Strasbourg, 21 November 2019

T-PD(2019)09

**CONSULTATIVE COMMITTEE OF THE CONVENTION
FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING
OF PERSONAL DATA**

**CONVENTION 108**

# OPINION ON THE DRAFT RECOMMENDATION ON THE HUMAN RIGHTS IMPACT OF ALGORITHMIC SYSTEMS

Directorate General Human Rights and Rule of Law

1. The Steering Committee on Media and Information Society (CDMSI) circulated on 5 November 2019 to its delegations (members and observers) the draft[1] Recommendation of the Committee of Ministers to member States on "the human rights impacts of algorithmic systems" (prepared and finalised by the Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence - the MSI-AUT), for possible comments by 24 November 2019, in view of its 17th Plenary meeting (3-5 December 2019).

2. The Committee of Convention 108 welcomes this important work, and the opportunity given to comment the draft finalised by the MSI-AUT. The scientific expert[2] who worked with the Committee of Convention 108 on artificial intelligence and big data attended all the meetings of the MSI-AUT with a view to ensuring full consistency of the draft with the data protection standards already adopted, the topic of the draft Recommendation being closely related to the work of the Committee of Convention 108.

3. The Committee of Convention 108 wishes to highlight from the outset its work of relevance to this topic. In particular, the Committee of Convention 108 recalls the recent modernisation of Convention 108 (Convention for the protection of individuals with regard to automatic processing of data) which aimed at responding to emerging challenges, notably in the context of algorithmic decision-making environments. A number of new provisions in the Amending Protocol CETS 223 specifically address those challenges (see for instance the increased transparency obligation, new rights for the data subject, the obligation of accountability, of privacy impact assessments, of privacy by design and by default).

4. The importance and relevance of Convention 108+ is acknowledged in recital 12 of the Preamble which reads "Reiterating particularly the importance of existing personal data protection standards, notably Convention 108 as modernised in the Amending Protocol (CETS 223)".

5. The Committee of Convention 108 adopted two sets of Guidelines of core relevance to algorithmic systems: the Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data (adopted on 23 January 2017) and Guidelines on data protection and artificial intelligence (adopted on 25 January 2019).

6. The Committee of Convention 108 considers necessary to comment on a series of specific provisions of the Appendix to the draft Recommendation ("Guidelines for States regarding the human rights impacts of algorithmic systems" which are, as stated in the first paragraph "designed to advise States, public and private sector actors in all their actions regarding the design, development and ongoing deployment of algorithmic systems".)

7. Paragraph 11 of the Guidelines highlights that "the application of an algorithmic system can 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". The Committee acknowledges that there are cases where it is more difficult to protect the rights potentially affected, but notes that this does not necessarily mean that the impact is higher.

[1] Document MSI-AUT(2018)06rev3, see in Appendix.

2 Professor Alessandro Mantelero, Politecnico Torino.

8. Paragraph 12 of the Guidelines refers to “algorithmic systems that are neither clearly public nor clearly private” […] “when parts of a public service are outsourced to private sector providers who may themselves depend on other service providers”. For any such outsourcing (possibly in chain), there should legally always be a clear organisational structure and a related allocation of responsibilities. In particular, from a data protection perspective, the legal definitions of controllers and processors (Article 2, litt. d) and f) of Convention 108+) help defining the different roles of the actors involved in the processing, and the legal consequences they carry.

9. With regard to the consultation of “all relevant stakeholders” foreseen in section 1.1, it may be necessary to proceed to sectorial consultations considering the incredibly wide range of AI applications.

10. Section 2.2 on datasets provides for the careful assessment of the quality of “outputted data” from the algorithmic system, which may be better specified by referring to the risk of decontextualisation of that data, i.e. the risk of ignoring the contextual information characterising the specific situations in which the proposed AI applications are meant to be used. Regarding the reference to the risk of possible identification of individuals using data processed based on pseudonymised data, it should be underlined that identification in that case is not a risk but rather an inherent possibility due to the particular nature of this data.

11. Section 3.3 on testing provides for the “evaluation of the legality and legitimacy of the goal that the system intends to achieve or optimise, and its possible human rights effects”. This should have already been assessed by the law allowing for this system and what would thus need to be assessed is the manner in which such goals are achieved, rather than the legality and legitimacy of the goal itself.

12. Regarding the testing on personal data (section 3.5) it should be noted that the requirement of diverse samples of population should not always be imposed, for instance when the application evaluated or tested concerns a specific group of population only. Finally, the reference to the “costs” for the individuals does not correspond to the standard terminology of risk assessment in respect of the potential infringement of human rights.

13. The Committee of Convention welcomes the central role played by human rights impact assessments, which is in line with the modernised Convention 108, for which no methodological guidance seems to be provided (the question of the small scale use of human rights impact assessments could be raised, noting the broad variety of development and use of AI applications while such impact assessments are usually performed on large-scale processes).

14. Moreover, it welcomes the importance given to the identifiability of algorithmic decision-making and meaningful contestability (sections 4.2 and 4.3), which are also in line with the transparency requirements and with the right of the individual not to be subject to purely automated decisions (without having his or her views taken into consideration) provided for by Convention 108+.

15. Section 5.1 provides an obligation to carry out human rights impact assessments "for all algorithmic systems with high risks to human rights". This requirement is narrower than the obligation imposed under the latest generation of data protection legal frameworks. For instance, Article 10.2 of Convention 108+ foresees that "controllers and, where applicable, processors, examine the likely impact of intended data processing on the rights and fundamental freedoms of data subjects prior to the commencement of such processing" without any reference to a high level of risks (also see the Guidelines on data protection and artificial intelligence). A risk-based approach is also contained in the European Union's General Data Protection Regulation, while formally imposing data protection impact assessments a processing is likely to result in a high risk to the rights and freedoms of natural persons (Article 35 of the GDPR).

16. Section 5.2 requires that "where private sector actors provide services that rely on algorithmic systems and that are considered essential in modern society for the effective enjoyment of human rights, member States should preserve the future viability of alternative solutions and ensure the continued access to such services by affected individuals and groups." The Committee of Convention 108 would take the example of the use of AI for image recognition by an hospital, to raise the issue of the implementation of alternative solutions, where the latter may actually be worse in terms of performance.

17. Regarding the second part (data management) of the Guidelines concerning the "Responsibilities of private sector actors with respect to human rights and fundamental freedoms in the context of algorithmic systems", the Committee notes the choice of the drafters to address two issues: consent and privacy settings. The Committee of Convention 108 recalls that consent is only one of the possible legal grounds on the basis of which personal data may be processed, and that AI applications can also be used on the basis of another legitimate ground laid down by law. This should be reflected in the text. Furthermore, it could be understood *a contrario* from the sentence "private sector actors should ensure that individuals who are affected by their algorithmic systems with potential for significant human rights impacts are informed of and empowered with the choice to give and revoke their consent regarding all uses of their data, with both options being equally easily accessible" that where there is no potential for significant human rights impacts, such a requirement is not applicable, which would be incorrect.

18. Privacy settings mentioned in Section 2.2 are an important feature of the system but ensuring respect for the right to data protection actually requires that many more conditions be satisfied, and this could be mentioned, referring the provisions of Convention 108+ which all apply where personal data is being processed.

19. Section 3.3 on systems and data security provides for measures to be put in place by the private sector in light of the action by third parties. The Committee of Convention 108 stresses that the controller's responsibility in that regard actually not only concerns third parties, but also any AI developer or any other employee working for that particular firm and could also be considered as a potential source of vulnerability in terms of system and data security (see Article 7 of Convention 108+).

20. In conclusion, the Committee of Convention 108 once again welcomes this important work, and any effort made in securing human rights compliant design, development, deployment and use of algorithmic systems and hopes that the present opinion will be helpful to the CDMSI.

MSI-AUT(2018)06rev3

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

***Preamble***

The Committee of Ministers, under the terms of Article 15.*b* of the Statute of the Council of Europe (ETS No. 1),

Considering that 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 and that this commitment stands throughout the continuous processes of technological advancement and digital transformation that European societies are experiencing;

Reaffirming that, as a result, member States must ensure that any design, development and ongoing deployment of algorithmic systems occur in compliance with human rights and fundamental freedoms, which are universal, indivisible, inter-dependent and interrelated, with a view to amplifying positive effects and preventing or minimising possible adverse effects;

Recognising the unprecedented rise in the use of digital applications as essential tools of everyday life, including in communication, education, health, economic activities and transportation, and their increasing role in governance structures and the management and distribution of resources;

Conscious therefore of the evolving impact, which may be positive or negative, that the application of algorithmic systems with automated data collection, analytics, decision making, optimisation or machine learning capacities has on the exercise, enjoyment and protection of all human rights and fundamental freedoms, and of the significant challenges attached to the increasing reliance on algorithmic systems in everyday life, also for democratic societies and the rule of law;

Underlining the need to ensure that racial, gender and other societal and labour force imbalances that have not yet been eliminated from our societies are not deliberately or accidentally perpetuated through algorithmic systems, as well as the desirability of addressing these imbalances through using appropriate technologies;

Bearing in mind that digital technologies hold significant potential for socially beneficial innovation and economic development, and that the achievement of these goals must be rooted in the shared values of democratic societies and subject to meaningful democratic participation and oversight;

Reaffirming therefore that rule of law standards that govern public and private relations, such as legality, transparency, predictability, accountability and oversight, must also be maintained in the context of algorithmic systems;

Considering that ongoing public and private sector initiatives intended to develop ethical guidelines and standards for the design, development and ongoing deployment of

algorithmic systems, while constituting highly welcome recognition of the risks that these systems pose for normative values, do not relieve Council of Europe member States from their obligations as primary guardians of the Convention;

Recalling the obligation of member States under the Convention to refrain from human rights violations through algorithmic systems, whether employed by themselves or as a result of their actions, and the obligation 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 and whether affecting relations between businesses, between businesses and consumers or between businesses and other affected individuals and groups;

Emphasising that 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, while promoting an environment in which technological innovation respects and enhances human rights and complies with the fundamental obligation that all human rights restrictions be necessary and proportionate in a democratic society, and implemented in accordance with the law;

Taking account of and building on existing Council of Europe, regional and international norms, standards, and recommendations related to the protection of human rights and fundamental freedoms in contemporary societies, as well as the evolving jurisprudence of the European Court of Human Rights;

Reiterating particularly the importance of existing personal data protection standards, notably Convention 108 as modernised in the Amending Protocol (CETS 223), while emphasising that the human rights impacts of algorithmic systems are broader and call for additional protections;

Recalling further that private sector actors, 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 and that, to this end, flexible governance models should be adopted that guarantee fast and effective 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 the proposal stage through to task identification, data selection, collection and analysis, system modelling and design, through to ongoing deployment, review and reporting requirements;

Acknowledging the fact that the fast-moving socio-technical developments require constant monitoring and adaptation of applicable governance frameworks to protect human rights effectively in a complex and global environment and recognising the need for regular guidance to be provided to all relevant public and private sector actors,

Recommends that member States:

1. review their legislative frameworks and policies as well as their own practices with respect to the procurement, design, development and ongoing deployment of algorithmic systems to ensure that they are in line with the guidelines set out in the appendix of this recommendation, promote their implementation in all relevant areas and evaluate the effectiveness of the measures taken at regular intervals, with participation of all relevant stakeholders;

2. ensure that this recommendation, including the guidelines in the appendix, be translated and disseminated as widely as possible and through all accessible means among competent authorities and stakeholders, including parliaments, independent authorities, specialised public agencies, civil society organisations and the private sector;

3. ensure, through appropriate legislative, regulatory and supervisory frameworks related to algorithmic systems, that private sector actors engaged in the design, development and ongoing deployment of algorithmic systems comply with applicable laws and fulfil their responsibilities to respect human rights in line with the United Nations Guiding Principles on Business and Human Rights and relevant regional and international standards;

4. endow their relevant national supervisory, oversight and enforcement institutions with the necessary resources and authority to investigate, oversee and coordinate compliance with their relevant legislative and regulatory framework, in line with this recommendation;

5. engage in regular, inclusive, meaningful and transparent consultation, cooperation and dialogue with all relevant stakeholders (such as civil society, human rights defence organisations, the private sector, the academic and professional community, media, education establishments, public libraries, infrastructure providers and basic public services, including welfare and policing), paying particular attention to the needs and voices of vulnerable groups, with a view to ensuring that significant human rights impacts stemming from the design, development and ongoing deployment of algorithmic systems be comprehensively monitored, debated and addressed;

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

7. encourage and promote the implementation of effective and tailored media, digital and information literacy programmes to enable all individuals and groups to (1) understand the functions and ramifications of systems employing automated decision making, (2)

make informed decisions in the use of such systems, (3) enjoy the benefits, and (4) minimise the exposure to threats and 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;

8. take account of the environmental impact of the development of large-scale digital services and take necessary steps to optimise the use and consumption of natural resources and energy;

9. review regularly, in consultation with all relevant actors, and report domestically and within the Committee of Ministers on the measures taken to implement this recommendation and its guidelines with a view to enhancing their effectiveness and adapting them to evolving challenges.

**Appendix to *Recommendation* CM(20xx)x**

**Guidelines for States regarding the human rights impacts of algorithmic systems**

## A. Scope and context

1. These guidelines are designed to advise States, public and private sector actors in all their actions regarding the design, development and ongoing deployment of algorithmic systems. To ensure that the human rights and fundamental freedoms of all individuals, 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 develop legislative and regulatory frameworks that foster an environment where all actors respect and promote human rights and seek to prevent possible infringements. Independently of State obligations and across jurisdictions, public and private sector actors have the responsibility to respect internationally recognised human rights.

2. For the purposes of this recommendation, algorithmic systems are understood as applications that, often using mathematical optimisation techniques, perform one or more tasks such as gathering, combining, cleaning, sorting, classifying and inferring 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 automate activities in a way that allows the creation of adaptive services at scale and in real time.

3. Operating typically by detecting patterns in large datasets, algorithmic systems offer the potential to improve the performance of services (particularly through increased precision, targeting and consistency), provide new solutions, and deliver returns in efficiency and effectiveness of 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 making possible novel forms of cooperation and coordination. As a result, they permeate many aspects of contemporary human life.

4. However, there are also significant human rights challenges attached to the increasing reliance on algorithmic systems in everyday life, such as relating to the right to fair trial, the right to privacy and data protection, the right to freedom of thought, conscience and religion, freedom of expression, freedom of assembly, the right to equal treatment, and economic and social rights. The functionality of algorithmic systems 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 the intrusion on individuals' privacy and the increasing potential of highly personalised manipulation, tracking at scale can have an important adverse effect on the exercise of human rights which must be considered from the proposal stage of algorithmic development or use onward.

5. While it is often argued that these costs are offset by gains in rationalisation and accuracy, it is important to note that most 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 biases, 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 large number of people affected by algorithmic systems, 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. Algorithmic systems do not process and generate outputs only based on personal data. They can also operate based on non-observational and non-personal data such as simulations, synthetic data, or generalised rules or procedures. However, human rights may still be negatively affected at the point of use of such algorithms. 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. Each use of the algorithmic system can 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. They prioritise certain values over others, for instance general gains over specific losses. This typically happens in ways that fail to be explicit, transparent, accountable or controllable by the affected individual, and may generate adverse effects, particularly for minorities and marginalised or disadvantaged groups.

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, accuracy, complexity, their effects 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 used, each of which associated with specific public and ethical values. Applications are very diverse, such as for e-mail spam filters, for health-related data analytics, or for rationalising traffic flows. They are also 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 ongoing deployment of an algorithmic system, it is therefore necessary to evaluate continuously and document the context, legal ground, purpose, accuracy, side effects and scale of the system's use. The inherent risks of these systems being attacked or confused via adversarial machine-learning or by other

means (including cyber-attacks) due to the scale, nature and possible value of the data that is processed should also be considered. The evaluation of the extent of the possible human rights impact of an algorithmic system should take account of the severity, scale and likelihood of giving rise to a human rights violation.

10. Many uses of algorithmic systems have minimal potential of creating an adverse human rights impact for an individual and therefore do not trigger corresponding State obligations or private actor responsibilities. Yet, the same system may have a collective impact on particular groups or the population at large, generating effects on human rights, democratic processes or the rule law that member States should consider. For the purposes of this recommendation, the term “significant human rights impact” denotes either relevant individual-level or collective impacts that engage State obligations or private sector responsibilities vis-à-vis human rights.

11. In some cases, the application of an algorithmic system can 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 or suffers negative consequences as a result of the decision 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 consequence. For example, the use of algorithmic systems in the judicial field for the purpose of legal analysis or individual risk assessment must be introduced with great care and in conformity with the guarantees of a fair trial as enshrined in Article 6 of the Convention. 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, including by amplifying or introducing distributive injustices.

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. Cases where 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 are also complicated. When such systems are then withdrawn for commercial reasons, the result can range from decrease in quality and/or efficiency to the loss of services that are considered essential 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 ongoing deployment 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, sometimes even deliberately. 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, will often not know what of the given information the system relies upon to make its decision, and 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. Given this complexity, 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 monitoring, assessment, review processes and redress for ensuing adverse effects or, where necessary, abandonment of processes that fail to meet human rights standards. Risk management processes should detect and prevent detrimental use of algorithmic systems and negative impacts. They should be based on a precautionary approach and require the refusal of certain systems when their deployment leads to high risks of irreversible damage or when, due to their opacity, human control and oversight become impractical.

## B. Obligation of states with respect to the protection and promotion of human rights and fundamental freedoms in the context of algorithmic systems

### 1 Principles of general application

1.1 **Legislation:** The process of drafting, enacting and evaluating policies and legislation or regulation applicable to the design, development and ongoing deployment of algorithmic systems should be transparent, accountable and inclusive. States should regularly consult with all relevant stakeholders and affected parties. States should ensure the enforceability and enforcement of laws, including by demanding that relevant actors produce adequate documentation to verify legal compliance. Where public and private sector actors fail to discharge their legal duties, they should be held responsible.

1.2. **Ongoing review:** Throughout the entire lifecycle of an algorithmic system, from the proposal stage through to the evaluation of effects, the human rights impacts of individual systems and their interaction with other technologies should be assessed regularly. This is necessary due to the speed and scale at which these systems function and the fast-evolving technological environment in which they operate. This should be done based on broad, meaningful consultations with those affected or likely to be affected.

1.3 **Democratic participation and awareness:** In order to ensure meaningful exercise of human rights and democratic freedoms, States should foster general public awareness of the capacity, power and consequential impacts of algorithmic systems, including their potential use to manipulate, exploit, deceive, or distribute resources, with a view to enhancing the ability of all individuals and groups to be aware of their rights and know how to act upon them and how to use digital technologies for their benefit. In addition, all relevant actors, including private, public and civil society actors across all sectors where algorithmic systems are contemplated or in use, should promote, encourage and support in a tailored and inclusive manner (taking account of diversity with respect to, for instance, age, gender, race, ethnicity, cultural or socio-economic background) a level of media, digital and information literacy that enables the competent and critical consideration of and use of algorithmic systems.

1.4 **Institutional frameworks**: States should identify and/or develop appropriate institutional and regulatory frameworks and standards that set general or sector-specific benchmarks and safeguards to ensure the human rights compatibility of the design, development and ongoing deployment of algorithmic systems. Efforts should ensure that direct or indirect human rights risks, including possible cumulative effects of discrete systems, are promptly identified and adequate remedial action initiated. States should invest in relevant expertise to be available in adequately resourced regulatory and supervisory authorities. They should further closely co-operate with independent authorities, equality bodies, universities, standard-setting organisations, operators of services, developers of algorithmic systems and relevant non-governmental organisations of diverse backgrounds, such as, particularly, those engaged in human rights defence.

## 2 Data management

2.1 **Informational self-determination:** States should ensure that all design, development, and ongoing deployment of algorithmic systems provide an avenue for individuals to be informed in advance of the data processing (including understanding its purposes and possible outcomes) and to control their data, which may not be limited to personal data. Deliberate efforts by individuals or groups to make themselves, their physical environment or their activities illegible to automation or other forms of machine reading or manipulation should be recognised as a valid exercise of informational self-determination, subject to possible exceptions or derogations necessary and proportionate in a democratic society and provided for by law.

2.2 **Datasets:** In the design, development, ongoing deployment and procurement of algorithmic systems for or by them, States should carefully assess what human rights and non-discrimination rules may be affected as a result of the quality of data that are being inputted or outputted into and from an algorithmic system, as these often contain bias and may stand in as a proxy for classifiers such as gender, race, religion, political opinion or social origin. The provenance and possible 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 identification of individuals using data that was previously processed based on anonymity or pseudonymity, and the generation of new, inferred, potentially sensitive data and forms of categorisation through automated means. Based on these assessments, States should take appropriate action to prevent and effectively minimise adverse effects.

2.3 **Infrastructure:** The increasing centralisation of data and data processing capacity (including in cloud processing) and the possibility of a lack of choice may negatively impact States' ability to discharge their human rights obligations under the Convention. Therefore, they should facilitate the development of alternative, safe and secure infrastructures to ensure that high quality data processing and computational capabilities remain available to public and private actors alike.

## 3 Analysis and modelling

3.1 **Computational experimentation:** States should ensure that computational experimentation that triggers the likelihood of significant human rights impacts be conducted only after a human rights impact assessment. The free, specific, informed and unambiguous consent of participating individuals should be sought in advance, with an accessible means of withdrawing consent. Experimentation designed to produce deceptive or exploitative effects should be explicitly prohibited.

3.2. **Embedding of safeguards:** States should ensure that algorithmic design, development, and ongoing deployment processes embed safety, privacy, data

protection, and security safeguards by design, with a view to preventing and mitigating the risk of human rights violations and other adverse effects on individuals and society. Certification schemes based on regional and international standards should be designed and applied for labelling provenance and quality assessment of datasets and models. 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.

3.3 **Testing:** Regular testing, evaluation, reporting and auditing against state-of-the-art standards related to completeness, relevance, privacy, data protection, other human rights, unjustified discriminatory impacts and security breaches before, during and after production and deployment should form integral part of testing efforts, particularly where automated systems are tested in live environments and produce real-time effects. State efforts should include public, consultative and independent evaluation of the legality and legitimacy of the goal that the system intends to achieve or optimise, and its possible human rights effects. 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.

3.4 **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, including with regard to bias and discriminatory outputs. Depending on the potential impact of the algorithmic system on human rights, testing should, where possible, be performed without using real personal data of individuals, and guided 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, 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.

3.5 **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 unintentional 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 specific individuals, groups, or populations and their environments. Relevant legislative frameworks should disincentivise the externalisation of risks or costs. Special care should be taken in relation to testing in live environments.

3.6 **Alternative and parallel approaches:** 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, methods such as alternative and parallel modelling should be performed in order to ensure that the decision to use or procure as well as the performance and the output of the algorithmic system can be adequately tested in comparison to other options.

## 4 Transparency, accountability and effective remedies

4.1 **Levels of transparency:** States should establish appropriate levels of transparency about the public procurement, 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. Transparency levels should be as high as possible and proportionate to the severity of adverse human rights impact. The use of algorithmic systems in decision-making processes that carry high risks to human rights should be accompanied by particularly high levels of explainability of processes and outputs.

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 at the initial interaction and in a clear and accessible manner.

4.3 **Meaningful contestability:** Affected individuals and groups should be afforded effective means to contest relevant determinations and decisions. As a necessary precondition, the existence, process, rationale, reasoning and possible outcome of algorithmic systems at individual and collective level should be explained and clarified in a timely, impartial, easily-readable and accessible manner to individuals whose rights or legitimate interests may be affected, as well as to relevant public authorities. Meaningful contestation should include an opportunity to be heard, actual review of the decision and the possibility to obtain a non-automated decision. It 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 **Consultation and adequate oversight:** States should ensure that adequate oversight is maintained by appropriately resourced independent institutions 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. They should ensure that the results do not only lead to remedial action in the specific case but are also fed into the systems themselves to avoid repetitions, seek improvement, and possibly discontinue the introduction or on-going deployment of certain systems due to the likelihood of negative human rights impacts. Information on these contestations and resulting follow-up action should be documented regularly and made publicly available.

4.5 **Effective remedies:** States should ensure equal, accessible, affordable, independent and effective judicial and non-judicial procedures that guarantee an impartial review, in compliance with Articles 6, 13 and 14 of the Convention, of all claims of 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 provided with access to prompt, transparent, functional and effective remedies with respect to their grievances. Judicial redress should remain available and accessible, when internal and alternative dispute settlement mechanisms prove insufficient or when either of the affected parties opts for judicial review 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 **Standards:** States should cooperate with each other and with all relevant stakeholders, including civil society, to develop and implement appropriate guidance (e.g. standards, frameworks, indicators, and methods) for state-of-the-art human rights impact assessment processes. These should be conducted with regard to all algorithmic systems with potentially significant human rights impacts at any stage of the lifecycle, with a view to evaluating potential risks and setting out measures, safeguards and mechanisms for preventing or mitigating such risks. Actual harms should be tracked, especially when such systems are applied for non-targeted, explorative purposes. Human rights impact assessments should be made mandatory for all algorithmic systems with high risks to human rights.

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 and consultatively conduct human rights impact assessments prior to public procurement, during development, at regular milestones, and throughout their context-specific deployment to identify risks of rights-adverse outcomes. Algorithmic systems should not be procured if confidentiality considerations or trade secrets frustrate the implementation of a meaningful human rights impact assessment. Where private sector actors provide services that rely on algorithmic systems and that are considered essential in modern society for the effective enjoyment of human rights, member States should preserve the future viability of alternative solutions and ensure the continued access to such services by affected individuals and groups. 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, and should contain clear recommendations on how to prevent or mitigate the high risks to human rights.

5.3 **Expertise and oversight:** States should ensure that all human rights impact assessments related to high-risk algorithmic systems be submitted for independent expert review and inspection. Tiered processes should be created for independent oversight. Human rights impact assessments conducted by or for States should be 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.

5.4. **Follow-up:** In circumstances where the human rights impact assessment identifies significant human rights risks that cannot be meaningfully mitigated, the algorithmic system should not be implemented or otherwise used by any public authority. If the risk is identified in relation to an algorithmic system that has already been deployed, implementation should be discontinued at least until adequate measures for risk mitigation have been taken. Identified human rights violations should immediately be addressed and remedied, and measures adopted to prevent further violations.

5.5 **Personnel management:** States should ensure that all relevant staff involved in the procurement, development, implementation, assessment and review of algorithmic systems with significant human rights impacts are adequately trained with respect to applicable human rights and non-discrimination 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 gender-equal and 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.

5.6 **Interaction of systems:** States should carefully monitor settings where multiple algorithmic systems operate in the same environment to identify and prevent negative externalities, particularly where their possible interdependencies and interactions require a precautionary approach. In their public service delivery, States should utilise the mechanism of procurement or engagement of private services with full regard of the need to maintain oversight, know-how, ownership and control over the use of algorithmic systems and their interaction with each other.

5.7 **Public debate:** States should engage in and support ongoing, 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 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 may include 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 steps 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, including to ensure that the interests of marginalised and vulnerable individuals and groups are adequately taken into account and represented. Where appropriate, this may require the discouragement of influences that may exclusively favour most commercially viable optimisation processes. State should ensure the adequate protection of whistle-blowing or other actions by employees engaged in the development or ongoing deployment of algorithmic systems who perceive a need to notify regulators and/or the public about present or possible failures to maintain human rights standards in the systems they have been tasked to build.

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 and climate challenges.

6.4 **Independent research:** States should initiate, encourage and publish independent research to monitor the societal and human rights implications of the ongoing deployment of algorithmic systems. In addition, such independent research should study 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.

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

### 1 Principles of general application

1.1 **Responsibility to respect human rights:** Private sector actors engaged in the design, development, sale, deployment, implementation and servicing of algorithmic systems, whether in the public or private sphere, must exercise human rights due diligence. They 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 actions, including innovation processes, respect human rights and be mindful of their responsibility towards society and the values that make it democratic. Efforts to ensure human rights compliance should be documented.

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:** Owing to the horizontal effect of human rights and given that design, development and ongoing deployment of algorithmic systems engage private sector actors in very close cooperation with public actors, some of the key provisions that are outlined in Chapter B as obligations of States translate into legal and regulatory requirements at national level and 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 relevant standards contained in provisions 1.2, 1.3, 2.1, 3.1, 3.3. and 4.2 of Chapter B related to ongoing review, participation and awareness, informational self-determination, computational experimentation, testing and identifiability of algorithmic decision-making.

1.4 **Discrimination:** Private sector actors that design, develop or implement algorithmic systems should follow a standard human rights due diligence framework to avoid fostering or entrenching discrimination through all lifecycles of their systems. They should seek to ensure that the design, development and ongoing deployment of their algorithmic systems do not have direct or indirect discriminatory effects 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 informed of and empowered with the choice to give and revoke their consent regarding all uses of their data, with both options 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 the 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, and should not be embedded in terms of services.

2.2 **Privacy settings:** Private sector actors should facilitate the right of data subjects to protect effectively their privacy while maintaining access to services. The possibility of choosing from a set of privacy setting options should be presented in an easily visible, neutral and intelligible manner and facilitate the use of privacy enhancing technologies. Default options should lead only to the collection of data that are necessary for and proportionate to the specific legitimate purpose of the data processing, while tracking settings should be set as default in optout mode. 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 and model quality:** 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 errors, bias and potential discrimination in datasets and models 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 sufficiently diverse and 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. **Systems and data security:** Private sector actors should configure their algorithmic systems in such a way that it prevents illegal access, system interference and misuse of devices, data and models by third parties in line with applicable standards.

## 4 Transparency, accountability and effective remedies

4.1 **Terms of service:** Private sector actors should ensure that the use of algorithmic systems that can trigger significant human rights impacts 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, prominent and plain language and in accessible formats. Adequate information about the nature and functionality of the algorithmic system should be provided to allow for meaningful contestation and objection. Terms of service should be reasonably concise, easily understandable and contain clear and succinct language about possibilities for users to manage settings. This should include information 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:** 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 relevant staff involved in the handling of customer complaints should be suitably versed in relevant human rights standards and benefit from regular training opportunities.

4.3. **Transparency:** 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, and the outcomes of the contests, 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.

4.4 **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 or ongoing use of a system with potential for human rights violations, or remedy a violation of rights. The scope of available remedies may not be limited. If prioritisation is necessary and as delays in response may affect remediability, the most severe human rights impacts should be addressed first. All complaints 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 and regulatory, review mechanisms. No waivers of rights or hindrances to the effective access to remedies

should be included in terms of service. Business associations should further invest – in cooperation with trade associations – in the establishment of model complaints mechanisms.

4.5 **Consultation:** 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, on the design, development, ongoing deployment and evaluation of algorithmic systems, as well as on their complaint mechanisms.

## 5 Precautionary measures

5.1 **Continuous evaluation:** Private sector actors should develop and document internal processes to ensure that their design, development and ongoing deployment 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 generate. Where the application of algorithmic systems carries high risks to human rights, including through processes of micro-targeting which they can avoid or mitigate themselves, private sector actors should have the possibility to 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 in question. Private sector actors should submit these algorithmic systems for regular independent expert review and oversight.

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 with the active engagement of affected individuals and groups. In case of deployment of high-risk algorithmic systems, the results of ongoing human rights impact assessments, 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 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 throughout the full lifecycle of the algorithmic system 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 assurance checks and real-time auditing through design, testing, and deployment stages. It further requires regular consultation with affected individuals to monitor algorithmic systems for human rights impacts in context and in situ, and to correct errors and harms appropriately and in a timely manner. This is particularly important given the risk of feedback loops that can exacerbate and entrench adverse human rights impacts.

## 6 Research, innovation and public awareness

6.1 **Research:** Private sector actors should engage in, fund and publish research, conducted in line with research ethics, 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 concern 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. Private sector actors should find effective channels of communication with local civil society groups, particularly in geographic areas where human rights concerns are high, in order to identify and respond to possible risks related to the deployment of algorithmic systems.

6.2 **Access to data:** For the purposes of analysing the impacts of algorithmic systems and digitalised services on the exercise of rights, on communication networks, and on democratic systems, private sector actors should extend access to relevant individual and meta-datasets, including access to data that has been classified for deletion, to appropriate parties, notably independent researchers, media and civil society organisations. This extension of access should take place in full respect of legally protected interests as well as all applicable privacy and data protection norms.