Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT)

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

26 June 2019

MSI-AUT(2018)06rev1

Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems

Preamble

1. Member States of the Council of Europe have committed themselves to ensuring the rights and freedoms enshrined in the Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 5, "the Convention") to everyone within their jurisdiction. This commitment stands throughout the continuous processes of technological advancement and digital transformation that European societies are experiencing. As a result, member States must ensure that the design, development and implementation of algorithmic systems occur in compliance with human rights, with a view to harvesting positive effects and preventing or minimising possible adverse effects.

2. Human rights and fundamental freedoms are universal, indivisible, inter-dependent and interrelated. The use of digital applications as essential tools of everyday life, including in communication, education, health, agriculture and transportation, is rising in an unprecedented manner. They also play an increasing role in governance structures and the management and distribution of resources. Therefore, the application of algorithmic systems that have automated data collection, analytics, decision, or machine learning capacities has an evolving impact, which may be positive or negative, on the exercise, enjoyment and protection of all human rights and fundamental freedoms.

3. For the purposes of this recommendation, algorithmic systems are understood as applications that, often using various optimisation techniques, perform one or more tasks such as gathering, combining, cleaning, sorting and classifying data, as well as selection, prioritisation, recommendation and decision-making. Relying on one or more algorithms to fulfill their requirements in the settings in which they are applied, algorithmic systems increasingly permeate many aspects of contemporary human life.

4. Operating principally by detecting patterns in large datasets, algorithmic systems offer the potential to improve the performance of services (particularly through increased precision and targeting), provide new solutions, and deliver enormous efficiency and effectiveness gains in task and system performance. They have led to immense improvements in the categorisation and searchability of digital information and have facilitated important advances in fields such as medical diagnostics, transportation and logistics, enabling the broader and faster sharing of information globally and allowing novel forms of coordination. Algorithmic systems can strengthen individual autonomy and self-determination and can enhance the exercise of human rights, for instance, by broadening access to information or by facilitating the enjoyment of the freedom of assembly and association, including by creating innovative ways of associating with others.

5. However, there are also significant human rights challenges attached to the increasing reliance on algorithmic systems in everyday life. Their functionality is frequently based on the systematic aggregation and analysis of data collected through the digital tracking of online and offline behaviour of individuals and groups at scale. In addition to personal data protection and privacy costs, tracking at scale can have an important chilling effect on the freedom of expression and other human rights. While it is often argued that these concerns are justified by gains in rationalisation and accuracy, it is important to note that algorithmic systems are based on statistical models of which errors form an inevitable part, sometimes with feedback loops that replicate, reinforce and prolong pre-existing errors and assumptions. Although it may seem as if larger datasets provide better chances of finding recurrent patterns and correlations, accuracy rates do not automatically increase with the size of the dataset. As a result of the abundance of data used in automated processes, the number of errors in the form of false

positives and false negatives, and of people who are affected by these errors and inbuilt bias, will also expand, triggering additional interferences with the exercise of human rights in multiple ways.

6. Data-driven algorithmic systems do not process and generate outputs only on the basis of personal information and data. Sometimes, they are also based on non-observational and non-personal data such as simulations, synthetic data, or generalised rules, norms, procedures or laws. However, human rights may still be negatively affected at the point of use of such algorithms, even if they are trained only on synthetic data. Individuals and groups whose data is not processed or who have not otherwise been taken into consideration may also be directly concerned and significantly impacted, particularly when algorithmic systems are used to inform decision-making, adjust recommendations, or shape physical environments.

7. Many algorithmic systems use optimisation techniques where development and implementation stages are tightly entangled, as each use of the algorithmic system can be used to prompt adjustments in its functioning towards better achievement of results that are based on a narrow range of pre-defined outcomes. Such processes can shape and disrupt environments, particularly when operating at scale, as they prioritise certain values over others, for instance profit orientation over accessibility, in ways that are often not transparent, not accountable, not controllable by the affected individual, and neither serving his or her interest nor promoting collective welfare.

8. Given the wide range of types and applications of algorithmic systems in everyday life, the level of their impact – positive and negative – on human rights will always depend on the specific purpose for which they are used, their functionality and the scale at which they are deployed. It will also depend on the broader organisational, thematic, societal and legal context in which they are implemented, each associated with specific public and ethical values. Applications may be very diverse, such as for e-mail spam filters, for health-related data analytics, or for rationalising traffic flows. They may also be applied for predictive purposes in the context of policing and border control, for the purposes of combatting money laundering and fraud, or in labour, employment and educational settings, including as part of public and private recruitment and selection processes.

9. When assessing a potential negative human rights impact stemming from the design, development and implementation of an algorithmic system, it is therefore necessary to evaluate continuously and document in what context, for what purpose, with what accuracy, with what performance indicators and at what scale the system is used.

10. In many instances, the human rights impact will not attain the ‘minimum level of severity’ for any given individual that renders it significant in terms of corresponding state obligations or private actor responsibilities. Yet the same system may impact collectively upon particular groups or the population at large, triggering substantial and systematic impacts on human rights that member States should consider. For the purposes of this recommendation, the term “significant human rights impact” thus denotes relevant individual-level or systematic impacts on human rights, that engage state obligations vis-à-vis human rights.

11. In some cases, the application of an algorithmic system may prompt a particular, higher risk to human rights, for instance because it is used by states for their public service or public policy delivery and the individual does not have a possibility to opt out. A similarly heightened risk ensues as a result of use in the context of decision-making processes, by either public authorities or private parties, in situations

that carry particular weight or legal consequences. For example, the automated classification and selection of applications for bank loans can lead to the social sorting of financially weak groups or to the disruption of housing and labour markets. In this recommendation, the term “high risk” is applied when referring to the use of algorithmic systems in processes or decisions that can produce serious consequences for individuals or in situations where the lack of alternatives prompts a particularly high probability of human rights infringement.

12. Deserving of particular attention in the assessment of potential negative human rights impacts — and resulting questions of responsibility allocation — is the wide range of uses of algorithmic systems that are neither clearly public nor clearly private. This may be the case when parts of a public service are outsourced to private sector providers, who may themselves depend on other service providers, when public entities procure algorithmic systems and servicing from the private sector, or when a company deploys an algorithmic system in order to achieve public policy objectives defined by States.

13. Complicated are also cases when functions traditionally performed by public authorities, such as related to transport or telecommunications, become reliant in full or in part on the provision of algorithmic systems by private parties. When such systems are then withdrawn for commercial reasons, the result can range from decrease in quality and/or efficiency to the loss of essential services by individuals and communities. States should have contingencies in place to ensure that essential services remain available irrespective of their commercial viability, particularly in circumstances where private sector actors dominate the market in ways that place them in positions of influence or even control.

14. The design, development, and implementation of algorithmic systems engages many actors, including software designers, programmers, data sources, data workers, proprietors, sellers, users or customers, providers of infrastructure, and public and private actors and institutions. In addition, many algorithmic systems, whether learning or non-learning, operate with significant levels of opacity. Even the designer or operator, who will usually establish the overarching aim and parameters of the system, including the input data, the optimisation target and the model, is likely to encounter uncertainty about the direct and indirect effects of the system on users and the broader environments in which these systems are intended to operate.

15. While digital technologies hold significant potential for economic growth and socially beneficial innovation, the achievement of these goals must be rooted in the shared values of democratic societies. Rule of law standards that govern public and private relations in the “analogue world”, such as transparency, predictability, accountability and oversight, must also be maintained in the context of algorithmic systems. While on-going public and private sector initiatives intended to develop ethical guidelines and standards for the design, development and implementation of algorithmic systems represent highly welcome recognition of the risks that these systems pose for normative values, they do not relieve Council of Europe member States from their obligations as primary guardians of the Convention.

16. In order to live up to their obligations under the Convention, member States must refrain from direct or indirect violations through algorithmic systems, whether employed by themselves or as a result of their actions. It is essential that member States be aware of the specific human rights impacts of these processes, and that any investment in such systems contain adequate contingencies for meaningful assessment, review processes and redress for ensuing adverse effects or, where necessary, abandonment of processes that fail to meet minimum human rights standards.

17. In addition to the above commitments, the Convention also contains positive obligations for member States to establish effective and predictable legislative, regulatory and supervisory frameworks that prevent, detect, prohibit and remedy human rights violations, whether stemming from public or private actors, whether affecting relations between businesses, between businesses and consumers or between businesses and other affected individuals and groups. Member States should ensure compliance with applicable legislative and regulatory frameworks and guarantee procedural, organisational and substantive safeguards and access to effective remedies vis-à-vis all relevant actors. They should further promote an environment in which technological innovation respects and enhances human rights and complies with the fundamental obligation that all human rights restrictions be necessary, proportionate and implemented in accordance with the law.

18. Private sector actors, due to the horizontal effects of human rights and in line with the UN Guiding Principles on Business and Human Rights, have the corporate responsibility to respect the human rights of their customers and of all affected parties. To this end, flexible governance models should be adopted that guarantee fast reparation and redress when incidents occur, ensuring that responsibility and accountability for the protection of human rights are effectively and clearly distributed throughout all stages of the process, from task identification to data selection, collection and analysis, to system modelling and design, through to deployment and implementation, review and reporting requirements. Risk management processes should detect and prevent detrimental use of algorithmic systems, negative impacts or disproportionately high risks, and include the possibility of refusing deployment of certain systems when this is proportional to the possible direct or indirect harms for human rights.

19. **Against this background, and in order to provide guidance to all relevant actors who are obliged to protect and respect human rights in the contemporary, global and technology-driven environment, the Committee of Ministers, under the terms of Article 15.b of the Statute of the Council of Europe (ETS No. 1), recommends that member States:**

- **fully implement the Guidelines set out in the Appendix of this Recommendation;**
- **in implementing the Guidelines, take account of their relevant obligations under the Convention, the European Social Charter (ETS No. 35 and ETS 163), the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data as modernised in the Amending Protocol (CETS No. 223, "modernised Convention 108"), the Convention on Cybercrime (ETS No. 185, "the Budapest Convention"), the Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (CETS No. 201, "the Lanzarote Convention") and the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (CETS No. 210, "the Istanbul Convention");**
- **in implementing the Guidelines, take account of the relevant case law of the European Court of Human Rights and previous Committee of Ministers' recommendations and declarations, notably:**
    - **Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries;**
    - **CM/Rec(2016)3 of the Committee of Ministers to member States on human rights and business**
    - **Recommendation CM/Rec(2014)6 on a Guide to human rights for Internet users;**
    - **Recommendation CM/Rec(2012)3 on the protection of human rights with regard to search engines;**

- **Recommendation CM/Rec(2012)4 on the protection of human rights with regard to social networking services;**
- **Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling,**
- **Recommendation CM/Rec(2007)16 on measures to promote the public service value of the Internet;**
- **Committee of Ministers Declaration CM/Decl(13/02/2019)1 on the manipulative capabilities of algorithmic processes;**
- **the Guidelines on Artificial Intelligence and Data Protection T-PD(2019)01;**
- **the 2017 Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data,**
- **the 2016 Venice Commission Rule of Law checklist CDL-AD(2016)007**

and related international standards.

- **fully comply with their positive obligation to ensure, when devising and implementing legislative, regulatory and supervisory frameworks related to algorithmic systems, that private sector actors engaged in the design, development and implementation of algorithmic systems, fulfil their responsibilities to respect human rights in particular with regard to the United Nations Guiding Principles on Business and Human Rights;**

- **promote the goals of this Recommendation at the national level and all relevant international and regional forums; engage in, and ensure the representativeness and balance of, a regular, inclusive, meaningful and transparent dialogue, paying particular attention to the needs and voices of vulnerable groups, with all relevant stakeholders, which may include civil society, the private sector, media, education establishments, academia, as well as infrastructure providers and basic public services, including welfare and policing, with a view to sharing and discussing information, coordinating initiatives, and monitoring and assessing the responsible use of algorithmic systems that impact the exercise and enjoyment of human rights and related legal and policy issues;**

- **prioritise the building of expertise in public and private institutions involved in integrating algorithmic systems into aspects of societies with a view to effectively protecting human rights;**

- **encourage and promote the implementation of effective and tailored media, digital and information literacy programmes to support all individuals and groups to enjoy the benefits and minimise the exposure to risks stemming from the use of algorithmic systems, in effective co-operation with all relevant stakeholders, including from the private sector, media, civil society, education establishments, academia and technical institutions;**

- **review regularly and report on the measures taken to implement this recommendation and its guidelines with a view to enhancing their effectiveness.**

Appendix to Recommendation CM(20xx)x

Guidelines for States regarding the human rights impacts of algorithmic systems

These guidelines are designed to advise states and private sector actors in all their actions regarding the design, development and implementation of algorithmic systems. To ensure that the human rights and fundamental freedoms of all individuals and affected parties, as enshrined in the Convention and other relevant treaties, be effectively protected throughout technological evolution, member States of the Council of Europe shall refrain from violating human rights through the use of algorithmic systems, and shall establish legislative and regulatory frameworks that foster an environment where all actors respect and promote human rights and seek to prevent possible infringements. Independently of that and across jurisdictions, private sector actors have the responsibility to respect internationally recognised human rights.

A – Obligation of states with respect to the protection and promotion of human rights and fundamental freedoms in the context of algorithmic systems

1 General principles

1.1 **Enacting legislation:** The process of drafting and enacting legislation or regulation applicable to the design, development and implementation of algorithmic systems should be transparent, accountable and inclusive. States should regularly consult with all relevant stakeholders and affected parties.

1.2 **Computational experimentation:** States should ensure that any form of computational experimentation, such as AB testing processes, be conducted only after a meaningful human rights impact assessment. The free, specific, informed and unambiguous consent of participating individuals should be sought in advance. Experimentation designed to produce deceptive or exploitative effects should be explicitly prohibited. An accessible means of withdrawing consent is also essential. Immature software applications should not be tested on individuals, groups, or populations.

1.3 **Empowerment:** States should consider media, digital and information literacy that enables the competent and critical use of digital technologies as an essential skill for all involved in and affected by the design, development and implementation of algorithmic systems. All relevant actors, including private sector actors, media, education establishments, academia and technical institutions, should promote, in a tailored and inclusive manner (taking account of diversity with respect to, for instance, age, gender, race, ethnicity or socio-economic background), appropriate levels of understanding of the functioning of algorithmic systems and of the human rights risks stemming from their use in everyday life, enhancing the ability of all users to be aware of their rights and freedoms and use these technologies for their benefit.

1.4 **Institutional frameworks**: States should identify appropriate institutional and regulatory frameworks and standards that set benchmarks and safeguards to ensure the human rights compatibility of the design, development and implementation of algorithmic systems. Efforts should ensure that direct or indirect human rights risks, including possible cumulative effects of discrete systems, can be promptly identified and adequate remedial action initiated. States should invest in relevant technical, legal and ethical expertise to be available in adequately resourced regulatory and supervisory authorities. They

should further closely co-operate with universities, standard-setting organisations, operators of services, developers of algorithmic systems and relevant non-governmental organisations of diverse backgrounds.

2 Data management

2.1 **Interoperability:** States should ensure that all design, development, and implementation of algorithmic systems provide an avenue for individuals to analyse, manage, export and transfer their data, including through the use of interoperable data and output formats. Deliberate efforts by individuals or groups to make themselves, their physical environment or their activities illegible to automation or other forms of machine reading should be recognised as valid exercise of informational self-determination, subject to possible restrictions provided for by law.

2.2 **Datasets:** In the design, development, implementation and procurement of algorithmic systems for or by them, States should carefully assess what human rights may be affected as a result of the types of data that are being inputted or outputted into and from an algorithmic system, as these may stand in as a proxy for classifiers such as gender, race, religion, or social origin. The shortcomings of the dataset, the possibility of its inappropriate use, the negative externalities resulting from these shortcomings and inappropriate uses as well as the environments within which the dataset will be or could possibly be used, should also be assessed carefully. Particular attention should be paid to inherent risks, such as the possible re-identification of individuals using data that was previously processed on the basis of anonymity or pseudonymity, and the generation of new, inferred, potentially sensitive personal data and forms of categorisation through automated means. Based on these assessments, States should take appropriate action to prevent, where possible, or otherwise effectively minimise adverse effects.

2.3 **Infrastructure:** States should invest in and develop infrastructures for data processing and storage that is safe and secure, with a view to achieving effective capacity to respond to the increasing centralisation of data and data processing capacity (including in cloud processing) in the hands of a few companies and ensuring that high quality data processing and computational capabilities remain accessible to public and private actors alike.

3 Analysis and modelling

3.1 **Embedding of safeguards:** States should ensure that, whenever appropriate, algorithmic design, development, and implementation processes embed safety, privacy, data protection, and security safeguards by design, with a view to preventing and minimising human rights violations and other adverse effects on individuals and society. Certification schemes based on international standards should be designed and applied for labelling provenance and quality assessment of datasets. Such safeguards should also form part of procurement processes and should be informed by and compliant with regulatory frameworks that ban certain uses of algorithmic systems.

3.2 **Testing:** Regular testing, evaluation, reporting and auditing against state of the art standards related to completeness, relevance, privacy, data protection and security infringements before, during and after production and deployment should form integral part of these efforts, in particular where automated systems produce real-time effects. Efforts should include an evaluation of the legality, desirability and legitimacy of the goal that the system intends to achieve or optimise. Such evaluation should also form part of procurement processes. Any significant restrictions on human rights that are identified during

testing of such systems should result in immediate rectification and, failing that, suspension of the system until such rectifications can take place.

3.3 **Evaluation of datasets and system externalities:** States should ensure that the functioning of algorithmic systems that they implement is tested and evaluated with due regard to the fact that outputs vary according to the specific context of the deployment and the size and nature of the dataset that was used to train the system, in particular with regard to bias and discriminatory outputs. Depending on the potential impact of the algorithmic system on human rights and in order to avoid compromising other human rights, testing should, where possible, be performed without using real personal data of individuals, and should be informed through a diverse and representative stakeholder process, taking due account of the externalities of the proposed system on populations and their environments before and after deployment. States should further be aware of the possibility and risks of testing samples or outputs being reused in contexts other than those for which the system was originally developed for, including when used for the development of other algorithmic systems. This should not be permitted without new testing and evaluation of the appropriateness of such uses.

3.4 **Testing on personal data:** States should ensure that the evaluation and testing of algorithmic systems on personal data of individuals be performed with diverse, sufficiently representative sample populations. Relevant demographic groups should be neither over - nor under - represented. States should also ensure that staff involved in such activities is from sufficiently diverse backgrounds to avoid deliberate or accidental bias. Furthermore, they should ensure that the development of algorithmic systems be discontinued if testing or deployment involves the externalisation of risks or costs on to particular individuals, groups, populations and their environments. Relevant legislative frameworks should disincentivise such externalisation of risks or costs.

3.5 **Parallel modelling:** As regards the use of algorithmic systems in the delivery of public services and in other high risk contexts in which States use such technologies, alternative and parallel modelling should be performed using other methods in order to ensure that the performance and output of the algorithmic model can be adequately tested in comparison to other options.

4 Transparency, contestability and effective remedies

4.1 **Levels of transparency:** States should establish minimum levels of transparency about the use, design and basic processing criteria and methods of algorithmic systems implemented by and for them or by private sector actors. The legislative frameworks for intellectual property or trade secrets should not preclude such transparency, nor should states or private parties seek to exploit them for this purpose.

4.2. **Identifiability of algorithmic decision-making:** States should ensure that all selection processes or decisions taken or aided by algorithmic systems that may significantly impact the exercise of human rights, whether in the public or private sphere, be identifiable as such and provide the necessary information to allow for meaningful human review and contestation, in both process and rationale. The use of algorithmic systems in decision-making processes that carry high risks to human rights should be accompanied by particularly high standards of explainability of processes and outputs.

4.3 **Meaningful contestability:** States should ensure that appropriate regulatory frameworks exist to guarantee a meaningful right to contest relevant determinations and decisions. As a necessary precondition, the existence, operation, reasoning and possible outcome of algorithmic systems at

individual and collective level should be explained and clarified in a timely, impartial, user-friendly and accessible manner to individuals whose rights may be affected, as well as to relevant public authorities. The right to contest may not be waived, and should be affordable and easily enforceable before, during and after deployment, including through the provision of easily accessible contact points and hotlines.

4.4 **Adequate oversight:** States should ensure that adequate oversight is maintained over the number and type of contestations made by affected individuals or groups against certain algorithmic systems that are directly or indirectly implemented by or for them, with a view to ensuring that the results do not only lead to remedial action in the specific case but are also fed into the systems themselves so as to avoid repetitions, seek improvement, and possibly discontinue the introduction or on-going deployment of certain systems due to their human rights risks. Information on these contestations and resulting follow-up action should be documented regularly and made publicly available.

4.5 **Effective remedies:** States should ensure accessible, affordable, independent and effective judicial and non-judicial procedures that guarantee the impartial review, in compliance with Articles 6 and 13 of the Convention, of all claims of direct and indirect violations of Convention rights through the use of algorithmic systems, whether stemming from public or private sector actors. Through their legislative frameworks, they should ensure that individuals and groups are afforded with access to prompt, transparent, functional and effective remedies with respect to their grievances, including apology, deletion or rectification of data, annulment of the automated decision or compensation for damages. Judicial review should remain available and accessible, when internal and alternative dispute settlement mechanisms prove insufficient or when either of the affected parties opts for judicial redress or appeal.

4.6 **Barriers:** States should proactively seek to reduce all legal, practical or other relevant barriers that could lead to directly or indirectly affected individuals and groups being denied an effective remedy to their grievances. This includes the necessity to ensure that adequately trained staff is available to review the case competently and take appropriate action effectively.

5 Precautionary measures

5.1 **Indicators:** States should cooperate with each other and with private sector actors and relevant rights groups to develop and implement appropriate indicators, criteria and methods for state of the art human rights impact assessment processes to be conducted with regard to all algorithmic systems with potentially significant human rights impacts, with a view to evaluating potential risks and tracking actual harms, especially when such mechanisms are applied for non-targeted, explorative purposes.

5.2 **Human rights impact assessments:** States should ensure that they, as well as any private actors engaged to work with them or on their behalf, regularly conduct such human rights impact assessments prior to public procurement, during development, at regular milestones, and throughout their context-specific use to identify risks of rights-adverse outcomes. For algorithmic systems with high risks to human rights, impact assessments should include an evaluation of the possible transformations that they may bring upon existing social, institutional or governance structures.

5.3 **Expertise and oversight:** States should ensure that human rights impact assessments conducted by of for them are publicly accessible, have adequate expert input, and are effectively followed up. This may be supported by conducting dynamic testing methods and pre-release trials and by ensuring that potentially affected individuals and groups as well as relevant field experts are consulted and included as

actors with real decision-making power, where appropriate, in the design, testing, and review phases. States should ensure that human rights impacts assessments related to high-risk algorithmic systems, whether produced in the public or private sphere, be submitted for independent expert review and inspection, and tiered processes should be created for independent oversight, including by judicial authorities when necessary.

5.4 **Staff training:** States should ensure that all relevant staff involved in the procurement, development, implementation and review of algorithmic systems with potentially significant human rights impacts are adequately trained with respect to applicable human rights norms and are aware of their duty to ensure not only a thorough technical review but also human rights compliance. Hiring practices should aim for diverse workforces to enhance the ability to consider multiple perspectives in the review processes. Such approaches should be documented with a view to promoting them beyond the public sector. States should also work together to share experiences and develop best practices.

5.5 **Interactivity of systems:** States should carefully monitor settings where multiple algorithmic systems operate in the same environment, or a given system bundles multiple algorithmic systems in order to identify and mitigate negative externalities, where responsibility is difficult to apportion. States should utilise the mechanism of procurement or engagement of private services in public service delivery in full consideration of the need to maintain relevant oversight capacity, know-how, ownership and control over the use of algorithmic systems in multiple aspects of societies, with a view to avoiding path dependencies and preserving the viability of alternative solutions. Insofar as private sector actors provide services that are considered essential in modern society or have a de facto monopoly in providing such services, member States should develop regulatory frameworks that ensure effective enjoyment of human rights by affected individuals and groups. They should publicly account for their efforts in this regard.

5.6 **Public debate:** States should engage in inclusive, inter-disciplinary, informed and public debates to define what areas of public services affecting the exercise of human rights may not be determined, decided or optimised through algorithmic systems.

6 Empowerment through research, innovation and public awareness

6.1 **Rights-promoting technology:** States should promote the development of algorithmic systems and technologies that enhance equal access to and enjoyment of human rights and fundamental freedoms through the use of tax, procurement, or other incentives. This includes the development of mechanisms to evaluate the impact of algorithmic systems, the development of systems to address the needs of disadvantaged and underrepresented populations, as well as necessary efforts to ensure the sustainability of basic services through analogue means, both as contingency and as an effective opportunity for individuals to opt out.

6.2 **Advancement of public benefit:** States should engage in and support independent research aimed at assessing, testing and advancing the potential of algorithmic systems for creating positive human rights effects and for advancing public benefit. This may require the anticipation and possible discouragement of influences that may exclusively favour most commercially viable optimisation processes.

6.3 **Human-centric and sustainable innovation:** States should promote innovative design and technological development in line with existing human rights norms, in particular with respect to social rights and

internationally recognised labour and employment standards, to enhance internationally agreed sustainable development goals, including as regards extraction and exploitation of environmental resources, and to address existing environmental challenges, such as through initiatives towards fair and human-centric innovation.

6.4 **Independent research:** States should encourage independent research into the development of effective accountability mechanisms and solutions to existing responsibility gaps related to opacity, inexplicability and related incontestability of algorithmic systems. Appropriate mechanisms should be put in place to guarantee the impartiality, global representation, and protection of researchers, journalists and academics engaged in such independent research.

6.5 **Control over data:** States should investigate strategies to prevent the monopolisation of control over data and data processing capacity with a view to ensuring the independence and vitality of the public and private sector, promoting the design and development of algorithmic systems in the public interest, and curbing concentration of market power.

B. Responsibilities of private sector actors with respect to human rights and fundamental freedoms in the context of algorithmic systems

1 General principles

1.1 **Responsibility to respect:** Private sector actors engaged in the design, development, sale, deployment, implementation and servicing of algorithmic systems, whether in the public or private sphere, have the responsibility to respect internationally recognised human rights and fundamental freedoms of their customers and of other parties who are affected by their activities. This responsibility exists independently of States' ability or willingness to fulfil their human rights obligations. As part of fulfilling this responsibility, private sector actors should take continuing, proactive and reactive steps to ensure that they do not cause or contribute to human rights abuses and that their innovation processes are human rights-friendly.

1.2 **Scale of measures:** The responsibility of private sector actors to respect human rights and to employ adequate measures applies regardless of their size, sector, operational context, ownership structure or nature. The scale and complexity of the means through which they meet their responsibilities may vary, however, taking into account their means and the severity of the potential impact on human rights by their services and systems. Where different sets of private sector actors co-operate and contribute to potential human rights interferences, efforts from all partners are required and should be proportional to their respective impact and abilities.

1.3 **Additional key standards:** Given that the design, development and implementation of algorithmic systems engages private sector actors at many levels and often in close cooperation with public actors, some of the provisions that are outlined in Chapter A as obligations of States also translate into corporate responsibilities for private sector actors. Irrespective of whether corresponding regulatory action has been taken by States and in addition to the below provisions, private sector actors should uphold the standards contained in provisions 1.2, 1.3, 2.1, 3.1, 3.2, 4.2, 5.2 and 6.3 of Chapter A.

1.4 **Discrimination:** Private sector actors should produce and provide their products and services without discrimination. They should seek to ensure that the design, development or implementation of their algorithmic systems do not have direct or indirect discriminatory effects or harmful impacts on individuals or groups that are affected by these systems, including on those who have special needs or disabilities or may face structural inequalities in their access to human rights.

2 Data management

2.1 **Consent rules:** Private sector actors should ensure that individuals who are affected by their algorithmic systems with potential for significant human rights impacts are empowered with the choice to give and revoke free and informed consent regarding all use of their data, with both processes being equally easily accessible. Users should be further empowered to know how their data is being used, what the real and potential impact of the algorithmic system in question is, how to object to relevant processing of their data, and how to contest and challenge specific outputs. Consent rules for the use of tracking, storage and performance measurement tools of algorithmic systems must be clear, simply phrased, meaningful and complete.

2.2 **Privacy settings:** Private sector actors should facilitate the right of users to protect effectively their privacy while maintaining access to services, including through the possibility of choosing from a set of privacy setting options, presented in an easily visible, neutral and intelligible manner, or through the use of privacy enhancing technologies. Default options should lead only to the collection of data that are necessary for the specific purpose of the data processing. Any application of mechanisms to block, erase or quarantine user data, such as for security purposes, should be accompanied with due process guarantees and rapid remedies available in case of erroneous or disproportionate use.

3 Analysis and modelling

3.1 **Data bias:** Private sector actors should be cognisant of risks relating to quality, nature and origin of the data they are using for training their algorithmic systems, with a view to ensuring that bias and potential discrimination in datasets is adequately responded to within the specific context.

3.2. **Sample populations:** The evaluation and testing of algorithmic systems on personal data of individuals should be performed with diverse, sufficiently representative sample populations and not draw on or discriminate against any particular demographic group. Development of algorithmic systems should be discontinued or adjusted if development, testing or deployment involves the externalisation of risks or costs on to particular individuals, groups, populations and their environments.

3.3 **Illegal access:** Private sector actors should configure their algorithmic systems in such a way that it prevents illegal access, system interference and misuse of devices by third parties in line with applicable standards.

4 Transparency, contestability and effective remedies

4.1 **Terms of service:** Private sector actors should ensure that the use of algorithmic systems in the products and services they offer is made known to all affected parties, whether individual or legal entities, as well as to the general public in clear and plain language and in accessible formats. Terms of service should be easily understandable, containing clear and succinct language about possibilities for users to influence settings, about available options to change the features of the system, about applicable complaint mechanisms, the various stages of the procedure, the exact competencies of the contact points, indicative time frames and expected outcomes. All affected parties, new customers or customers of products and services whose application rules have been amended should be notified of relevant changes in a user-friendly format, and requested to consent to the changes where relevant. Failure to consent should not lead to essential services becoming unavailable.

4.2 **Contestability:** Private sector actors should make public information about the number and type of contests made by affected individuals or groups regarding the products and services they offer, with a view to ensuring that the results do not only lead to remedial action in the specific case but are also fed into the systems themselves to draw lessons from complaints and correct errors before harm occurs at massive scale.

4.3 **Human review:** In order to facilitate meaningful contestability, private sector actors should ensure that human reviewers remain accessible and that direct contact is made effectively possible, including through the provision of easily accessible contact points and hotlines. Individuals and groups should be allowed not only to contest but also to make suggestions for improvements and provide other useful

feedback, including with respect to areas where human review is systematically required. All staff involved in the handling of customer complaints should be suitably versed in relevant human rights standards and benefit from regular training opportunities.

4.4 **Effective remedies:** Private sector actors should ensure that effective remedies and dispute resolution systems, including collective redress mechanisms, are available both online and offline to individuals, groups and legal entities, who wish to contest the introduction of a system with potential for human rights violations or remedy a violation of rights. The scope of available remedies may not be limited. All remedies should allow for an impartial and independent review, should be handled without unwarranted delays and should be conducted in good faith, with respect for due process guarantees. Relevant mechanisms should not negatively impact the opportunities for complainants to seek recourse through independent national, including judicial, review mechanisms. No waivers of rights or hindrances to the effective access to remedies may be included in their terms of service.

4.5 **Participation:** Private sector actors should actively engage in participatory processes with consumer associations, human rights advocates and other organisations representing the interests of individuals and affected parties, as well as with data protection and other independent administrative or regulatory authorities, for the design, implementation and evaluation of their complaint mechanisms, including collective redress mechanisms. Business associations should further invest – in cooperation with trade associations – in the establishment of model complaints mechanisms.

5 Precautionary measures

5.1 **Continuous evaluation:** Private sector actors should develop internal processes to ensure that their design, development and implementation of algorithmic systems is continuously evaluated and tested not only against possible technical errors but also against the potential legal, social and ethical impacts that the systems may carry. Where the application of algorithmic systems carries high risks to human rights, including through processes of micro-targeting, private sector actors should notify and consult supervisory authorities in all relevant jurisdictions to seek advice and guidance on how to manage these risks, including through the redesign of the services that led to the problematic outcome. Private sector actors should submit these algorithmic systems for regular independent expert review, and create tiered processes for independent oversight, including by judicial authorities when necessary.

5.2 **Staff training:** All relevant staff involved in human rights impact assessments and in the review of algorithmic systems should be adequately trained and aware of their responsibilities with respect to human rights including, but not limited to, applicable personal data protection and privacy standards.

5.3 **Human rights impact assessments:** Human rights impact assessments should be conducted as openly as possible and encourage active engagement of affected individuals and groups. In case of implementation of high-risk algorithmic systems, the results of human rights impacts assessment, identified techniques for risk mitigation, and relevant monitoring and review processes should be made publicly available, without prejudice to secrecy safeguarded by law. When secrecy rules need to be enforced, any confidential information should be provided in a separate annex to the assessment report. This annex shall not be public, but should be accessible by relevant supervisory authorities.

5.4 **Follow up:** Private sector actors should ensure appropriate follow-up to their human rights impact assessments by taking adequate action upon the findings and monitoring the effectiveness of identified

responses, with a view to avoiding or mitigating adverse effects on and risks for the exercise of human rights. Identified failures should be resolved as quickly as possible and related activities suspended where appropriate. This requires regular and continued quality assurances checks and real-time auditing through design, testing, and deployment stages to monitor algorithmic systems for human rights impacts in context and in situ, and to correct errors and harms as appropriate. This is particularly important given the risk of feedback loops that can exacerbate and entrench negative outcomes.

6 Empowerment through research, innovation and public awareness

6.1 **Research:** Private sector actors should engage in ethical research aimed at assessing, testing and advancing the potential of algorithmic systems for creating positive human rights impacts and for advancing public benefit. They should also support independent research with this aim and respect the integrity of researchers and research institutions. This may include the development of mechanisms to evaluate the impact of algorithmic systems and the development of algorithmic systems to address the needs of disadvantaged and underrepresented populations.

6.2 **Access to data:** Private sector actors should provide access to relevant individual and meta-datasets in full respect of data protection legislation and principles, as well as access to data that has been classified for deletion, to independent researchers, journalists and academics engaged in analysing the impacts of algorithmic systems and digitalised services on the exercise of rights, on communication networks, and on democratic systems.