alt.AI

A.I.MPACT

An overview of resources on AI impact assessments

A.I.MPACT
An overview of resources on AI impact assessments
ALT Advisory, 2022

This report forms part of a research series produced for **ALT AI**, an ALT Advisory Special Project focused on safeguarding fundamental rights in an AI-driven future. Access all publications at ai.altadvisory.africa.

The research team includes Tara Davis, Murray Hunter, Zahra Abba Omar, Wendy Trott and Christy Chitengu. This report was designed by Wilna Combrinck.

Suggested citation
ALT Advisory, *A.I.MPACT*, September 2022, accessible at ai.altadvisory.africa.

Acknowledgements
This work was carried out in the context of the Africa Digital Rights Fund ("ADRF") with support from the Collaboration on International ICT Policy for East and Southern Africa ("CIPESA").

Disclaimer
Please note that while every attempt has been made to ensure that the information in this report is up-to-date and accurate, there may be errors and omissions. The research in this report is provided for general guidance on matters of interest and does not constitute legal advice. ALT Advisory is not responsible for any errors or omissions, or for the results obtained from the use of this information.

Connect with ALT Advisory
altadvisory.africa
@altadvisory
connect@altadvisory.africa

TABLE OF CONTENTS

ACRONYMS

ADRF	The Africa Digital Rights Funds
AI	Artificial Intelligence
AIA	Algorithmic Impact Assessments
CAHAI	The Council of Europe's Ad Hoc Committee on AI
CIPESA	The Collaboration on International ICT Policy for East and Southern Africa
EU	European Union
FRIA	Fundamental Rights Impact Assessments
FTC	The Federal Trade Commission
HRESIA	Human Rights, Ethical and Social Impact Assessment
HRIA	Human Rights Impact Assessment
RIA	Risk and impact assessments
UN	United Nations
UNHCR	United Nations Human Rights Committee

GLOSSARY

Artificial Intelligence	A broad term for a computer or software system's ability to be programmed to 'think' like a person, for example, to analyse information, look for patterns, or make decisions.
Algorithm	A set of rules or instructions that a computer or software system is programmed to follow in order to process information or perform a task.
Automated processing	Any tech-enabled processing of personal information without ongoing human involvement.
Biometrics	The recording of a person's physical or biological information, such as fingerprints, retinal scans, voice features, or facial features, as a way of identifying them.
Data Subject	The person to whom personal information relates.
Impact Assessment	An impact assessment is a mechanism that has been developed to identify, predict, understand, assess, and address adverse consequences of something such as a project, tool, or policy. They vary in form and name and are sometimes referred to as Human Rights Impact Assessments ("**HRIAs**"), Algorithmic Impact Assessments ("**AIAs**"), or Fundamental Rights Impacts Assessments ("**FRIAs**"). We use 'impact assessment' throughout this report to denote a broad scope that includes different types of assessments.
Machine learning	An AI technique that uses existing datasets to teach a computer or software system how to solve a specific problem, by detecting patterns and similarities in existing and new data.
Personal information / personal data	Information relating to a data subject that identifies the data subject. This includes but is not limited to contact information, information relating to race, gender, sex, pregnancy, national, ethnic, or social origin, information relating to medical, financial criminal, or employment history, and biometric information.
Processing	Any operation or activity concerning personal information which includes but is not limited to the collection, recording, collation, storage, alteration, and use of personal information.

INTRODUCTION

Physicist Stephen Hawking** famously warned **about the dangers of artificial intelligence ("AI"), saying that "[t]he development of full artificial intelligence could spell the end of the human race."

This statement might be hyperbole. Yet AI discourse is littered with examples of real-life harms caused by its deployment, such as an Amazon recruitment tool that favoured men, a healthcare tool that unfairly denied Black patients access to kidney transplants in the US, or security camera software in South African suburbs that seemed to flag (Black) pedestrians as 'suspicious'. The resultant harm of these AI systems impacted the lives of individuals and undermined their rights to equality and life.

The potential harm posed by AI has prompted various responses and policy proposals aimed at mitigation – including the use of impact assessments. Impact assessments evaluate any potential or actual harm caused by something, such as a project or plan, and aim to correct or mitigate the harm. Impact assessments are already a common policy feature in many sectors internationally (for example, in environmental or economic affairs, and in human rights governance) and have been proposed as a mechanism to evaluate the impacts of AI systems.

Critics have argued that impact assessments are a limited tool that do not adequately protect against human rights abuses in the domain of AI. However, the complexity and challenges for regulating AI technologies suggest that effective oversight and regulation is only possible through multi-layered approaches that use a range of tools and mechanisms in tandem, to offset each one's limitations. In addition, the growing field of research and policy work on AI oversight offers new and improving guidance and resources on effective implementation of impact assessments in the domain of AI.

The purpose of this report is to collate this existing work to equip practitioners with the necessary information to develop an effective and context-specific impact assessment for artificial intelligence. Our hope is that these available resources will enable practitioners to guard against the possible harms and rights violations posed by certain AI technologies – and, perhaps, to avert Hawking's fear of the *end of the human race*.

THE IMPACT OF AI ON HUMAN RIGHTS

The European Commission has broadly defined the constellation of technologies that make up AI as systems "that display intelligent behaviour by analysing their environment and taking actions—with some degree of autonomy—to achieve specific goals."

The ability of a machine to perform multiple tasks across several domains is called artificial general intelligence. While such examples of machines with highly sophisticated human-like intelligence have been realised in the popular imagination, spanning from Fritz Lang's *Metropolis* to Alex Garland's *Ex Machina*, current technology available to us has yet to achieve this.

Instead, the applications of AI today are still limited to what is called *narrow* or *weak* AI – artificial intelligence that can only operate within a limited, predefined range of functions. Under the umbrella of AI are technologies that rely on massive amounts of real-time data to emulate the human capacity for sense (through computer vision and audio processing), for comprehension (such as natural language processing systems), and for action (such as machine learning, an AI technique that uses training data consisting of many input/output examples to teach systems how to reliably and accurately solve specific problems in a given domain).

AI technologies as enablers of rights

One dominant narrative regarding AI frames its technologies as enabling opportunities for the African continent and beyond, with AI as a driver of growth, development, and for realising human rights. Rachel Adams notes that "this narrative is perhaps even stronger in the so-called developing world, where ideas of transformation are central to the project of state-making."

Healthcare, for instance, is undergoing significant transformation through the implementation of AI technologies, which have the potential to improve diagnostics and make the provision of healthcare more affordable and accessible. In South Africa, the Institute for Intelligent Systems at the University of Johannesburg has deployed AI technology through a service that automatically and remotely monitors patients' vital signs in real-time and collects, processes and analyses the resultant data. The device is called "e-mutakalo" (Tshivenda for "health") and alerts medical professionals in the case of an emergency. The Ugandan start-up Wekebere, provides belts to expectant mothers that provide readings of their vital signs and that of their unborn children. The belt produces results through the handheld device, a mobile app as well as stores the information from the mobile app to the Wekebere cloud platform. To date, 15 000 women have used the Wekebere to provide insights on their maternal health journey that are generated from the app's identifying of patterns within pregnancy. Such deployments of AI advance the right to healthcare as recognised in the African Charter on Human and Peoples' Rights ("**African Charter**") under Article 16, which states that every individual shall have the right to enjoy the best attainable state of physical and mental health.

AI technologies' potential harms to human rights

AI has the potential to radically change our societies and play a role in addressing inequity. Considering this, it is important to maintain a critical lens of artificial intelligence, to discern who is designing these technologies, whose interests the technologies protect, who is left out of their design, and how this might inhibit the realisation of fundamental human rights.

In June 2020, Tendayi Achiume, United Nations Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia, and related intolerance, issued a report to the UNHCR that found "emerging digital technologies exacerbate and compound existing inequalities, many of which exist along racial, ethnic, and national origin grounds."

The rights enshrined in Articles 2 and 3 of the African Charter – the right to freedom from discrimination and the right to equality before the law and equal protection of the law – are essential principles that make up the Charter's anti-discrimination and equal protection provisions. Per the United Nations Human Rights Committee ("**UNHCR**"), an individual is unfairly discriminated against if, without just cause, and on a prohibited ground, they are conferred treatment that is less favourable which results in a negative effect on them.

The use of AI can, and has been proven to, impact on the enjoyment of the right to equality and protection from discrimination. This generally happens where AI-driven decision systems reflect and reinforce human biases, and result in outcomes which are discriminatory while being accepted, or presented, as objective. Where a dataset is unrepresentative or skewed, it can lead to an algorithm having less favourable outcomes for certain groups. This allows for the transference of prejudices from the human participants and the data to create an algorithm that is inherently biased – often referred to as coded bias. In sum, datasets reflect biases, and the choice to use discriminatory code, whether in the realm of employment, healthcare, education, or criminal justice, can reproduce and sediment prejudices.

As AI technologies are increasingly deployed, there is growing policy urgency in ensuring protections for the human rights of those rendered marginalised and vulnerable by digital change. It is for this reason that the use of impact assessments has been encouraged, and in some instances legally required, to identify and mitigate against the potential harm of AI.

WHAT IS AN IMPACT ASSESSMENT?

An impact assessment is a mechanism that has been developed to identify, predict, understand, assess, and address adverse consequences of something such as a project, tool, or policy. Such assessments generally aim to determine the likely impact the intervention would have on human rights; if the potential harm outweighs the benefit, it could result in a project or initiative being discontinued, or subject to redesign and further review. Impact assessments are often a regulatory requirement for the public and private sectors as part of their due diligence processes. Common examples include environmental impact assessments and data protection impact assessments.

Impact assessments may be *ex-ante* (conducted before something is developed, deployed, or procured) or *ex-post* (after something has been completed or deployed). Importantly, impact assessments are grounded in the context of the thing that is being assessed, and the assessment parameters and any remedial steps in an impact assessment should always be adapted to the relevant context.

According to recommendations by Data & Society and the European Centre for Not-for-Profit-Law, impact assessments serve the following purposes:

1. "Providing an ex-ante or ex-post assessment of the potential or actual impacts of a technology, policy, or business practice.
2. Creating a document of these impacts that can be shared with stakeholders, regulators, or members of the public.
3. Providing a reflexive exercise for developers of a policy or technology to question what intended outcomes they hope to achieve, and what mitigative measures they may need to put in place to address potentially harmful outcomes.
4. Providing a mechanism for developers and policy makers to engage with a range of stakeholders who may be affected by that policy or technology."

A framework developed by AI Now for AI impact assessments identifies four policy goals:

1. "Respect the public's right to know which systems impact their lives by publicly listing and describing automated decision systems that significantly affect individuals and communities;
2. Increase public agencies' internal expertise and capacity to evaluate the systems they build or procure, so they can anticipate issues that might raise concerns, such as disparate impacts or due process violations;
3. Ensure greater accountability of automated decision systems by providing a meaningful and ongoing opportunity for external researchers to review, audit, and assess these systems using methods that allow them to identify and detect problems;
4. Ensure that the public has a meaningful opportunity to respond to and, if necessary, dispute the use of a given system or an agency's approach to algorithmic accountability."

COLLATED RESOURCES

There is a growing body of research on how to effectively develop and implement an impact assessment in the context of AI. We overview a selection of these resources, sourced from global experts on the human rights implications of AI, that provide guidance and insight to practitioners looking to design and implement impact assessments of their own.

Table 1 | Resources on impact assessments

Date	Organisation/ Author and Title	Description
2011	United Nations Office of the High Commissioner on Human Rights United Nations Guiding Principles on Business and Human rights	Part II of the guiding principles provides for the corporate responsibility to respect human rights, as defined in the International Bill of Human Rights and the International Labour Organisations Declaration on Fundamental Principles and Rights at Work. It recommends that businesses conduct human rights due diligence processes to identify, prevent, mitigate, and account for how they will address the adverse human rights impacts of their activities. It provides that these processes should draw on internal and/or independent human rights expertise and involve meaningful consultation with potentially affected groups. Assessments should be taken at regular intervals, and the findings should be integrated across relevant internal functions and processes. Appropriate action should be taken and the effectiveness of responses tracked with feedback from both internal and external sources. Businesses should also communicate concerns externally and provide for remediation where they have caused adverse impacts.

2018	United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression Report A/73/348 Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression	This report addresses Artificial Intelligence technologies and their implications for the information environment and highlights that HRIAs can serve as a tool to enable transparency throughout the lifecycle of AI. It recommends that they should be performed before procurement, development, or use and involve both self-assessment and external review. It also states that public sector procurement of AI technologies from private vendors should be accompanied by a public consultation to elicit societal views and input on the design and implementation of the AI system. It recommends that this should be done before it is acquired and that the external review of AI systems provides a critical guarantee of rigour and independence in transparency and ongoing independent audits should therefore supplement pre-procurement human rights impact assessments. It recommends that States should ensure that human rights are central to private sector design, deployment, and implementation of AI systems by pursuing regulatory or co-regulatory schemes designed to require businesses to undertake impact assessments and audits of AI technologies and ensuring effective external accountability mechanisms. Finally, it recommends that the results of human rights impact assessments and public consultations should themselves be made public.
2018	AI NOW Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability	This report provides a practical framework to assess automated decision systems used by public agencies and to assess the impacts of these systems. The AIA framework proposed by the authors is designed to support affected communities and stakeholders to assess the claims made about these systems and to determine where – or if – their use is acceptable. The key elements of a public agency AIA proposed are that: • "Agencies should conduct a self-assessment of existing and proposed automated decision systems, evaluating potential impacts on fairness, justice, bias, or other concerns across affected communities; • Agencies should develop meaningful external researcher review processes to discover, measure, or track impacts over time; • Agencies should provide notice to the public disclosing their definition of "automated decision system," existing and proposed systems, and any related self-assessments and researcher review processes before the system has been acquired; • Agencies should solicit public comments to clarify concerns and answer outstanding questions; and • Governments should provide enhanced due process mechanisms for affected individuals or communities to challenge inadequate assessments or unfair, biased, or otherwise harmful system uses that agencies have failed to mitigate or correct."
2019	Draft Bill introduced in the US House of Representatives, sponsored by Rep. Yvette Clarke (D-NY-9) ("**the Bill**") 2019 Algorithmic Accountability Act ("**AAA**")	The Bill mandates the Federal Trade Commission ("**FTC**") to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments. The Bill defines what it perceives to be an adequate impact assessment, which includes that it must (1) describe the system in detail, (2) assess the relative costs and benefits of the system, (3) determine the risks to the privacy and security of personal information, and (4) explain the steps taken to minimise those risks, if discovered. Assessments of high-risk information systems involving personal information must evaluate the extent to which the system protects the privacy and security of such information. The entities covered should conduct these assessments, preferably with the consultation of third-party, independent auditors and technology experts. Entities have the option to publish the results after submitting them to the FTC.
2019	Government of Canada Canadian Directive on Automated Decision-Making, promulgated under section 7 of the Financial Administration Act and section 4.4.2.4 of the Policy on Service and Digital ("**the Directive**").	The directive applies to Federal Government Departments that currently or intend to use automated decision systems. The systems to be assessed are those that recommend or make administrative decisions about end-users. The directive requires the Assistant Deputy Minister responsible for the program deploying the system to complete an AIA prior to implementation and to comply with the requirements specified in the Act. These requirements differ depending on the level of risk presented by the system and include peer review, notice of the use of the system to be published, that humans must be in the loop for certain decisions, meaningful explanations are to be provided along with training, contingency plans are to be put in place, and, in some circumstances, that approval is given for the system to operate. It also requires that the final results of the impact assessment be released in an accessible format via Government of Canada websites.

2020	European Commission Independent High-Level Expert Group on AI ("**HLEG**") Ethics Guidelines for Trustworthy Artificial Intelligence ("**Ethics Guidelines**") and Assessment list for Trustworthy AI ("**ALTAI**")	The HLEG published the Ethics Guidelines in 2019 to enable organisations and companies to conduct self-assessments on AI systems being developed, deployed, procured, or used. Following that, in 2020, the ALTAI was developed which translates the ethics guidelines into an accessible and dynamic checklist that guides developers and deployers of AI in implementing such principles in practice. The Assessment List includes seven requirements for trustworthy AI systems, which are: Human agency and oversight; • Technical robustness and safety, including resilience to attack and security, as well as accuracy, reliability, and reproducibility; • Privacy and data governance; • Transparency, including traceability, explainability, and the communication of a system's capabilities and limitations to users; • Diversity, non-discrimination, and fairness, including accessibility and universal design as well as stakeholder participation; • Societal and environmental wellbeing, which incorporates the impact on work and skills; and • Accountability, which means auditability and risk management. The ALTAI also recommends that before the self-assessment using ALTAI, companies, and organisations should perform "Fundamental Rights Impact Assessments" ("**FRIAs**") by asking questions drawn from the specific protections of the European Charter on Fundamental Rights, the European Convention on Human Rights, the European Social Charter and their Protocols, of which four examples are given.
2020	Access Now Access Now's submission to the Consultation on the "White Paper on Artificial Intelligence - a European approach to excellence and trust"	Access Now recommends that the use of HRIAs be a mandatory requirement for both private and public sector entities deploying AI. This includes the creation of a publicly viewable pre- and post-hoc conformity HRIA and requiring actors that deploy AI to sign on to a publicly viewable AI register. Access Now also provides guidelines on how to design the mandatory HRIA requirement.
2021	Data & Society and the European Centre for Not-for-Profit Law Recommendations for Assessing AI Impacts to Human Rights, Democracy, and the Rule of Law	Noting that the Council of Europe's Ad Hoc Committee on AI ("**CAHAI**") has proposed the development of a "uniform model" for human rights impact assessments of AI systems, the authors assess the limitations of HRIAs for algorithmic accountability and advance a framework for evaluating impact assessment processes that will help to create more accountable and rights-respecting AI systems. The authors make the following recommendations: • Impact assessments should be legitimised either through legislation or within a set of norms that are officially recognised and publicly valued; • Impact assessments should establish an accountability relationship between actors that design or deploy a system and a forum that is empowered to assign responsibility for potential consequences and demand change in the design, deployment, and operation of such systems; • There should be clearly defined points in the development and/or procurement process that trigger a requirement to conduct impact assessments; • The time frame in which an impact assessment is conducted must cover all the stages of the AI lifecycle, starting with the ideation stage and running through post-deployment. • The public must have the ability to scrutinise and contest an impact assessment's process and documentation; • The conditions for solicitation of feedback should be from the broadest possible set of stakeholders in a system; • The methods deployed in an impact assessment process should be developed further in partnership with affected communities and stakeholder groups; • Assessors must be independent and have expansive expertise, and must have full access to all the information required to make an informed assessment; • A holistic and open approach must be taken to assessing the impacts of AI systems on a wide range of human rights.

2021	Data & Society Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts	This paper analyses the history and nature of impact assessments from other domains, the emerging and existing proposals for AIAs, and findings from past AIAs. It makes the following recommendations: • Designers, operators, and maintainers of algorithmic systems should be obliged to identify, explain and justify the potential harms of algorithmic systems; • The assessment metrics should dynamically change and not be rigid templates; • Assessors should consider that harms are context-dependent and affect different communities and individuals differently; • External expertise should be sought to police the system to avoid the potential for bias; • Deployers of AI systems should assemble multi-disciplinary expertise; • The individuals and communities affected and potentially affected by the algorithmic systems are the 'foremost experts' and therefore should be included in assessment design and implementation; and • The outcomes of AIAs should be presented and considered by the people who have the power to change the system.
2021	Louis Au Yeung, Center for Long-Term Cybersecurity, UC Berkeley Guidance for the Development of AI Risk and Impact Assessments	This paper provides a comparative analysis of AI risk and impact assessments ("**RIA**s") from five regions around the world: Canada, New Zealand, Germany, the European Union, and San Francisco, California. It makes the following recommendations, amongst others: • Certain risk mitigation measures are emphasised across all the surveyed frameworks and should be considered essential, such as human oversight, external review and engagement, documentation, testing, and mitigation of bias, alerting those affected by an AI system of its use, and regular monitoring and evaluation; • In addition to assessing impacts on safety and rights, it is important to account for impacts on inclusiveness and sustainability to protect the wider interests of society and ensure that marginalised communities are not left behind; • Individuals and communities affected by the use of AI systems should be included in the process of designing risk and impact assessments to help co-construct the criteria featured in the framework; • Risk and impact assessments should include banning the use of specific AI systems that present unacceptable risks, to ensure that fundamental values and safety are not compromised; • Periodic risk and impact reassessments should be required to ensure that continuous learning AI systems meet the standards required after they have undergone notable changes, and should be performed as early as the design stage; • Risk and impact assessments should be tied to procurement and purchase decisions to incentivise the use of voluntary frameworks; • In assessing harm, sustainability has to be a factor, including taking into account the adverse impacts of computationally intensive models and energy-intensive data centres.
2021	European Commission The European Union Artificial Intelligence Act ("**AI Act**")	The AI Act is a proposed European law on AI. It uses a risk-based approach in which AI is classified according to its relative threat to fundamental rights: low/minimal risk, high risk, and unacceptable risk. AI systems considered as unacceptable risks are prohibited and include systems that have "significant potential to manipulate persons through subliminal techniques" or to "exploit vulnerabilities of specific vulnerable groups" as well as social scoring systems and real-time biometric identification systems. The AI Act requires providers of high-risk systems to create Risk Management Systems that run throughout the algorithmic lifecycle, and which identify and analyse foreseeable risks and adopt measures to manage the risk. They are also required to carry out data protection impact assessments under Article 35 of EU Regulation 2016/679 or Article 27 of EU Directive 2016/680.
2021	Latonero and Agarwal, Carr Center for Human Rights Policy, Harvard Kennedy School Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar	By reviewing a specific example of an HRIA, the paper assesses whether HRIAs are fit for purpose for AI. It finds that HRIAs should be updated if they are to be used for AI. Further, interdisciplinary expertise is needed to determine the appropriate methods and criteria for specific contexts where AI systems are deployed. HRIAs should be conducted at appropriate times relevant to critical stages in an AI development lifecycle, and a mix of voluntary and mandatory measures may be needed to incentivise organisations to incorporate HRIAs for AI in a more effective, transparent, and accountable way.

2021	Data & Society and the European Centre for Not-for-Profit Law Mandating Human Rights Impact Assessments in the AI Act	In a commentary on the European Union's efforts to develop a legally binding framework on AI based on the EU's standards on fundamental rights, the authors provide a set of recommendations for establishing an HRIA process for AI: • Require developers and deployers to conduct HRIAs; • Determine the criteria for assessing impacts on human rights; • Ensure public access; • Establish an oversight mechanism; • Develop methods for participatory inclusion, public consultation, and appeal; • Establish an HRIA research agenda; and • Integrate HRIAs with other accountability mechanisms.
2022	Alessandro Mantelero Human Rights Impact Assessment and AI	This chapter is part of the Information Technology and Law Series book series. It proposes a model for (HRIAs) as part of the broader Human Rights, Ethical and Social Impact Assessment ("**HRESIA**") model. The authors argue that the model was developed in line with the existing practices in HRIAs, but in a way that better responds to the specific nature of AI applications in terms of scale, impacted rights and freedoms, prior assessment of production design, and assessment of risk levels, as required by several proposals on AI regulation. The model relies on the following building blocks commonly used in traditional HRIAs: planning and scoping, data collection (including rightsholder and stakeholder consultation), and analysis.
2022	Microsoft Microsoft Responsible AI Standard, v2: General Requirements ("**the Standard**")	The Standard is intended to serve as concrete and actionable guidance to product development teams on how to develop responsible AI in the absence of effective laws and norms. Goal A1 of the Standard is that all Microsoft AI systems are assessed using impact assessments, which include an evaluation of the system's impact on people, organisations, and society early in its development and documenting these efforts using a standardised template. The assessment must be reviewed and approved by a pre-defined set of reviewers (determined by the organisation's compliance process) before development starts. The Impact Assessment must be updated and reviewed at least annually when new intended users are added, and before advancing to a new release stage. The Standard also deals with reviews to identify systems that may require additional oversight and requirements and with how to ensure appropriate human oversight and control over AI systems. Further, it deals with mechanisms to ensure that identified stakeholders can intelligibly interpret the systems and provide oversight. It provides specific requirements for providing information about the capabilities and limitations of AI systems to stakeholders and notifying users that they are interacting with an AI system. Finally, it provides procedures to minimise the risk of bias or discrimination and steps to be taken to prepare for failures and remediations.

A SUMMARY OF RECOMMENDATIONS

Below is a selection of the recommendations included in the collated resources.

Procedural Recommendations

When should AI impact assessments be conducted?

- Assessments should be conducted before procurement, development, or use (*United Nations Special Rapporteur on freedom of expression and opinion*);
- Periodic impact reassessments should be required to ensure that continuous learning AI systems meet the standards required after they have undergone notable changes, and should be performed as early as the design stage(*Louis Au Yeung*);

Who should be involved?

- The individuals and communities affected and potentially affected by the algorithmic systems are the 'foremost experts' and therefore should be included in assessment design and implementation (*Data & Society*);

- Individuals and communities affected by the use of AI systems should be included in the process of designing risk and impact assessments to help co-construct the criteria featured in the framework (*Louis Au Yeung*);
- Deployers of AI systems should assemble multi-disciplinary expertise (*Data & Society*);

Governance and accountability

- Assessments should entail self-assessment and external review (*United Nations Special Rapporteur on freedom of expression and opinion);*
- The outcomes of AIAs should be presented and considered by the people who have the power to change the system (*Data & Society*);
- Public bodies should develop meaningful external researcher review processes to discover, measure, or track impacts over time (*AI NOW*);
- External expertise should be sought to police the system to avoid the potential for bias (*Data & Society*);
- The public sector should conduct public consultations before procuring AI technologies from the private sector (*United Nations Special Rapporteur on freedom of expression and opinion*);
- Develop methods for participatory inclusion, public consultation, and appeal (*Data & Society and The European Centre for Not-for-Profit Law*);
- Establish an oversight mechanism (*Data & Society and The European Centre for Not-for-Profit Law*);
- Mechanisms should be developed to enable individuals to challenge inadequate assessments or to note consequences that were not considered and/or corrected (*AI NOW*);
- Integrate HRIAs with other accountability mechanisms (*Data & Society and The European Centre for Not-for-Profit Law*).

Transparency

- The results of human rights impact assessments and public consultations should be made public (*United Nations Special Rapporteur on freedom of expression and opinion);*
- Public bodies should provide notice to the public disclosing their definition of "automated decision system," existing and proposed systems, and any related self-assessments and researcher review processes before the system has been acquired (*AI NOW*);
- Ensure public access (*Data & Society and The European Centre for Not-for-Profit Law*).

Additional

- Risk and impact assessments should be tied to procurement and purchase decisions to incentivise the use of voluntary frameworks (Louis Au Yeung).

Substantive Recommendations

- The assessment metrics should dynamically change and not be rigid templates (Data & Society);
- Be cognisant that harms are context-dependent and affect different communities and individuals differently (Data & Society);
- Account for impacts on inclusiveness and sustainability to protect the wider interests of society and ensure that marginalised communities are not left behind (Louis Au Yeung);
- Risk and impact assessments should include banning the use of specific AI systems that present unacceptable risks, to ensure that fundamental values and safety are not compromised; (Louis Au Yeung)
- Sustainability has to be a factor when assessing harm, including taking into account the adverse impacts of computationally intensive models and energy-intensive data centres (Louis Au Yeung);
- Designers, operators, and maintainers of algorithmic systems should be obliged to identify, explain and justify the potential harms of algorithmic systems (Data & Society); and
- Determine the criteria for assessing impacts on human rights (Data & Society and The European Centre for Not-for-Profit Law).

CONCLUSION

An impressive and extensive body of work exists to enable the development and implementation of effective impact assessments for AI. While acknowledging the limitations of impact assessments, these resources may inform vital safeguard against potential harms caused by the deployment of AI. Yet, there remains a troubling lack throughout Africa of AI regulation that would mandate the use of impact assessments, or any other measures to limit possible AI-based rights infringements.1 We will continue to monitor developments in this area and hope it assists practitioners as they work to identify, predict and correct harmful consequences of AI.

1 See ALT Advisory, *AI Governance in Africa*, July 2022, accessible at ai.altadvisory.africa.