cyber security and crime laws to set up multidisciplinary cybersecurity response teams and implement cybersecurity response strategies and plans to respond to old and emerging cybersecurity threats such as cybercrime and terrorism and distinguishing matters in the dual use of AI in the realm of security, defence, and critical national infrastructure.
States shall not: —

(a) use national security concerns as a blanket justification to excuse unwarranted privacy breaches, for example, unjustifiably increasing their legal and technical capabilities to closely monitor citizens and introducing measures that enable the collection of personal data through surveillance alongside invasive data retention modalities.

(b) Unjustifiably collect and store vast amounts of personal and intimate information about an individual or group's past or future actions. States should not subject journalists, lawyers, human rights defenders, and political activists to arbitrary and unlawful surveillance, either because they are being actively singled out for monitoring or simply because the Internet, often their primary means of communication, is subject to extensive monitoring.

(c) Perpetuate intentional suppression of legitimate dissent, curtailment of the right to free speech, and restriction of other citizens' right to access information.

States shall uphold the right to anonymity for online users as this gives them the confidence to report incidents without the fear of double victimisation.

States shall, therefore, ensure that encryption technologies are legally permitted and not take measures to undermine them or adopt real-name registration policies.

5. OPERATIONALISATION OF HUMAN RIGHTS AND ETHICS.

This part outlines some steps the states and the private sector may adopt in implementing digital tools, particularly AI and machine learning tools, to protect users' fundamental rights and safety.

(i). RISK ASSESSMENT AND IMPACT ANALYSIS

In the development, deployment, and implementation of AI, Southern African Governments must conduct a thorough and comprehensive risk assessment whereby they systematically evaluate the potential risks associated with the development, deployment, and use of AI systems by doing the following:

(a) Identify and analyse various risks, such as biases, security vulnerabilities, discriminatory outcomes, privacy breaches, and adverse societal impacts that AI may pose.

(b) Evaluate the likelihood and potential impact of identified risks, considering the context of the AI application and its intended use.

(c) Prioritise risks based on severity, potential harm, and likelihood of occurrence to effectively focus regulatory efforts and resources.

Insofar as the Impact of AI is concerned, states must evaluate the broader effects and consequences of AI on individuals, communities, economies, and societies at large by doing the following:

(a) Identify and categorise the impact areas, such as economic, social, ethical, legal, environmental, and political, that AI technologies may influence.

(b) Evaluate both positive and negative impacts of AI, considering advancements, job displacement, privacy enhancement, fairness, and equity, among other factors.

(c) Analyse AI's immediate and potential long-term effects on societal structures, employment patterns, education systems, and public services.

(ii). THE ROLE OF TECHNICAL OVERSIGHT.

(a) Independent Audits

African governments need to support a collaborative ecosystem of technical oversight and governance that includes independent parties, not just companies and governments, to bolster the trustworthiness of advances in AI innovation.

The technical oversight should include a data protection impact assessment (DPIA) and a fundamental rights impact assessment (FRIA). States must take the following steps to mitigate and reduce the harms of human rights violations from machine learning in public sector systems:

(b) Identify risks

Any state deploying machine learning technologies must thoroughly investigate systems for their potential to pose a risk to human rights before development or acquisition, where possible, before use, and on an ongoing basis throughout the lifecycle and contexts of the technologies.

This investigation may include:
a) Conducting regular impact assessments before public procurement, during development, at regular milestones, and throughout the deployment and use of machine learning systems to identify potential sources of discriminatory or other rights-harming outcomes — for example, in algorithmic model design, in oversight processes, or data processing.

b) Taking appropriate measures to mitigate risks identified through impact assessments — for example, mitigating the risk for misuse in amplifying tensions, undermining privacy, and controlling information; conducting dynamic testing methods and pre-release trials; ensuring that potentially affected groups and field experts are included as actors with decision-making

power in the design, testing, and review phases; submitting systems for independent expert review where appropriate.

c) Subjecting systems to live, regular tests and audits; interrogating markers of success for bias and self-fulfilling feedback loops; and ensuring holistic independent reviews of systems in the context of human rights harms in a live environment.

d) Disclosing known limitations of the system in question — for example, noting measures of confidence, known failure scenarios, and appropriate use limitations.

(III). ENSURING TRANSPARENCY AND ACCOUNTABILITY.

States must conduct a realistic assessment of the capabilities and limitations of AI and must ensure and require accountability and maximum transparency around public sector use of machine learning systems.

States should:
(a) Publicly disclose to the public sphere use of machine learning systems and provide information that explains in clear and accessible terms how automated and machine learning decision-making processes are reached—document actions are taken to identify, document, and mitigate against human rights harming impacts.

(b) Enable independent analysis and oversight by using auditable systems.

(c) Avoid using 'black box systems' and only use systems that meet meaningful standards of accountability and transparency, and refrain from using these systems in high-risk contexts.

(IV). ENFORCING OVERSIGHT

States must adopt oversight mechanisms that identify, examine, resolve, and test biases in datasets and the machine learning model throughout the designing and development phases.

Implementing oversight mechanisms may ensure that the datasets used are not deficient, outdated, or insufficient.

States should:
(a) Proactively adopt diverse hiring practices and engage in consultations to assure diverse perspectives so that those involved in designing, implementing, and reviewing machine learning represent a range of backgrounds and identities.

(b) Ensure that public bodies carry out training in human rights and data analysis for officials involved in the procurement, development, use, and review of machine learning tools.

(c) Create mechanisms for independent oversight, including by judicial authorities.

(d) Ensure that machine learning-supported decisions meet internationally accepted standards for due process.

Any state authority procuring machine learning technologies from the private sector should maintain relevant oversight and control over the use of the system and require the third party to carry out human rights due diligence to identify, prevent, and mitigate against discrimination and other biases.

Transparency requires greater engagement with digital rights organisations and other relevant civil society sectors.
Internet platforms, mainly social media, should adhere to open communication, follow an open and transparent decision-making process, and openly publicise findings in contested cases given internet platforms such as Facebook's impact on the public sphere.