Towards Afro-feminist AI:

A Handbook for approaching Governance of AI in Africa

POLLICY

Research and Writing by Amber Sinha and Bobina Zulfa.

Review by Neema Iyer

Contents

01 **Background**

02 **The need for an Afro-feminist Approach to AI**

04 **An Afro-feminist Playbook for regulating AI**

- 04 **Forms of Regulatiton**
- 07 **Regulatory Strategy**
  - 08 Mandatory v. Voluntary
  - 08 Modelling of Human Behavior
  - 09 Equality and Non-discrimination
  - 09 Safety and Human Impact
  - 11 Data Exploitation
  - 11 Design Principles for Afro Feminist AI

Background

A recent opinion piece in the New York Times cited a survey of over 700 top academics and researchers, half of which were of the opinion that there was a “10 per cent or greater chance of human extinction (or similarly permanent and severe disempowerment) from future A.I. systems.”[1] Other reputable sources including The New Yorker[2] and Vox[3] covered this news. The fear-mongering about Artificial Intelligent systems (AI) is perhaps only matched by its hype, both of which ascribe sentient, superhuman abilities to schemes which are essentially statistical models running on very large quantities of data. These over-inflated fears turn our much-needed attention away from the real-life dangers and benefits of AI systems, which are numerous and far more prosaic. As we write this, machine learning algorithms (ML) intermediate our use of everyday services, determining how we spend our time, what content we read, view and consume, what opportunities we pursue, and perhaps, what we think.

The global promotion of using AI for decision-making in public and private functions has been celebrated as a way to enhance human abilities, eliminate bureaucratic obstacles, and bring advantages to society. However, there is growing acknowledgement that algorithms can worsen existing structural inequality and pose a threat to fundamental constitutional values, due to concerns about bias, fairness, and the lack of algorithmic accountability.

Across Africa, with the involvement of policymakers, universities, large companies, start-ups, and multi-stakeholder partnerships to varying degrees, we see a large uptake in the development and deployment of AI. Their use cases and spread over regions are diverse from chatbots to provide healthcare services to people without visiting doctors in Kenya[4] to online shopping services such as TakeAlot in South Africa and Konga in Nigeria.[5]

AI and other automated technologies broadly described as the Fourth Industrial Revolution (4IR) are often positioned as an antidote to problems of disparity and inequity. However, several studies have documented women’s lack of access to the myriad opportunities afforded by the 4IR in Africa.[6] It is a fiction that the old order will simply be replaced by post-gender structures. On the contrary, AI is located very much within, reinforces and often amplifies the old gendered and racial structures of power.

---

[1] Yuval Harari, Tristan Harris and Aza Raskin, “You Can Have the Blue Pill or the Red Pill, and We’re Out of Blue Pills.” New York Times. Published on March 24, 2023. Available at https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html.
[2] Jaron Lanier, “There is no AI.” The New Yorker. Published on April 20, 2023. Available at https://www.newyorker.com/science/annals-of-artificial-intelligence/there-is-no-ai.
[3] Kelsey Piper, “AI experts are increasingly afraid of what they’re creating.” Vox. Published on November 28, 2022. Available at https://www.vox.com/the-highlight/23447596/artificial-intelligence-agi-openai-gpt3-existential-risk-human-extinction.
[4] Phiri M, Munoriyarwa A. Health Chatbots in Africa: Scoping Review. J Med Internet Res. 2023 Jun 14;25:e35573. doi: 10.2196/35573. PMID: 35584083; PMCID: PMC10337242.
[5] Tom Jackson, “The dawn of an African e-commerce goldrush may be a false one.” Quartz. PublishedMay 23, 2016. Available at https://qz.com/africa/689864/the-dawn-of-an-african-e-commerce-goldrush-may-be-a-false-one.
[6] Celine Mulrean, “Women in the Fourth Industrial Revolution: A Gendered Perspective in DIgitalization in Kenya, Nigeria and South Africa.” Master’s Thesis. Centre International de Formation Europeenne Institut Europeen. Available at https://www.ie-ei.eu/Ressources/FCK/image/Theses/2020/MULREAN_Thesis_GEGPA_2020.pdf.

The need for an Afro-feminist Approach to AI

In our longer paper, we touch upon the idea of VELAI () drawing from a variety of non-exhaustive examples where interventions to a new approach to AI are sorely needed. Afro-feminism has emerged as a fresh lens "to consider decolonisation and decoloniality because it gives voice to and perspective from African women."[7] The focus is on centring the advocacy and research of women in the fight against the injustices of colonial effects. We believe that the ideals of Afro-feminism are critical guiding principles with which to understand and govern AI in Africa. Below, we highlight a few ways in which AI can be grounded in Afro-feminism.

01 Harmony between collective and individual data rights

This sense of community, exemplified by the African concept of Ubuntu, is acknowledged by scholars like Sylvia Tamale in stark contrast to the prevailing universalism and individualism that dominate discussions on rights. These are parallel processes that have accompanied formal, legal, rights-based approaches in Africa. Among these processes is the Gacaca Process, which is juxtaposed with the International Criminal Tribunal for Rwanda (ICTR). Both initiatives aimed to address the aftermath of the genocide in Rwanda in 1994. Gacaca, meaning "judgement on the lawns" in Kinyarwanda, represents a traditional African justice system that emphasises reconciliation and reparations. Unlike the ICTR, it operates within the context of the entire family and community, rather than focusing solely on individuals.[8]

The African Union Data Policy and Framework draws on this idea to encourage Member States to adopt novel approaches, including prioritising collective privacy rights and the need for data stewardship and other forms of data trusts.[9]

02 Confronting AI Exploitation

The data economy which drives AI is inherently exploitative. The global data economy today is both structured around and dependent on entrenched power asymmetries, further exacerbated by the control asserted by powerful entities. As the digital lives of individuals are transformed into computable data to be used for the aggrandisement of capital, users face a systematic deprivation of control. The erosion of control is not uni-dimensional — it is intersectional and dynamic.

[7] Corinne Knowles, (2021). Decolonization and Afro-Feminism. Journal of Contemporary African Studies. Doi:10.1080/02589001.2021.1938976.
[8] Under the Gacaca, the process of justice engaged the participation of the victims, the offenders and their respective community members who determined the guilt or innocence of the suspect before them since it was the same community which had witnessed and participated in the killing of their own members. In other words, Rwandan themselves were responsible for dealing with suspects of crimes committed in Rwanda by Rwandans against fellow Rwandans. Unlike the ICTR process, rather than simply punishing the perpetrator, justice was aimed more at reconciling the parties and reintegration of offenders.
[9] AU Data Policy Framework, African Union. Available at https://au.int/en/documents/20220728/au-data-policy-framework

Afro Feminist AI builds on the recognition of a need[10] to redistribute power over data in a manner that is fair and equitable to all actors involved in its generation. Big tech companies located in the North effectively block African countries' attempts to design and deploy their own AI technologies. For instance, the capture of advertising revenue in the media sector by Google and Facebook in South Africa has severely impaired the traditional media companies and their ability to compete.[11]

03 Visibilising workers

The myth of seamless, frictionless automation is supported by efforts of "ghost" workers often from the South who work as informal, short-term labour in a range of tasks such as data labelling. The skewed power structures between corporations and workers in the gig-economy devalues and invisibilises their contribution to fit the narrative of automation.[12] There is a need for regulation to illuminate, recognise and protect workers from exploitative labour practices.

04 Centering decisional autonomy

The datafication and extraction characterising information capitalism stifles human dignity and agency by colonising and commodifying one's innermost self.[13] The normative idea of dignity is premised on the inherent value of human life- requiring that all human beings be treated with respect at all times, regardless of status. In order for individuals and communities to have autonomy in the data and AI ecosystem, we need to view the data that drives AI itself as an intended extension of an individual's and community's decisional autonomy.

[10] Nick Couldry and Mejia U (2019), "Data colonialism: rethinking big data's relation to the contemporary subject" Television and New Media. Available at https://eprints.lse.ac.uk/89511/1/Couldry_Data-colonialism_Accepted.pdf.

[11] Financial mail, "How Google and Facebook are the biggest threat to South African news media." Published on November 16, 2017. Available at https://www.businesslive.co.za/fm/features/cover-story/2017-11-16-how-google-and-facebook0-are-the-biggest-threat-to-south-african-news-media/

[12] Billy Perrigo, "Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic." Time Magazine. Published on January 18, 2023. Available at https://time.com/6247678/openai-chatgpt-kenya-workers/

[13] Zuboff, Shoshana. 2019. The Age of Surveillance Capitalism. London, England: Profile Books.

An Afro Feminist Playbook for Regulating AI

This playbook will centre values of agency, human dignity, privacy and non-discrimination. The question of governing AI leads to several questions on regulatory scope, strategy and mode most appropriate for it. In this playbook, we provide guidance on principles which can help policymakers approach regulation, provide civil-society actors with a framework to hold governments and industry to account, and industry bodies to contemplate situations where self and co-regulation are relevant.

Forms of Regulatiton

In essence, 'regulation' involves governing with an intention in complex situations where there are competing interests. Sovereign entities traditionally determine regulation, but market actors are increasingly devising their own frameworks. This is due to the fact that governments lack complete information, expertise, and resources to devise, implement, and enforce regulation for emerging technologies that bring rapid change and uncertainty.

Primary Regulation
Governments use their power, funding, and organisation to enforce laws. They have ways to detect breaches and apply punishments. They have the power to enforce punishments in a way that people respect. The government has many resources to do its job. The bureaucratic structure is what makes it all possible. However, regulating AI solutions in different economies, in Africa, is not easy.

Self Regulation
Jessop defines self-regulation as a system where private actors limit regulatory bodies from the bottom-up. Independent actors work together to develop mutually beneficial projects by sharing resources. Self-regulation can be standardised or voluntary. Standardised self-regulation means industry-wide organisations set rules, standards, and codes for everyone in the industry.
Voluntarism is when a company regulates itself and creates its own code of conduct. Recently, companies like Microsoft, Google, and IBM proposed ethical AI principles. However, these guidelines don't explicitly follow domestic or international law standards. Ethical frameworks may fail to regulate AI solutions because of different goals, professional history and norms, proven methods to translate principles into practice, and legal and professional accountability mechanisms. Relying on ethical AI guidelines as a form of self-regulation may not be enough to regulate the variety of ways AI could be used in public functions and to protect core values and human rights.

Co-regulation
Decentralised regulation recognizes that states are not the only regulators and that regulation is often indirect, multi-faceted, and complex due to power struggles. Co-regulation, also known as

"regulated self-regulation," involves private entities governing their affairs through codes of conduct or rules within a legal framework that links state and non-state regulation. The European Commission identifies four elements of co-regulation, including public policy objectives, a connection between state and non-state regulation, discretionary power for non-state regulation, and state supervision. In a co-regulatory framework, governments and private actors share responsibilities, with the government setting high-level goals and the industry setting standards while still being supervised by the state.

Table 1: Modes of regulation

REGULATORY TYPE	ENFORCEABILITY	RIGIDITY	CREATION	APPLICABILITY
LEGISLATION	**Highest.** Binding law, along with clearly defined sanctions for non-compliance.	**Highest.** Clearly defined standards of municipal law with any ambiguity ideally being resolved by the judiciary.	**Top-down.** Devised by the legislator with optional consultation.	**Lowest common denominator.** Would only prevent directly identifiable harms resulting from AI. Would also require production of adequate evidence and causality.
CO-REGULATION	**Middle.** Decentralised regulatory process may lead to a binding outcome.	**Not unique.** Could be clearly defined or vague depending on the outcome.	**Participatory** with government, civil society and industry meaningfully engage in this process.	May have wide or narrow applicability to actors, situations, and individuals depending on the context.
SELF-REGULATION	**Lowest.** Enforceable at the organisational level but not binding. Reliance on 'soft sanctions' with no clearly defined sanctions for non-compliance.	**Lowest.** Clearly articulated frameworks with greater ambiguity and more scope for manipulation.	**Participatory.** Devised through high-level consultations among industry and civil society but with an absence of government actors.	All AI that is ethical is necessarily legal. However, ethical frameworks have a broader applicability to harms that are outside the rigid confines of the law.

Understanding Impact

Regulatory decisions in this case prioritise safety, security, and human impact. When AI has the potential for direct, adverse, or large-scale human impact, more regulatory intervention is required. According to a Berkman-Klein study, 81% of ethical AI documents emphasise safety and security of AI systems. Therefore, we need to ask if there is a high likelihood or severity of potential adverse human impact of the AI solution. We also need to determine if the likelihood or severity of adverse impact can be reasonably ascertained with existing scientific knowledge.

When considering the impact, it is important to look at both the severity and likelihood of the adverse impact. If the severity is high enough, even low likelihood may warrant greater regulatory scrutiny. If the likelihood or severity of harm cannot be reasonably ascertained, we recommend not implementing the solution until scientific knowledge reaches a stage where it can be reasonably ascertained.

The following table lists possible impact scenarios and regulatory strategies.

OUTCOME	EXPLANATION OF OUTCOME	RECOMMENDED REGULATORY STRATEGY
A) High Likelihood, High Severity	Scenarios where the state is involved in predicting human behaviour (predictive policing/credit-rating/predicting school dropouts) but training data is incomplete and a thorough impact assessment has not been conducted.	Ban or proscribe until underlying issues are solved to reduce likelihood of harm. If likelihood or severity cannot be gauged, then the solution must not be deployed.
B) Low Likelihood, High Severity	Scenarios where training data is robust but individuals relying on use cases (flood prediction, crop price forecasting) may face dire economic consequences if the solution works incorrectly.	State run human rights impact assessment that externally verifies compliance.
C) High Likelihood, Low Severity	Possible in pilot cases where data, methodology, and funding are not yet clear and safeguards have not been appropriately devised, or where AI is not directly impacting civil liberties or socio-economic rights (traffic management).	Strong redressal mechanisms that enable even one impacted individual to receive compensation, particularly if the initial estimation of severity is too low.
D) Low Likelihood, Low Severity	Where data is robust, methodology, troubleshooting, and outreach have been clearly devised, and use cases are not directly impacting civil liberties or socio-economic rights.	Possible regulatory forbearance with strong industry-driven codes for standardisation, evaluation, and redressal if the private sector is involved.

Table 2: Impact thresholds

ii Regulatory Strategy

Below, we articulate clear values that must underpin any regulation of AI systems, and accompanying questions which can help policymakers arrive at regulatory positions. The questions are not intended to be exhaustive, as designing such a model would be informed by local context, state of AI deployment, knowledge of other supporting regulations on data protection, exclusions and discrimination. However, the below table will help policymakers ask the right questions which can lead to a comprehensive regulatory strategy

VALUE	QUESTION
AGENCY	Is adoption of the solution mandatory?
	Does the solution allow for end-user control?
	Is there a vast disparity between the primary user and the impacted party?
EQUALITY, DIGNITY, AND NON-DISCRIMINATION	Is the AI solution modelling or predicting human behaviour?
	Is the AI solution likely to impact women, protected, or at-risk groups?
SAFETY, SECURITY, AND HUMAN IMPACT	Is there a high likelihood or high severity of potential adverse human impact as a result of the AI solution?
	Can the likelihood or severity of adverse impact be reasonably ascertained with existing knowledge?
ACCOUNTABILITY, OVERSIGHT, AND REDRESS	To what extent is the AI solution built with 'human-in-the-loop' supervision prospects?
	Are there reliable means for retrospective adequation?
	Is the private sector partner involved with either the design of the AI solution, its deployment, or both?
PRIVACY AND DATA PROTECTION	Does the AI solution use personal data, even in anonymised form?

To further illustrate how responses to these questions lead to specific regulatory outcomes, we highlight below examples of use of AI in different sectors.

a) Mandatory v. Voluntary

The use of AI in public functions will often lead to mandatory outcomes. For instance, the adoption of predictive policing software will be mandatory for all law enforcement agents if adopted by the state. In such cases, it is essential for regular consultation and feedback from all levels within the police hierarchy, in particular officers who directly engage with victims and defendants on the ground.

On the other hand, the use of AI in agriculture will often be optional for farmers, and pros and cons of adopting the solution should be clearly communicated in an understandable format to the farmer. In such cases, encouraging the evolution of a self-regulatory framework may often be suitable.

In Africa, there has been increasing reliance on use of AI in credit scoring using alternative data. While not mandatory by law, their high adoption will inevitably lead to a lack of other options for consumers. Given the nature of high stakes, it may be worthwhile for the state to explore regulation which clearly regulates the adoption of AI solutions for provision of such financial services.

b) Modelling of Human Behavior

This issue speaks to the nature of data that is driving the development and deployment of AI. The modelling of human behaviour poses a higher case of regulation and governance because supply (data input) and demand (decisions about human) both indicate domains which need to be scrutinised carefully.

By modelling human behaviour, we mean the use of datafication and AI technologies which collect, process, analyse and infer data about human beings. Let us consider the above use-case for modelling of human behaviour. Use of AI in law enforcement will inevitably attempt to model likelihood of criminal intent. In such cases, any deployment must be preceded by an assessment of why modelling human behaviour is proportionate to the objective of reducing crime and also demonstrating why no other reasonable alternatives exist.

Credit scoring decisions or any other kinds of scoring or rating of human beings also clearly model human behaviour and arrive at assessments about a consumer's likelihood of satisfying a desired goal. Such deployments also need mandatory reasoning from the sector clarifying why algorithmic decision making is more accurate than traditional credit-scoring methods, as well as full transparency on data being used and curation methods.

On the other hand, use of AI in agriculture might model weather data and crop patterns, and may not be interested in modelling human behaviour based on analysis of personal data.

c) Equality and Non-discrimination

The use of AI when it models human behaviour is often likely to lead to reinforcement and amplification of social inequities and discriminatory impacts on women and sexual minorities, racially marginalised groups, rural populations, and linguistic, socio-economic and other disadvantaged groups. In most cases, the likely governance responses are often limited to awareness, sensitisation, and the creation of grievance redressal mechanisms targeting vulnerable groups. However, striving for Afro-feminist AI must necessarily require that we move from reactive to proactive governance mechanisms.

These require structural changes such as greater representation of minorities and disadvantaged groups in datasets, more hiring of women and minorities in fields of data science as computer scientists, user experience designers and actors who perform fail-safe functions such as human-in-the-loop agents.

The inclusion of Afro-feminism in AI must also extend to radical changes to the ideas of who gets to call themselves a technologist. Inclusion of social scientists studying technology, inclusivity designers and critical theory scholars so that the issues of race, gender and language can be centred in the discourse. In concrete governance and regulatory terms, these can translate into an expansion of funding criteria to include a wider range of scholars, affirmative action mandates which privilege lived experiences in hiring decisions and data collection practices which are not exploitative but inclusive in nature.

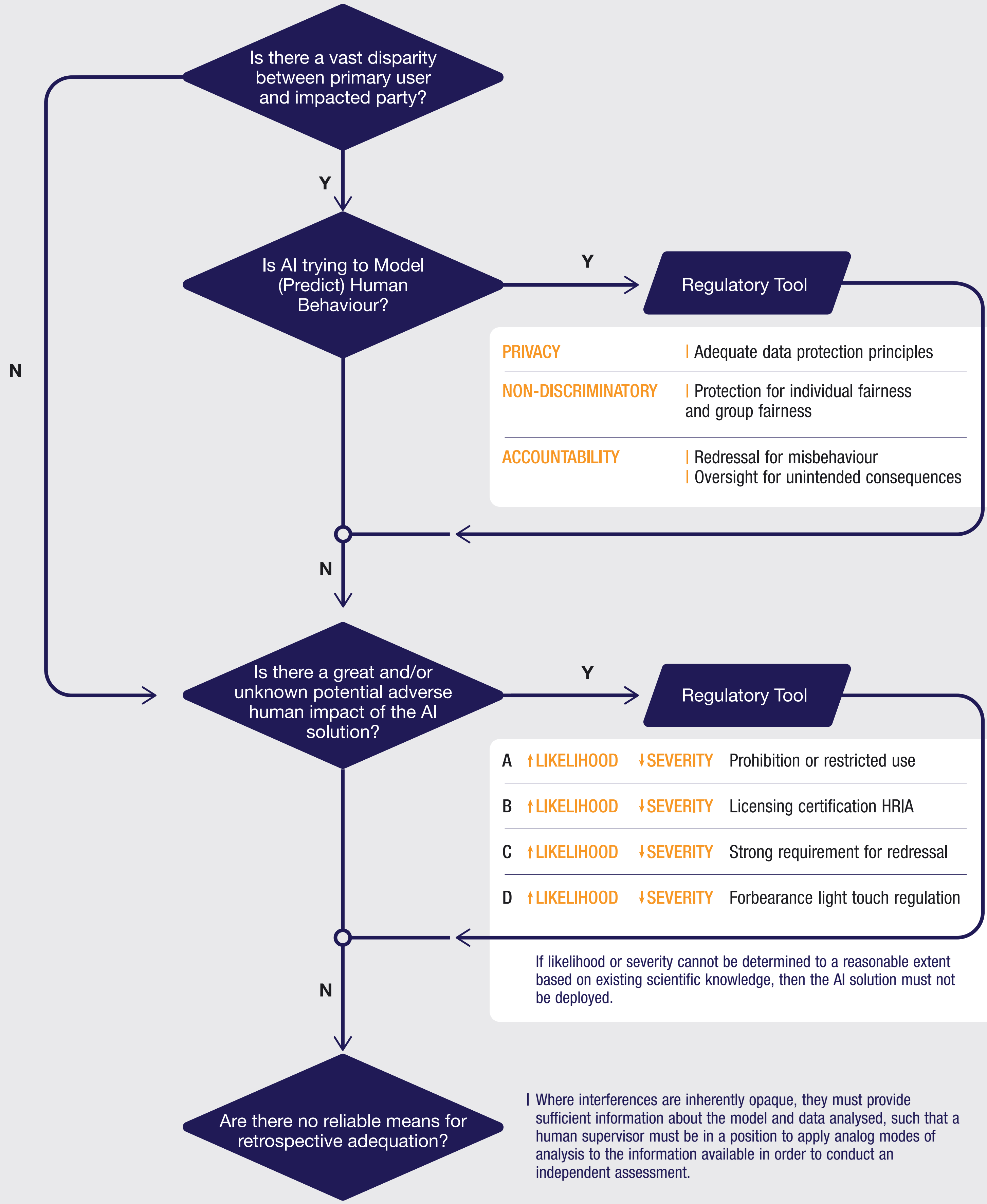
In high-impact sectors such as law enforcement and financial services, the regulatory responses can draw from newer legal innovations such as equality duty. Equality Duty is an innovation in non-discrimination law which is of extreme relevance to implementing AI systems. In simple terms, it requires that all public authorities must, in the exercise of their functions, “have due regard to the need to” eliminate discriminatory conduct. Usually, the duty would refer to discriminatory behaviour which pertains to protected characteristics such as age, disability, gender, race, religion, sex and sexual orientation. More specifically, it imposes an obligation on public bodies to take steps towards the elimination of discrimination towards groups which are disadvantaged or discriminated against on the bases of the above-mentioned characteristics. Further, it requires the public authorities to advance equality of opportunity between persons who share a relevant protected characteristic and persons who do not share it. The intersectional nature of the public sector equality duty incorporates feminist principles which can serve to challenge social inequality and dismantle structures of power.

d) Safety and Human Impact

One of the questions we proposed is whether there is a high likelihood or high severity of potential adverse human impact as a result of the use of the AI solution. For sectors such as predictive policing, there is a reasonable risk of both high likelihood and high severity of adverse impact, unless data collection practices are improved. The obvious regulatory response in such cases is the proscription of AI solution until data curation and analysis is improved and standardised.

For credit scoring also, there is a reasonable risk of both high likelihood and high severity of adverse impact. Along with greater regulation on permitted direct and proxy data points, mandatory pilot projects and standardisation of data curation practices certified by a co-regulatory committee could be suitable potential governance measures. With the increased use of energy-intensive technologies such as large language models, it is likely that the high adverse impact on climate change may be another parameter through which we may judge impact.

Thresholds for Regulation



e) Data Exploitation

Data exists within a complex matrix of relationships between different key stakeholders — individuals, communities, corporations and governments — who may have a beneficial interest in data, and how this interest may be exercised as and against other stakeholders. The need for identification of such inherent beneficial interests in data outside the purview of contractual rights arises from the power structures that have emerged in the current data economy. The exploitative nature of mining personal data creates an imbalance in the benefits accrued by those whose data is utilised for financial gain and those monetizing on having access to personal data. There has been an upswell of discontent, particularly in the Global South with several commentators claiming that excessive focus on consent has skewed the discussion in favour of US based technology corporations who reap monetary dividends from data gathered from Global South citizens, thereby leading to accusations of 'data colonialism'. There is a dire need to redistribute power over data in a manner that is fair and equitable to all actors involved in its generation. In Africa, this specifically must involve drawing from traditional modes of collective bargaining.

Data stewardship models which centre the interests of individuals and community against the two dominant actors— state and corporations can help marginalised communities participate in and reap the benefits of the data economy.

f) Design Principles for Afro Feminist AI

Key design principles which need to be integrated in the development and deployment of AI in Africa are illustrated below.

Participation
Whilst participation is an approach that is often proposed as a design approach for more inclusive models, this often isn't fully achieved because it isn't genuine. This is in the sense that the involvement of all persons across design is typically a myth or pipedream for now with the existence of limited choice for participants to opt in or out. Additionally, accessibility due to cost and time isn't usually considered when looking at participatory design which many times alienates critical voices to the conversation. Lastly, the influence the feedback participants offer is rarely considered which deems this lack of following through not fully participatory.

Inclusivity in Procurement
How AI systems are procured can ultimately lead to more equitable outcomes of these systems for all people. Being inclusive in this process speaks to the diverse procurement teams which are able to conduct a more comprehensive evaluation of the system's harms and benefits, especially including people from the community the system is designed for. In addition to this, AI procurement budgets must be intentional to train more marginalised groups such as women to be able to have the capacity to understand these systems and evaluate them aptly consequently.

Gender Impact Assessments
Gender Impact assessments are a way of critically thinking about how policies, programs and services will meet the different needs of women, men and diverse gender groups.

This is according to the Equality Institute which also proposed four steps to conduct these i.e. define the challenges or issue, understand your context, options analysis and lastly make recommendations. We add an intersectionality approach to conducting these as well engendering this throughout design and adoption as opposed to having a one-off check-off.

<table>
<tr><td colspan="3">Story 1: Travel service chatbot X

A now popular travel service chatbot has been adopted by several companies in the travel and hospitality industry over the last year. This chatbot is typically the primary communication interface between customers and the service providers helping deliver full time and real time solutions to the customers. Unknowingly to the customers however, this chatbot is also fitted with affect AI tools that assess customer’s emotions when engaging with the bot which information is then shared with the service providers as ‘customer insights’. These insights are then used to further nudge customers into acting a certain way in relation to purchase of the service provider’s services.

It is also important to note that the chatbot’s safety and privacy notice does not mention these affect tools to the users, however the service providers are sold the tech differently as an added analytics feature.</td></tr>
<tr><td>Lessons learned

The drive towards quicker and more profit made possible by AI tools can enable opaque practices that are harmful to users.

The service providers leaving out the affect tech in the chatbot’s privacy notice is deliberate.

The creator of the chatbot using unclear terminology such as analytics as opposed to emotion recognition is deliberate as well.</td><td>Harm level

Per say harmful case since emotion recognition and the resultant subliminal behaviour nudging erodes many fundamental rights of the users.</td><td>Possible interventions

The owner of the chatbot ought to ensure transparency of the chatbot’s capabilities from marketing to licensing of its adopters.

Updating the chatbot’s privacy notices so users can consent to using it or not.

Having human oversight along the chatbot as customers engage the service providers.</td></tr>
<tr><td colspan="3">Questions that can be asked:

How can the drive towards more efficiency and profit be balanced out with transparency of AI systems by developers of these tools?

What does intended information hoarding by AI developers or adopters signal?</td></tr>
</table>

Story 2: Mutually beneficial AI

A certain developer has been building a cancer diagnostics and care model. In pursuit of making the tool more robust and therefore useful to a diverse range of people, they require more datasets to train their model more effectively. To this end, the developer has opened up an intervention to access more datasets. Essentially,they are informing the public about the tool they are developing and its benefits once it is more use-ready as a way of getting them to willingly share their data to help improve the tool. Fortunately, enough people are sharing their data which means they have got to appreciate the utility of this model and would not therefore mind sharing their data with the developer.

Lessons learned	Harm level	Interventions
Developing models with tangible utility enhances mutual relations between the developer and prospective users. Transparency greatly fosters reciprocal relations between developers and users.	Per se not harmful since there is no harm in being transparent about the developer's data needs to develop a model that is/will be of great value to the larger public.	Compliance with data protection measures by the developer even when they have sourced the necessary datasets responsibly.

Questions that can be asked:

How can developers realise meaningful reciprocal relations between themselves and their users beyond assumed benefits to users?

How can users of an AI tool be meaningfully engaged in its development?

Story 3: Inclusion as a trap

Non-profit X has designed a program to enrol children in a refugee camp to smart education services. However, whilst this program has the transformational capability to impact these children's lives positively, this non profit also simultaneously is owned by a multinational tech corporation which is taking data from these students and their guardians to build a race detection model for border control much to the ignorance of the camp or country operators. This poses a risk of coming up with a model that is discriminatory and oppressive to such an already marginalised group.

Lessons learned	Harm level	Interventions
Even the most useful models are susceptible to bad actors.	Per say harmful/ Sensitive use cases.	Discontinuity of the program.

Questions that can be asked:

How can the full extent of any philanthropic interventions be understood?

How could such blatant harm be mitigated?

How could such a bad actor be held accountable?

Story 4: Gen Z loss of cultural identity through Tiktok

Tiktok is increasingly the platform of online media consumption by all gen Z's globally. Its recommendation algorithms which are said to personalise videos of one's interests also in many ways go against that with the viral videos or trends (challenges) which are said to be high interest videos for all users. The catastrophe in this is finding similarity in language, tone of expression and at worst behaviour of young persons influenced by these algorithms over time. This gradual erosion of cultural identity based on 'virality' is a problem we are yet to contend with particularly where certain culture's may be overshadowed by others feeding into the overall existential threat of certain cultures extinction versus preservation of others.

Lessons learned	Harm level	Interventions
What goes viral over and over again has an impact on the behaviour of especially younger users.	Per se harmful.	Tiktok ought to find ways to find equitable virality of content from all parts of the globe.

Questions that can be asked:

What are the impacts of consuming content that has no reflection of one's indigeneity over and over again?

Are these impacts desirable for different communities?

How may these possibly be mitigated?