

# Principles on Equality by Design in Algorithmic Decision-Making



EQUAL RIGHTS TRUST



These Principles on Equality by Design in Algorithmic Decision Making were developed by the Equal Rights Trust, in consultation with experts and activists from across the globe. The Principles are endorsed by a group of international equality organisations - the Alliance for Universal Digital Rights, Equality Now, Help Age International, the International Lesbian, Gay, Bisexual, Trans and Intersex Association, Minority Rights Group International and Women leading in AI. The Equal Rights Trust acknowledges and thanks all those who contributed to the development of the Principles and those who have endorsed them.

## **Equality by Design: A framework for securing equality and non-discrimination in the use of algorithmic decision-making systems**

As the use of algorithmic decision-making systems has spread, it has become clear that while these systems have the potential to enable the equal enjoyment of human rights, they can have – and are having – adverse effects, including on the rights to equality and non-discrimination.

**Algorithmic decision-making systems (algorithmic systems)<sup>i</sup> are creating, replicating, and exacerbating patterns of discrimination on a wide range of grounds, in a variety of sectors and in many different areas of life.** Affected communities, civil society organisations, United Nations experts and others have exposed how algorithmic systems are denying or limiting the equal participation of groups including, but not limited to, women; racial, ethnic and religious minorities; persons with disabilities; LGBTI persons; older persons; and young persons.<sup>ii</sup> These systems are having discriminatory impacts in areas ranging from employment to education and from access to services to freedom of movement.

**Because of the way in which algorithmic systems are developed and designed, trained and evaluated, deployed and used, they are frequently discriminatory by default.** Systems which many believe are inherently objective and fair in fact reinforce existing patterns of discrimination, reflect stereotypical assumptions and replicate bias. Discrimination can and does occur at every point in the lifecycle of these technologies – because of the way in which problems are identified or defined; because of the data used to create or train systems; because of the design of algorithms and systems; and because of the manner in which these systems are deployed and used.

**States and businesses have obligations to prevent the discriminatory impacts of algorithmic systems.** These impacts may be unintended, unforeseen, or challenging to identify or understand. However, this does not limit the obligations of States and responsibilities of business to prevent discrimination. International law requires that States do not discriminate in law, policy and practice and that they take effective measures to prevent discrimination by business and other actors, including through the adoption of comprehensive anti-discrimination legislation. Even where a State falls short of these duties, businesses have responsibilities to respect the right to non-discrimination which exists independently of States' obligations. All of these duties apply to the use of algorithmic systems.

**The potential of algorithmic systems to increase equality of access, opportunity and participation is not being explored or exploited.** Algorithmic systems have the potential to contribute to the elimination of discrimination and advancement of equality. These systems could be used to assess and remove barriers which prevent equal participation; to enable or facilitate adaptations to meet specific needs; and to identify and eliminate bias in decision-making. These opportunities are not being explored or seized, despite States' positive obligations to use all appropriate means to ensure the enjoyment of rights on an equal and non-discriminatory basis.

**States and businesses are failing to meet their obligations of equality and non-discrimination in the regulation, development and use of algorithmic systems.** The speed with which these technologies are being developed and the scale on which they operate; the complexity and opacity of the systems themselves; misconceptions about legal obligations and liabilities; and a host of other challenges demand a new approach. States and businesses will only be able to meet their obligations to eliminate discrimination and advance equality if they take a proactive approach to identifying, assessing and addressing the equality impacts of algorithmic systems during their development.

## Principles on Equality by Design in Algorithmic Decision-Making

1. **In order to meet their legal obligations and responsibilities to eliminate discrimination and achieve their ambitions to advance equality, States and businesses must adopt an Equality by Design approach to the development of algorithmic systems.**
2. Equality by Design is an approach to system design which requires and enables the potential equality impacts of an algorithmic system to be identified, assessed and addressed as an integral part of the development process. It is an intentional, proactive approach through which potential equality impacts are identified and addressed on a pre-emptive and precautionary basis. It requires informed consultation and meaningful engagement with those exposed to discrimination and disadvantage to identify potential equality impacts at all stages in the lifecycle of the system, and to design and integrate appropriate measures and adaptations within the system to eliminate discriminatory impacts and maximise opportunities to advance equality.
3. The objective of Equality by Design is to ensure that the use of algorithmic systems does not result in discrimination and that potential positive equality impacts are maximised. This requires that potential discriminatory impacts are identified and eliminated and that the potential of a system to remove barriers, eliminate bias, increase diversity and advance equality of participation is assessed and maximised. Through the process, equality is consciously designed into the process of development, deployment and use of algorithmic systems.
4. Equality by Design is an essential means for businesses to ensure compliance with their obligations under national laws and their responsibilities under the UN Guiding Principles on Business and Human Rights to ensure non-discrimination in their operations. It allows those involved in the development and design of these systems to identify and eliminate discriminatory impacts before the system is adopted or deployed. It acts as a critical safeguard for those involved in the procurement and application of these systems, enabling them to ensure that the use of a system will not result in discrimination. It allows all those involved in developing and using these systems to identify and eliminate potential discriminatory impacts and to consider how they can be used to remove barriers to participation and advance equality.
5. Equality by Design is a necessary but not sufficient condition for States to ensure compliance with their international law obligations to prevent and eliminate discrimination and ensure equal enjoyment of human rights. As an essential part of their measures to prevent discrimination and advance equality, States must impose and enforce obligations of Equality by Design on those involved in the development and use of algorithmic systems. The adoption of Equality by Design should be part of a comprehensive, system-wide approach which includes the adoption and enforcement of comprehensive equality legislation, access to justice and remedial mechanisms, and positive measures to advance equality.

### A Legal obligations and responsibilities

6. Equality by Design is a process which enables states and businesses to meet their obligations, responsibilities and ambitions to prevent and eliminate discrimination and advance equality of participation. Imposing and enforcing obligations of Equality by Design is a necessary, but not sufficient, condition for states to ensure that they are respecting, protecting and fulfilling the right to non-discrimination in the regulation and oversight of algorithmic systems. Adopting an

Equality by Design approach is an essential means for businesses – both developers and users – to ensure that they do not violate their legal obligations of non-discrimination.

7. The rights to non-discrimination and equality are “among the most fundamental principles and rights of international human rights law.”<sup>iii</sup> Almost every State in the world has accepted international law obligations to ensure the equal exercise and enjoyment of human rights without discrimination, and to ensure equality before the law and equal protection to all.<sup>iv</sup> These obligations are immediate and cross-cutting and they underpin all civil, political, economic, social and cultural rights.<sup>v</sup> The rights are guaranteed to all persons, citizen and non-citizen,<sup>vi</sup> irrespective of status, identity, characteristics or beliefs.<sup>vii</sup>
8. At a minimum, the right to non-discrimination requires that States do not discriminate, either in action or through failure to act, in law, policy or practice.<sup>viii</sup> This obligation applies to impacts which are both direct and indirect and those which are intended and unintended.<sup>ix</sup> Meeting this obligation requires States to identify and eliminate discriminatory impacts as an integral part of their law, policy and decision-making processes.
9. Alongside their preventive duty, states have a positive obligation to “pursue by all appropriate means and without delay a policy of eliminating” discrimination and to ensure the enjoyment of rights on an equal and non-discriminatory basis.<sup>x</sup> As with the duty to prevent discrimination, this obligation necessitates proactive measures. To fulfil this obligation, States must take a range of actions, including, but not limited to, measures to: identify and remove barriers which prevent equal participation; accommodate difference; combat prejudice and stereotype; and advance and promote equality.<sup>xi</sup>
10. States’ obligations to eliminate discrimination include prohibiting and preventing discrimination by private actors.<sup>xii</sup> To fulfil this obligation, States should adopt a range of measures, including adopting and enforcing laws.<sup>xiii</sup> Specifically, States must adopt dedicated, comprehensive anti-discrimination legislation.<sup>xiv</sup> Such legislation should prohibit all forms of discrimination, on the basis of all grounds recognised at international law, in both public and private sectors and in all areas of life regulated by law; provide means for rights-holders to challenge violations, access justice and secure remedy and sanction; and impose positive obligations on both state and private actors to prevent discrimination, redress disadvantage and advance and promote equality.<sup>xv</sup>
11. States’ anti-discrimination laws should impose obligations of non-discrimination on private actors, with effective, proportionate and dissuasive sanctions where these obligations are violated.<sup>xvi</sup> Irrespective of whether a State adopts or enforces national laws prohibiting discrimination by private actors, businesses have responsibilities to respect human rights which exists independently of States’ legal obligations.<sup>xvii</sup> This responsibility requires that businesses anticipate and address adverse human rights impacts,<sup>xviii</sup> including discrimination.
12. Human rights, including the rights to equality and non-discrimination, apply in the digital environment as well as the offline environment: State and business obligations to ensure the enjoyment of these rights apply and must be complied with in all contexts, including in the development, design, deployment and use of algorithmic decision-making systems.<sup>xix</sup> As the *Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems* underlines, when developing, using or regulating the use of machine learning systems,

“governments have obligations and private sector actors have responsibilities to proactively prevent discrimination in order to comply with existing human rights law and standards.”<sup>xx</sup>

13. States’ duty to ensure the enjoyment of the right to non-discrimination and to ensure the equal enjoyment of human rights imposes an immediate obligation to eliminate and prevent discrimination and its consequences. It also imposes an obligation to use all appropriate measures to ensure the equal enjoyment of human rights. In the context of algorithmic systems, these obligations can only be fulfilled by the imposition of a duty of Equality by Design on those involved in the development, design, deployment and use of these systems.
14. Businesses have obligations and responsibilities to ensure that they do not discriminate. These obligations apply to acts and omissions which are both intentional and unintentional, foreseen and unforeseen, and to decisions and actions which are taken by employees, agents and proxies. In order to ensure that they do not violate their obligations, businesses need to pre-emptively identify and eliminate potential discriminatory impacts of their decisions and actions. In the context of algorithmic systems, this requires the adoption and integration of an Equality by Design approach in the development, design, deployment and use of these systems.

## **B Identification and assessment of equality impacts**

15. Equality by Design requires those involved in the development, design and deployment of algorithmic systems to intentionally and proactively identify, assess and address potential equality impacts. It requires consideration of both potential negative and positive impacts, arising at any point in the lifecycle of the system, from inception through to use.
16. **Equality impact** refers to any impact on the enjoyment of the rights to equality and non-discrimination which may be caused by, or result from, the use of an algorithmic system. Equality impacts can be both positive and negative.
  - a. **Negative equality impacts** (discriminatory impacts) are those which cause, perpetuate or exacerbate any form of discrimination on the basis of any one or any combination of grounds. Negative equality impacts include impacts which are direct and indirect, intended and unintended and foreseen and unforeseen. Negative, discriminatory impacts can arise through omission, including the failure to accommodate different needs.
  - b. **Positive equality impacts** are those which contribute to challenging or combating discrimination, removing barriers to participation, actively accommodating difference, increasing diversity, countering or overcoming bias or stereotype, or otherwise advancing or enabling equal access or participation.
17. The process of identifying and assessing potential equality impacts should be undertaken in close consultation with groups exposed to discrimination and inequality on different grounds, following the principles on informed consultation and meaningful engagement set out below.

### **B1 Negative equality impacts**

18. The developers and users of algorithmic systems should undertake an evidence-based, participatory and consultative process to identify whether and how the development, design, deployment or use of such system could result in discrimination.

19. **Discrimination** is any unfavourable treatment or disproportionate impact that arises in connection with one or more protected characteristics or grounds of discrimination.<sup>xxi</sup> Discrimination takes different forms: direct discrimination, indirect discrimination, harassment, failure to make reasonable accommodation and segregation.
20. Discrimination can arise on any one or any combination of the wide range of **grounds** recognised under international law, including – but not limited to: age; descent, including caste; disability; gender expression or gender identity; health status; indigenous origin; political or other opinion; race or ethnicity; religion or belief; sex or gender; sexual orientation or other status.<sup>xxii</sup> The inclusion of “other status” allows for the recognition of grounds which have not yet been recognised but which are analogous to those explicitly listed in extant legal instruments. Discrimination can occur on the basis of the perception that a person has a characteristic or status related to a ground, due to the association of an individual with a ground, and due to the interaction or intersection of grounds.
21. Discrimination takes different forms:<sup>xxiii</sup>
- a. **Direct discrimination** occurs where a person is treated less favourably or subjected to a detriment because of one or more protected characteristics.
  - b. **Indirect discrimination** occurs where a provision, criterion or practice has or would have a disproportionate negative impact on persons having a status or a characteristic associated with one or more grounds of discrimination.
  - c. **Denial of reasonable accommodation** occurs where there is a failure to make necessary and appropriate modifications or adjustments or support to ensure the participation, on an equal basis with others, of persons sharing a particular protected characteristic or characteristics. Denial of reasonable accommodation will occur *inter alia* where there is no option or possibility for a person to request a modification to enable their equal participation.
  - d. **Harassment** occurs where unwanted conduct has the purpose or effect of creating an intimidating, degrading, humiliating or offensive environment in connection with a protected characteristic or characteristics.
  - e. **Segregation** occurs where persons sharing a particular protected characteristic are, without their full, free and informed consent, separated and provided different access to institutions, goods, services and rights.
22. Equality by Design requires that consideration is given to the potential of an algorithmic system to result in any form of discrimination on any one or any combination of grounds in any area of life regulated by law. The process should consider potentially discriminatory impacts arising at any point in the lifecycle of the system, from inception through to operation.
23. It is the responsibility of the developers and users of algorithmic systems to ensure that the operation of these systems does not result in discrimination. Ignorance of potential discriminatory impacts or lack of intent do not limit this responsibility. Accordingly, the Equality by Design process should consider any direct or indirect, intended or unintended, actual or

potential discrimination which may be caused by, or consequences following from, a decision of to design or deploy an algorithmic system.

## **B2 Positive equality impacts**

24. The developers and users of algorithmic systems should undertake an evidence-based and consultative process to identify whether and how the use of such a system could result in potential positive equality impacts.
25. Positive equality impacts include, but are not limited to:
  - a. **Removal of barriers:** Development of processes which overcome or remove spatial, linguistic, cultural, financial or other barriers which prevent equal participation in any area of life.
  - b. **Elimination of bias:** Identification and elimination of stereotype or bias in data or systems used in existing human or algorithmic decision-making processes.
  - c. **Replacement of discriminatory practices:** Replacement of directly or indirectly discriminatory rules, policies or practices in existing human or algorithmic decision-making processes.
  - d. **Promotion of equal participation:** Creation of systems which promote equality of access, opportunity and participation by underrepresented groups in any area of life.
26. Equality by design requires that consideration is given to the full range of ways in which a system has the potential to achieve a positive equality impact or promote or accelerate equal participation. This includes consideration both of potential positive equality impacts through the planned and intended development, design, deployment and use of the system, and the potential for adaptation or modification at any point in the lifecycle to secure or increase a positive equality impact.

## **C Addressing negative equality impacts**

27. Equality by design requires that where potential discriminatory impacts are identified, they are addressed. As a first step, developers and users of algorithmic systems should consider whether and how the system could be adapted or modified to eliminate its discriminatory effects. Where adaptation or modification is not possible, consideration should be given to whether the discriminatory impact meets the requirements under international law for justification, as set out below.
28. Where negative impacts cannot be eliminated through adaptation, nor justified under law, the development or deployment of the system should be discontinued. The deployment and use of an algorithmic system where potential discriminatory impacts have been identified but not eliminated nor justified under law will be a violation of the right to non-discrimination.
29. The process of assessing and addressing potential discriminatory impacts should be undertaken in close consultation with groups exposed to discrimination and inequality on different grounds, following the principles on informed consultation and meaningful engagement set out below,



## C1 Adaptation

30. Where actual or potential discriminatory impacts of an algorithmic system are identified, an assessment should be undertaken to determine whether the system, its deployment or use can be adapted or modified to eliminate the discriminatory effect. Adaptation is an essential element of the Equality by Design process, enabling those developing or using the system to mitigate and eliminate discriminatory impacts, and acting as a guardrail to ensure compliance with obligations and responsibilities of non-discrimination.
31. Assessment of potential adaptations should include consideration of every stage in the lifecycle of the system, assessing, *inter alia* potential modifications to the definition of the problem which the system is developed to address, the data on which the system is trained or which it uses, the design of the algorithm, the way in which the system may be deployed and used and any other aspect of the design or use of the system.
32. In order to discharge obligations of reasonable accommodation, systems should be adapted to ensure that where persons with one or more grounds of discrimination may require adjustment, adaptation or replacement of an existing or planned system in order to participate on an equal basis, there is a mechanism through which a request for such accommodation can be made. Failure to provide an effective mechanism for the request and consideration of accommodation – including through providing access to a human decision-maker – will constitute a failure to make reasonable accommodation.
33. Equality by Design requires that **where a system can be adapted or its deployment or use can be modified to eliminate discrimination, the relevant adjustment must be made.**

## C2 Justification

34. Where an identified discriminatory impact cannot be eliminated through adaptation to the system or modification to its planned deployment and use, consideration should be given to whether the discriminatory impacts can be justified in line with international legal standards. Equality by Design requires that actual or potential discriminatory impacts are assessed against clear criteria governing the potential for justification as established by international law.
35. An otherwise discriminatory impact of an algorithmic system will be justified where (a) it pursues a **legitimate aim**; (b) through means which are **appropriate, necessary and proportionate**.<sup>xxiv</sup>
36. A legitimate aim may never be justified by reference to discriminatory stereotypes. Direct discrimination and segregation may only be justified exceptionally, on the basis of strictly defined criteria. Harassment cannot be justified.
37. A system which results in indirect discrimination may serve a legitimate aim and may in principle be justified. However, the means employed to achieve this aim must be appropriate, necessary and proportionate.
38. Where an algorithmic system can be adapted or modified to eliminate or reduce a discriminatory impact, the means employed by that system will not be appropriate and necessary. Failure to make the necessary adjustments in such cases will constitute discrimination.



39. Where adaptation is not possible, all actual and potential discriminatory impacts arising throughout a system's lifecycle must be subjected to the justification test. This should include consideration of the wider social context and how the system will interact with its environment. A system that is permitted in one context may be impermissible in another, where the underlying social conditions are different.

### **C3 Implementation**

40. Decisions on whether and how to proceed with the development, design, deployment or use of an algorithmic system should be determined by the outcomes of the identification, assessment, adaptation and justification stages of the Equality by Design process, as follows:
- a. Where the process identifies no actual or potential discriminatory impacts, the system may be **permitted**, and the development, design, deployment and use of the system can proceed.
  - b. Where the process identifies discriminatory impacts, but these can be addressed through adaptation or modification, the system must be **modified or adapted** through adjustments to the development, design and deployment process, prior to use.
  - c. Where the process identifies discriminatory impacts but these pursue a legitimate aim through means which are appropriate, necessary and proportionate, the system may be **permitted**, and the development, design, deployment and use of the system can proceed.
  - d. Where the process identifies discriminatory impacts which cannot be addressed through adaptation or modification, and which cannot be justified by reference to the criteria set out above, the development, design, deployment and use of the system must be **halted** and the further development or use of the system must not proceed.
41. In cases where the development, design, deployment or use of a system is halted, the system or the particular use of the system should be prohibited unless and until sufficient evidence is provided that any actual or potentially discriminatory impacts can be eliminated.
42. There is growing evidence that certain uses of algorithmic systems are inherently discriminatory based on one or more protected characteristics in ways which cannot be justified and cannot be addressed through adaptation or modification. Such uses should be **prohibited** unless and until sufficient evidence is provided that any actual or potentially discriminatory impacts can be eliminated.
43. In all cases where the development, design, deployment or use of a system is permitted, a plan for ongoing monitoring and periodic review should be adopted and implemented, as set out below.

### **C4 Monitoring and review**

44. Equality by Design requires that those developing and using algorithmic systems develop and implement a plan for continuous monitoring and cyclical review of potential discriminatory impacts as part of the system design and deployment process.
45. Continuous monitoring should form an integral part of the maintenance and operation of any algorithmic system. Mechanisms should be designed into systems to gather data on system access

and use, and to track outcomes and impacts for groups sharing protected characteristics. Such mechanisms should be used to identify and investigate disproportionalities in access, use, outcomes and impacts. Measures and metrics for identifying and monitoring potential discriminatory impacts should be designed into the system. Monitoring systems should provide for pre-emptive and precautionary “early warning” of potential discriminatory impacts.

46. Cyclical review should be designed into the procedures for deployment and use. Proactive review should be undertaken on a periodic basis, with the frequency determined through risk assessment. Proactive review should also be undertaken in advance of each stage in the lifecycle of a system or at any point where the system is adapted, adjusted or re-deployed. Reactive review should be undertaken in response to complaints, concerns or risks identified.
47. Monitoring and review plans should be informed by a risk assessment of the system. Such risk assessment should be informed by the outcomes of the identification, assessment, adaptation and justification stages of the Equality by Design process, and subject to consultation with those exposed to discrimination and disadvantage, following the principles on consultation and engagement below. Systems or elements of systems which have been subjected to adaptation to eliminate discriminatory impacts and those where discriminatory impacts have been assessed as justified should be identified as high risk. While levels of risk should inform the frequency, scale and nature of monitoring and review plans, they should not be determinative: monitoring and review procedures should be designed into all systems.
48. Where monitoring or review identifies disproportionate or potentially disproportionate impacts, these should be subject to investigation, following the process set out above. Where monitoring and review identifies discriminatory impacts, the use of the system should be halted until such impacts can be assessed and addressed.

## **D Addressing positive equality impacts**

49. Equality by Design requires that where potential positive equality impacts of an algorithmic system are identified, these are addressed. At a minimum, this requires those involved in the development and use of these systems to have due regard to how to achieve the greatest possible positive equality impact, without imposing a disproportionate or undue burden.
50. Positive equality impacts may arise both as a result of the intended development and use of the system, and as a result of adaptation or modification at any point in the lifecycle.
51. Where the Equality by Design procedure identifies that the planned and intended development, design, deployment or use of an algorithmic system will have a positive equality impact, steps should be taken to adopt and implement the system with a view to having the most immediate and greatest positive equality impact possible, within available resources.
52. Where the Equality by Design procedure identifies that an algorithmic system or its planned use could be adapted or modified to achieve a positive equality impact, due regard should be had to how such adaptations can be made. Where possible, within available resources, the system or its planned use should be adapted or modified with a view to having the most immediate and greatest positive equality impact possible.

53. The identification of an actual or potential positive equality impact in the development, design, deployment or use of an algorithmic system does not negate or limit the obligation on the developer or user of such system to fully identify, assess and address negative or discriminatory impacts.
54. Where an Equality by Design procedure identifies a potential positive equality impact, the responsible entity has a duty of disclosure. At a minimum, developers and users of such systems should inform relevant state entities.
55. It is the responsibility of the state to take all appropriate measures to eliminate discrimination and advance equality of participation. Pursuant to this obligation, where an algorithmic system has a potential positive equality impact, the state should have due regard to how to maximise its use, including through investment. Where the achievement of a positive equality impact requires adaptation or would otherwise impose a disproportionate or undue burden on the developer or user, the state should have due regard to how to support and enable the achievement of such impact.

## **E Process and principles**

56. Equality by Design should be undertaken with the purpose of eliminating discrimination, with a concern to address and prevent disadvantage, and an ambition to realise equality outcomes.
57. The process for identifying, assessing and addressing potential equality impacts must adhere to the following principles:
  - a. **Proactive:** Assessment of equality impacts should be undertaken proactively, intentionally and purposefully as an integral part of any decision to develop, design or deploy an algorithmic system. It is the responsibility of the developer and user of the system to undertake the assessment; it should not be reactive or responsive to a complaint, though complaints should also trigger an assessment or reassessment.
  - b. **Pre-emptive:** Assessment of actual or potential equality impacts must be undertaken on a pre-emptive basis, meaning before any decision to develop, design, deploy or a system is made, and at a point where adjustments can occur.
  - c. **Precautionary:** Assessment of impacts must be precautionary, in that where potential discriminatory impacts are identified, steps should be taken to eliminate these impacts, even where uncertainty remains.
  - d. **Participatory:** Assessment of impacts should be undertaken by providing for informed consultation and meaningful engagement with groups exposed to discrimination or disadvantage. The assessment must be substantive, and the results must drive change, in the form of adaptations or other measures to address the equality impacts identified.

## **F Consultation and engagement**

58. An Equality by Design approach requires the systematic identification and assessment of equality impacts through informed consultation and meaningful engagement with those exposed to

discrimination or disadvantage. In line with an Equality by Design approach, engagement with these groups and individuals should be proactive, pre-emptive, precautionary and participatory. At the same time, the approach should be open, transparent and inclusive, allowing all interested actors and potentially impacted groups to self-identify and participate.

59. Any engagement should be preceded by research to identify and assess relevant data on the experiences of groups exposed to discrimination. Resources consulted may include the research of relevant academic, civil society and inter-governmental bodies, including the results of any previous engagement and consultation with potentially affected groups. Research should be systematic and comprehensive. The purpose of research should be to identify groups exposed to discrimination and gather evidence relating to the actual or potential impacts arising from the algorithmic system.
60. Following the completion of research, those responsible for the Equality by Design process should seek to engage with groups exposed to discrimination or disadvantage, both directly and through their representative organisations. Engagement should be open and inclusive, and should not be limited to representative bodies only.
61. The **identification** of groups exposed to discrimination and disadvantage, and any representative organisations, should observe the following key principles:
  - a. **Intentional:** Identification of potentially affected groups should be systematic, methodical, comprehensive, data-driven and consultation-led. Groups should be identified based on preliminary research, undertaken in line with this guidance, and complemented by meaningful consultation with any representative organisations. Assumptions should be avoided; the relevance and impact of an algorithmic system on all groups exposed to discrimination and disadvantage should be considered in a systematic manner.
  - b. **Inclusive:** Groups which may be affected by an algorithmic system should be understood as any group exposed to discrimination on any one, or any combination of, the grounds recognised by international law, and any other grounds which may be the basis of significant disadvantage arising from an algorithmic system. It is essential to identify those groups which are less visible or whose interaction with the system may not be immediately clear.
  - c. **Intersectional:** In identifying groups which may be affected by an algorithmic system particular consideration should be given to those at risk of multiple or intersectional discrimination as a result of the intersection of more than one ground of discrimination. Consideration should be given to the fact that the operation of algorithmic systems can give rise to new forms of intersectionality.
62. When identifying individuals, groups and representative organisations, those leading the process should be mindful of diversity within groups and characteristics, including different views and opinions within the group, and the different needs of particular members of any group. Consideration should be given to the fact that not all affected groups have representative organisations, and that no single organisation can be considered as representative of all those sharing a particular characteristic. Those leading the process should consistently question who may be missing or excluded from the process and take steps to proactively ensure their participation.

63. While ensuring that they identify and engage with those exposed to discrimination and disadvantage in a proactive and systematic manner, those leading the process should ensure that it is open and transparent at all stages. It is essential to enable groups and individuals with a potential interest in the system to self-identify and participate in an open, consultative and inclusive process.
64. The approach to **engaging** with potentially affected groups and any representative organisations should adhere to the following core principles:
  - a. **Accessible:** Any barriers to participation should be identified and removed. This includes physical, communicative, linguistic, geographical, financial and other barriers.
  - b. **Sensitive:** Some groups may be discouraged from engaging due to social or cultural concerns. Decision-makers should work to identify any social and cultural barriers and take proactive measures to ensure that the means of engagement are appropriate.
  - c. **Safe:** Some groups may be at risk of violence or repression as a result of their status, identity or beliefs. Particular care should be taken to identify these groups and the risks which they face, and to design means of engagement which are suitable to ensure their safety and security.
65. Engagement should be undertaken with attention to the resource and capacity limitations of those engaged. Those leading the process should ensure their means of engagement do not place any undue burden, including financial or evidentiary burdens, on those they engage with. Assessors have a duty to enable the participation of those they seek to engage with by, for example, meeting any financial costs associated with their participation.

## **G Responsibility and oversight**

66. It is the responsibility of the developers and users of algorithmic systems to ensure that the operation of these systems does not result in discrimination. Accordingly, it is the responsibility of these entities to undertake the Equality by Design process. The duty to identify, assess and address equality impacts remains with the decision-maker and cannot be delegated.
67. The assessment of potential impacts must be participatory and consultative; while the assessor is responsible for the process, the assessment itself should be based upon and derived from the inputs provided by those consulted. Consultation and engagement should follow the principles of intentionality, inclusivity and intersectionality, accessibility, sensitivity, and safety, outlined above. The process should be open and transparent.
68. The entity undertaking the Equality by Design process should keep evidence of who is engaged through the process, how they were engaged and the rationale for why certain groups were engaged; the means of engagement and the inputs provided; and the outcomes of each stage of the process. Records should be kept of any adjustments or modifications made to the system or its use; any consideration of the justification of potential discriminatory impacts; and the outcomes of the risk assessment, monitoring and review procedure. These records should evidence that the entity has fulfilled its duty to identify, assess and address any actual or potential equality impacts.

69. The outcomes of the Equality by Design process should be subject to external, independent review by an expert with sufficient knowledge of the algorithmic system in question to undertake a meaningful evaluation of the accuracy and completeness of the assessment. The evaluator should be capable of effectively critiquing the process and outcomes.
70. Those undertaking Equality by Design should ensure that they report back to the groups and representative organisations consulted on outcomes of the process. The assessor should establish mechanisms for those consulted to raise a concern or make a complaint. Any complaints should trigger a re-assessment.

## **H Compliance, sanction and remedy**

71. In order to meet their obligations under international human rights law, states must impose legal duties of Equality by Design on those involved in the development, design, deployment and use of algorithmic systems. These duties must be enforceable. To ensure enforceability, the state should develop and implement an effective system of compliance, sanction and remedy.
72. States should establish suitable monitoring mechanisms to ensure that developers and users are compliant with their Equality by Design obligations. Such mechanisms should require periodic reporting by those with Equality by Design duties on procedures, assessments and outcomes. They should also allow, at minimum, for periodic, external and independent review and assessment of Equality by Design systems, procedures and assessments. Review should be both quantitative and qualitative, allowing for assessment of the quality, comprehensiveness and effectiveness of the procedure.
73. States should ensure the existence of appropriate institutions to ensure compliance with Equality by Design obligations. States may establish new monitoring and enforcement institutions or mandate existing institutions to discharge these functions. States have discretion as to the structure and form of such institutions, but must ensure that they are independent, effective and accessible. These bodies must be afforded the resources and given the functions and powers necessary to fully and effectively discharge their mandate.
74. Monitoring and enforcement institutions should be empowered to apply sanctions for failure to comply with the Equality by Design duty. Sanctions should apply to any breach of the duty, including procedure, assessment and reporting, irrespective of whether a failure results in discriminatory impact. Such sanctions must be effective, dissuasive and proportionate.
75. The objectives of Equality by Design include the proactive, pre-emptive and precautionary identification and elimination of discriminatory impacts of an algorithmic system. Nevertheless, where discriminatory impacts are not eliminated prior to use, the State should ensure effective remedy for victims of any discriminatory impact. Remedy includes, but is not limited to: sanctions for those found responsible for discrimination; reparations, including recognition, compensation and restitution for victims of discrimination; and institutional and societal measures designed to address the social causes and consequences of discrimination and prevent repetition.

## **I Immediate obligations**

76. States have obligations under international law to respect, protect and fulfil the right to non-discrimination. Compliance with these obligations requires, *inter alia*, that they adopt, enforce and

implement comprehensive anti-discrimination legislation. Such laws should impose obligations of non-discrimination on actors in both the public and private sector.

77. In addition to states' obligations to enact, enforce and implement comprehensive anti-discrimination law, and obligations imposed on businesses under such laws, the adoption of an Equality by Design approach to the development and use of algorithmic systems is essential if states and businesses are to meet their obligations to prevent discrimination and advance equality.

78. **States should:**

- a. Establish an enforceable, legal duty of Equality by Design applicable to anyone involved in the development, design, deployment or use of algorithmic systems;
- b. Develop and disseminate guidance on the application of the Equality by Design approach, reflecting these Principles, and foster the development of good practice in this regard;
- c. Establish and implement an effective system to ensure compliance with the Equality by Design, duty including periodic reporting and review; sanctions for failures to follow procedures; and effective remedy in cases of discriminatory impacts;
- d. Ensure that where Equality by Design procedures identify potential positive equality impacts, due regard is had to how to maximise such positive impacts, including through adaptation or investment;
- e. Ensure that where those fulfilling public functions engage in the development, design, deployment or use of algorithmic systems, they integrate and apply an Equality by Design approach at all stages in the system lifecycle.

79. **Businesses should:**

- a. When developing and designing algorithmic systems, adopt, integrate and implement an Equality by Design approach, compliant with these Principles.
- b. When commissioning, deploying or using algorithmic systems, ensure that they require evidence of the satisfactory completion of an Equality by Design procedure and ensure compliance with the Principles established here in the use and implementation of any system.



## Notes

---

<sup>i</sup> We use the term algorithmic decision-making system (algorithmic system) to refer to any system or process through which an automated system is used as part of a decision-making process. Algorithmic system is a broad term to describe any system which uses data and statistical analyses to make decisions or propose solutions. Algorithmic systems include a broad range of tools, systems and processes including both simple automated systems and different types of artificial intelligence (AI), including rule-based AI and machine learning. The level of human involvement in algorithmic systems can vary significantly. Some algorithmic systems simply gather data and present it in a readable format to help a human operator come to their own decision, while others produce fully automated decisions, with little or no human intervention. As we use the term, algorithmic systems include both those systems which are fully automated as well as those where humans are in the decision-making loop. This report presents case studies where a variety of different types of algorithmic systems, with varying levels of human intervention, have been used.

<sup>ii</sup> For examples of documented patterns of discrimination arising from the use of algorithmic decision-making systems, see: Equal Rights Trust, *Discriminatory by Default? Case studies on the discriminatory impacts of algorithmic decision making systems*, 2023

<sup>iii</sup> Committee on the Rights of Persons with Disabilities (CRPD), General comment No. 6 (2018), para. 4.

<sup>iv</sup> Article 2 (1) of the International Covenant on Civil and Political Rights (ICCPR) and Article 2 (2) of the International Covenant on Economic, Social and Cultural Rights (ICESCR) each require States to respect and to ensure those rights guaranteed by the Covenants without discrimination. Article 26 of the ICCPR further provides that “all persons are equal before the law and are entitled without any discrimination to the equal protection of the law.” In its guidance, the Human Rights Committee has noted that this provision establishes an “autonomous right” to non-discrimination that applies in all areas of life “regulated and protected by public authorities.” (Human Rights Committee (CCPR), General Comment No. 18, 1989, para. 12). In total, 173 States are party to the ICCPR, whilst 171 States are party to the ICESCR. For the ratification status of the core United Nations human rights treaties, see: <https://indicators.ohchr.org/>.

<sup>v</sup> See: Committee on Economic, Social and Cultural Rights (CESCR), General comment No. 20 (2009), para. 7; and CRPD, General comment No. 6 (2018), para. 12.

<sup>vi</sup> Both the Human Rights Committee and the Committee on Economic, Social and Cultural Rights have affirmed in their general comments that the right to non-discrimination should be guaranteed to “all persons” irrespective of their citizenship status. For further information, see: CCPR, General comment No. 15 (1986), para. 2; CESCR, General comment No. 20 (2009), para. 30.

<sup>vii</sup> Discrimination is prohibited on the basis of particular grounds. Some of these grounds are expressly listed within the text of a convention, whilst others have come to be recognised as protected forms of “other status”. A wide range of grounds have been recognised under international law, including (non-exhaustively) age; birth; civil, family or carer status; colour; descent, including caste; disability; economic status; ethnicity; gender expression; gender identity; genetic or other predisposition towards illness; health status; indigenous origin; language; marital status; maternity or paternity status; migrant status; minority status; national origin; nationality; place of residence; political or other opinion; pregnancy; property; race; refugee or asylum status; religion or belief; sex; sex characteristics; sexual orientation; social origin; social situation; or any other status. For further information, see broadly: CESCR, General comment No. 20 (2009); and CRPD, General comment No. 6 (2018), para. 21. See also: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 2, Section I.A.1.

<sup>viii</sup> By ratifying a human rights treaty, States agree to respect, protect and fulfil the rights contained therein. As part of States' obligations to respect, they must refrain from discrimination, and amend or remove any laws, policies and practices that produce discriminatory impacts. See, illustratively, International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), Article 2 (1) (a) and (c); Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), Article 2 (d), (f) and (g); Convention on the Rights of Persons with Disabilities, Article 4 (1) (b) and (d).

<sup>ix</sup> Human rights treaty bodies have consistently recognised both direct and indirect discrimination as forms of prohibited conduct falling within the scope of the right to non-discrimination. In its interpretation of article 2 (2) of the ICESCR, for instance, the CESCR has defined indirect discrimination as “laws, policies or practices which appear neutral at face value, but have a disproportionate impact ... as distinguished by prohibited grounds of discrimination”. The Committee on the Elimination of Racial Discrimination (CERD), the CRPD and the CCPR have all held that motive or intent is not necessary to the determination of whether discrimination has occurred. For further information, see, for instance: CRPD, General comment No. 6 (2018), para. 18 (a); CERD, *V.S. v. Slovakia*

---

(CERD/C/88/D/56/2014), para. 7.4; CRPD, General comment No. 6 (2018), para. 18 (a); and CCPR, *Simunek et al. v. the Czech Republic* (CCPR/C/54/D/516/1992), para. 11.7.

<sup>x</sup> See, for example, ICCPR, art. 2; ICESCR, art. 2; ICERD, art. 2; CEDAW, art. 2; Convention on the Rights of Persons with Disabilities, art. 4; and Convention on the Rights of the Child, art. 2. See also: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, p. 115.

<sup>xi</sup> See: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 2, Sections I.C and 5.

<sup>xii</sup> See, for example: CRPD, General comment No. 3 (2016), para. 18; UN Office of the High Commissioner of Human Rights (OHCHR), *Guiding Principles on Business and Human Rights*, 2011, principle 3; CEDAW, General recommendation No. 28 (2010), para. 9; CESCR General comment No. 20 (2009), para. 11; CCPR, General comment No. 28 (2000), para. 31.

<sup>xiii</sup> See, for example: CESCR, General comment No. 20 (2009), para. 11; and CCPR, General comment No. 28 (2000), para. 31.

<sup>xiv</sup> Over the last two decades, an international consensus has developed that, in order to discharge their obligations to respect, protect and fulfil the right to non-discrimination, States must adopt specific, comprehensive anti-discrimination legislation. For discussion of this process, see: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 1, Section 1C. For statements by United Nations treaty bodies on the requirement to enact comprehensive anti-discrimination legislation, see, for example: CESCR, General comment No. 20 (2009), para. 37 and CRPD, General comment No. 6 (2018), para. 22.

<sup>xv</sup> As defined by the United Nations Human Rights Office, "comprehensive anti-discrimination legislation is law adopted with the purpose and effect of eliminating all forms of discrimination and promoting equality for all." The core necessary contents of such law are discussed in: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Executive Summary.

<sup>xvi</sup> For discussion of sanction and remedy under comprehensive anti-discrimination legislation, see United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 2, Section 2.

<sup>xvii</sup> United Nations Human Rights Office, *Guiding Principles on Business and Human Rights*, 2011, principle 11; Access Now and Amnesty International, *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems*, 2018, para. 42.

<sup>xviii</sup> See: UN General Assembly, *Report of the Secretary General*, UN Doc. A/74/821, 2020, paras. 39-40; United Nations Human Rights Office, *Guiding Principles on Business and Human Rights*, 2011, principle 13; Access Now and Amnesty International, *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems*, 2018, para. 42.

<sup>xix</sup> See, for example: UN Human Rights Council, A/HRC/47/23, 2021, para. 2; UN Human Rights Council, A/HRC/47/16, 2021, para. 1; UN General Assembly, *Report of the Secretary General*, UN Doc. A/74/821, 2020, para. 38; UN General Assembly, *Note by the Secretary-General*, UN Doc. A/70/174, 2015, paras. 24-26.

<sup>xx</sup> Access Now and Amnesty International, *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems*, 2018, para. 14.

<sup>xxi</sup> For a detailed discussion of the definition of discrimination under international law, see: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 2, Section I.A.

<sup>xxii</sup> For a list of grounds recognised at international law, see above, note 7. See also: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 2, Section I.A.

<sup>xxiii</sup> For a detailed discussion of the forms of prohibited conduct under international law, see: United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 2, Section I.A.2

<sup>xxiv</sup> For a detailed discussion of justification under international law, see United Nations Human Rights Office, *Protecting Minority Rights: A Practical Guide to Developing Comprehensive Anti-Discrimination Legislation*, New York and Geneva, 2023, Part 2, Section I.A.4.