Ethics by design and international soft and hard standards on the nexus gender-artificial intelligence*

Cristiana Carletti

Abstract

The contribution is intended to debate over the nexus between gender and artificial intelligence as for programs and systems based on AI which could produce, if the ethics by design is not approached according to a gender perspective, gender biases. The need for overcoming this criticality rests upon the need for improving the presence and participation of the female component in the design, development and implementation of the aforementioned programs and systems, in digital teams as members or leaders, to contribute for the elaboration of technical solutions within a legal framework which is aimed to translate current soft standards in force into hard laws.

Summary

Keywords

gender - artificial intelligence - ethics by design - soft law - hard standards

1. Setting the scene and the need for a gendered ethics by design towards hard laws

The speediness in the automated processing and use of digital technologies, with particular reference to artificial intelligence, is a social reality. Public and private actors, in charge for the advancement of studies, research and analysis on methodologies of data collection, storage and management, have embraced this challenge since early

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

years of this century and, in pandemic times, have confirmed this approach for the identification of organizational solutions for restructuring and making operational their respective systems through a series of basic algorithms in artificial intelligence programs[1]. Regardless of the meaning attributed to algorithms used in such programs[2], it is clear that their functionality rests on the quantitative and qualitative nature of data under analysis as input data, for producing result as output data[3]. It is the first component of digital data, its volume, that has a decisive impact on the creation of biases, due to subjective choices when selecting, collecting and analyzing data, which is not isolate but rather continuous along the setting of algorithms that allow artificial intelligence programs to operate[4].

Indeed, next to the definition of data categories analyzed by algorithms as for their generality, accuracy, reliability in an objective way, it is the setting of this operation by the designer that could contribute to the occurrence of technical or intentionally dependent biases in the subjective analysis and interpretation of data. This depends upon the setting of algorithms' inquiry to produce categorical output data, which is limiting as to information in terms of quantity - as opposed to the analysis on overall data - as well as quality. At the same time the subjective factor affects the technological design, implying a potential duplication of the analysis of the phenomenon in a biased and discriminatory perspective.

The nexus between the potential development of digital knowledge and its use in such a way as to result in a discriminatory impact in a broad sense, if not also in the gender dimension, has not yet promoted a careful legal analysis such as to anticipate the dynamics inherent in the management of rights and freedoms compressed in an individual capacity as well as the activation of the competencies of judicial and para-judicial bodies for the purpose of remedy in favor of the injured parties.

Some attempts at normative production have been characterized by a soft relevance, distinguishing the commitment to the protection of standards referable to the protection of fundamental rights in charge of public and private actors, although always in a collaborative perspective.

These preliminary exercises are leading, with some effort, to the compilation of secondary legislation in the European Union system and complex legal instruments with binding impact in the framework of the Council of Europe, focused on the best ways to regulate technological apparatuses, particularly those generated and managed through artificial intelligence.

In the first case, the proposed Regulation on Artificial Intelligence (so-called Artificial Intelligence Act) proposed by the European Commission in April 2021, on which the European Parliament approved its negotiating position on June 14, 2023, appears particularly important[5].

---

[1] L. Downey, *Algorithms*, Investopedia, 2021.

[2] J. Guszcza, *Smarter together: Why artificial intelligence needs human-centered design*, Deloitte, 2018.

[3] J. Denny, *What is an algorithm? How computers know what to do with data*, The Conversation, 2020.

[4] A. Manasi - S. Panchanadeswaran - E. Sours - S.J. Lee, *Mirroring the bias: gender and artificial intelligence*, in *Gender, Technology and Development*, 26, 2022, 295 ss.

[5] EU Commission, Communication from the Commission to the European Parliament, the Council, the

The human-centric interpretation offered by the EU institutions rests on the *acquis* of rights and freedoms set forth in the Charter of Fundamental Rights, subject to severe limitations depending upon the use of artificial intelligence in 'high-risk' situations: respect for human dignity, private and family life, protection of personal data, freedom of expression and information, assembly and association, the principle of non-discrimination, and several set of individual and collective rights also relevant in the social and economic domains as well as the judicial system, including in these cross-references gender equality.

It is precisely the gender component (encompassing gender identity and sexual orientation, race, ethnic origin, migratory status, political or religious orientation, or otherwise other discriminatory factors) to be expressly mentioned since the degree of risk and intrusiveness of artificial intelligence-based systems determines profiling by reason of highly sensitive elements that are altered with extreme ease: for example biometric data, personal choices related to the educational and training system up and professional opportunities preferred and pursued by women and girls.

Following these considerations, the European Parliament asserted that «diversity, non-discrimination and fairness' means that AI systems shall be developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law».

On the other hand, the Council of Europe system has set its reasoning in a similar and parallel way, resting on the need for a human rights-based approach in the process of compiling a binding legal instrument dedicated to artificial intelligence and algorithmic technologies[6].

In order to ensure a reinforcing shift for human rights' protection from soft to hard law, the Committee tasked with analysing legal prerequisites underpinning such an instrument recommended the need to include «a provision on respect of equal treatment and non-discrimination of individuals in relation to the development, design, and application of AI systems to avoid unjustified bias being built into AI systems and the use of AI systems leading to discriminatory effects».

The translation of this recommendation into a high-impact binding provision could be set up firstly by introducing an obligation on public and private actors to ensure that artificial intelligence and algorithmic systems are designed to promote the principle of non-discrimination, thus also gender equality; additionally, the obligation requires a wider substantial perspective to include the attribution of a definite mandate to equality bodies, ombudspersons, and independent national human rights institutions

European Economic and Social Committee and the Committee of the Regions, Fostering a European approach to Artificial Intelligence; EU Commission, Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts; as for the EP position, Draft Compromise Amendments on the Draft Report, Doc. KMB/DA/AS, 16 May 2023, Decision by Parliament, 1st reading, 14 June 2023.

[6] As examined further on in the contribution, see Council of Europe-CAHAI Feasibility Study on a legal framework on AI design, development and application based on Council of Europe standard (2020) and Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law (2021) – see note 23.

in monitoring and identifying the impact of algorithms in a discriminatory logic; the adoption of a proportionality test that technically and punctually assesses algorithmic metrics from a discriminatory perspective; the introduction of a general obligation of transparency to value the fairness of technical solutions adopted by digital professionals in the definition of artificial intelligence-based mechanisms; and finally the formulation of an indirect obligation – otherwise positive obligations - for the purpose of adopting preventive assessing tools over algorithmic discriminatory biases in the framework of policies aimed at achieving *de jure* and *de facto* equality.

Such operational proposals, when transposed into a legally binding instrument, could be the best legal prerequisite for incentivizing not only public authorities but also private actors to collaborate for the elaboration and adoption of 'equality by design' technical models.

In this latter perspective, gender biases emerge when, in identifying artificial intelligence programs running, possibly similar to human reasoning, the preeminent reference model is a male one[7] since it is theoretical and does not ponder intuitive and emotional factors generally attributed to female thinking and aptitudes[8]. Moreover gender biases are even more evident in relation to technological knowledge: this research field has always been considered a preferred one for a predominantly male presence, as opposed to a female component that is not sufficiently ready to acquire technological skills and abilities[9]. This has fostered a vision that is no longer only factual but also perpetuated in the digital context in favor stereotypical and prejudicial social models of the female role and image, especially when they refer to limited opportunities for women and girls to participate in and contribute to the development of technological knowledge related to artificial intelligence[10].

More specifically, gender biases are a category to be further carefully explored calling for the above mentioned need for a proper comprehensive legal framework: they can be produced either in the process of setting algorithms through the use of word embeddings in the programming language that contain sexist formulas, or in data collection and storage even as it relates to the monitoring and evaluation of results of such activities - since they can incentivize a discriminatory appraisal based on factors such as gender, age, ethnic origin, religion, political or sexual orientation, or even in decision-making on the basis of the artificial intelligence programs in place[11]. It is precisely at these stages, with particular emphasis placed on the latter one, that the no gender-neutral consideration about knowledge and application of digital technologies clearly emerges.

In fact, structural and operational disparities related to the lack of the gender component in science and technology fields result in the design and implementation of appli-

[7] H. Schelhowe, *Paradigms of computing science: The necessity for methodological diversity*, in *Gender, Technology and Development*, 8, 2004, 321 ss.

[8] D.M. Sutko, *Theorizing femininity in artificial intelligence: a framework for undoing technology's gender troubles*, in *Cultural Studies*, 34, 2020, 567 ss.

[9] D. Johnson, *Sorting out the Question of Feminist Technology*, University of Illinois, 2010.

[10] A.R. Fryxell, *Artificial Eye: The Modernist Origins of AI's Gender Problem*, in *Discourse*, 43, 2021, 31 ss.

[11] A. Manasi, *Addressing Gender Bias to Achieve Ethical AI*, IPI Global Observatory, 2023.

cation of artificial intelligence-based programs and systems that are at all gender-neutral. The limited presence of the female component in ICTs-related fields is for sure a consequence of the not at all gender-inclusive approach to human capital - in the face of a growing female percentage in STEM disciplines[12] - and the allocation, according to traditional norms, of care commitments to women and girls, this implying fewer opportunities to access professional careers in ICTs.

The need to incentivize a gender appraisal with respect to ICTs and, in particular, to programs and systems based on artificial intelligence, arose visibly during a debate promoted by UNESCO in 2020 dedicated to the relationship between gender equality and artificial intelligence, for an interpretation designed to promote the definition of principles and legal instruments inspired by an ethical approach aimed at preventing and countering any form of gender discrimination in the digital field.

In this framework the ethical factor takes on a proper connotation: in general, linked to artificial intelligence, it is instrumental for a clear distinction between the human being and the machine: oftentimes, artificial intelligence programs and systems are seen as tools to overcome this distinction on the basis of the large volume of data to be collected, managed, and analyzed - a difficult task for the human mind and, at the same time, artificially solvable through a series of complex machine functions – including also developmental potential and operational flexibility of artificial intelligence[13], which seems quite objective and able of introducing social biases[14].

As above reported, in relation to an ongoing process to negotiate hard standards at the EU/Council of Europe level, in the larger debate promoted at UNESCO[15], since the adoption on November 24, 2021 of the Recommendation on the Ethics of Artificial Intelligence, the active involvement and reception of female contribution in the definition and implementation of a set of ethical principles for the elaboration of artificial intelligence-based programs and systems has been considered crucial. In this sense, the expression ethics by design has been introduced: the dual function of regulation through standards and design of artificial intelligence mechanisms inspired first and foremost by respect for human rights in terms of autonomy, dignity, and freedom, with specific reference to the right to privacy and protection of personal data. Ethics is also essential for the operation of the aforementioned programs and systems in such a way that everyone has equal access to them, enjoying equal rights and equal opportunities, and feels protected as individuals or members of a community, albeit a digital one. Programs and systems, ethically, must be made transparent and accessible

[12] E. Davila Dos Santos - A. Albahari - S. Díaz - E.C. De Freitas, *Science and Technology as Feminine': Raising awareness about and reducing the gender gap in STEM careers*, in *Journal of Gender Studies*, 31, 2022, 505 ss.

[13] A. Gilli - M. Pellegrino - R. Kelly, *Intelligent machines and the growing importance of ethics*, in A. Gilli (ed.), *The brain and the processor: Unpacking the challenges of human-machine interaction*, NATO Defense College, 2019, 45 ss.

[14] R. Benjamin, *Race after technology: Abolitionist tools for the new JIM code*, in *Social Forces*, 98, 2020, 1 ss.; *contra* the idea of flexibility, translated into clarity of purpose and choice, undoubtedly attributable to the human being and not to the machine see above, note 13.

[15] L. Hogenhout, *A Framework for Ethical AI at the United Nations*, UN Office for Information and Communications Technology, Unite Paper, 2021(1).

so that all digital users can know their design and application as well as verify their functioning by assigning specific responsibility to designers if it does not comply with ethical principles. So far ethics by design is indispensable for the compliance with these principles, as a key operational prerequisite for artificial intelligence-based programs and systems and also for the future compilation of dedicated legally binding instruments – moving from the regional to the global level.

For the ethical principles on which the design process rests to be concretely validated, the functioning of any program or system based upon artificial intelligence requires: the predetermination of objectives; the definition of technical and non-technical requirements; a complex design that is nevertheless qualitatively such as to ensure compliance with principles; the operation of data collection, storage and management for its integrity and reliability; the possible development of additional design elements, sufficiently flexible and adaptable to the model; and interventions to verify and evaluate the program or system during its functioning.

While these elements may appear to be primarily technical or otherwise abstract, the specific relevance of the ethical component emerges, even from a gender perspective, when the artificial intelligence-based program or system is able to operate in full compliance with them through the cognitive contribution of the female component to its design and proper functioning.

2. Recommending a gender-based approach in the digital space

The elaboration of digital tools, including those based on artificial intelligence, that are truly gender-neutral is a topic addressed in the United Nations framework to promote a process of legal regulation supported by both member states and actors of a non-institutional nature, particularly ICTs' companies. This process is yet framed along the lines of soft law documents, due to the legal fatigue and eventual barriers for a global support for a dedicated binding treaty over digital issues at large from States but also to the need for a motivated inclusion of private actors to provide their contributions and to accept their 'compliance' to hard standards.

In the most recent considerations shared by the Secretary-General anticipating the 67th session of the Commission on the Status of Women in 2023[16], it is suggested that the compilation of voluntary ethical standards could be a starting point for such a process: they will be able to identify conducts and activities of digital actors as producers of programs and systems so that they are instrumental to both the development and proper functioning of technological tools, particularly those based on artificial intelligence.

In order to ensure the effective impact of these ethically-driven voluntary standards, it will be essential to correlate them with monitoring and evaluation procedures that

[16] For further details see UN, Commission on the Status of Women, Sixty-seventh session, *Innovation and technological change, and education in the digital age for achieving gender equality and the empowerment of all women and girls,* Report of the Secretary-General. E/CN.6/2023/3, 30 December 2022.

cannot be internal to digital actors' frameworks (especially if private) as they could not sufficiently guarantee independent assessments, nor external since they cannot quickly activate the removal of digital content that does not meet the standards or decisively affect ethics by design to correct digital contents.

However, as for the development of digital technologies, a further regulatory step can be shaped, as in the European context, by providing for the compilation of a binding legal framework by digital actors so that: any content that does not comply with it is removed, due diligence is introduced to prevent and manage any risks arising from technological devices to the detriment of users - also from a gender perspective, transparency is ensured in the sharing of information from programs and systems based on artificial intelligence, and moderation methodologies are used about digital contents.

The ultimate goal of this process lies in the introduction and implementation of mandatory standards, defining in a timely manner obligations and responsibilities on digital actors, including the attention paid to the truly gender-neutral dimension of ICTs. It is precisely the volume of digital data that presents, along a gender perspective, limited quantitative relevance as users suffer from an obvious gender digital divide as well as unequal to other categories for processing the same data in disaggregated form. These two factors affect the functioning of artificial intelligence-based programs and systems, making cognitive biases permanent, compressing the quality of digital services, and incentivizing discriminatory appraisal of data.

With particular reference, once again, to the ethical component of programs and systems based on artificial intelligence, it is usual to recall benefits and criticalities produced by different types of biases, including gender biases[17]. The latter, in particular, depend on transferring prejudicial behaviors and stereotype-based conducts from factual to virtual reality and are produced by the limited female promotion and access to STEM disciplines, digital careers and the opportunity to enter technical teams designing automated or artificial intelligence-based technological tools.

The digital gender divide encompasses all forms of obstacles in accessing and attending educational and training paths - from the primary level up to the academic and postgraduate specialization ones - dedicated to technologies and in gaining cultural and social experience about concrete limitations of developing knowledge and skills in order to benefit from the progress stemming from digital transformation and innovation, available through all devices - from smartphones to laptops and access to the web.

The primary consequence of this consideration are concrete limited opportunities for women and girls as scientific or entrepreneurial components or team-leaders to enter the professional field of new digital technologies, and to experiment in the sub-sector of artificial intelligence. Thus, there is not only a segregation of a horizontal nature, if statistical data confirm digital educational and professional disparity between men and women, but also a segregation of a vertical nature related to overcoming the main ob-

[17] E. Lamm - G. Ramos - E. Ronchi - M. Squicciarini, *The Gendered Impacts of AI: Policies and Safeguards to Regulate New Technologies, Mitigate Risks and Protect Rights*. UN Women, Expert Group Meeting 'Innovation and technological change, and education in the digital age for achieving gender equality and the empowerment of all women and girls', 10-13 October 2022.

stacle of access to the digital sector, that is the concrete overcoming the glass ceiling by reaching top positions in the tech sector. In addition to these observations, there is also an additional risk for female workers in the labor market: due to their compressed representation, they could suffer a further form of exclusion resulting from the wider use of automated production mechanisms that will mainly affect professional figures with low and medium levels of education and skills.

3. Innovation and technological change, and education in the digital age for achieving gender equality and the empowerment of all women and girls

While the issue under consideration has been addressed in the United Nations system through a comprehensive analysis of the evolution of the digital sector and the need to introduce regulatory measures of both a voluntary and mandatory nature to prevent and manage different types of biases of artificial intelligence-based programs and systems, specific attention was paid to the gender dimension on the occasion of the 67th session of the United Nations Commission on the Status of Women, held in March 2023 in New York.

The annual session of the Commission requires the predetermination of the so called priority theme, on which debates will be scheduled, including the political and high-level level as well as in technical and interactive format and the dialogue with civil society organizations, which are called upon to provide their input in the framework of informal events and meetings.

The priority theme of the 67th session of the Commission focused on innovation and evolution of technologies, linked to education in the digital age with the goal of achieving gender equality and empowerment of all women and girls.

In the document adopted at the end of the session, the so-called Agreed Conclusions[18], which does not have a binding nature and yet contains interesting recommendations addressed to member states as well as all non-institutional actors who are affected by the issues examined and discussed by the Commission, the priority theme is articulated in its general scope and in order to propose an appraisal that, in the case under consideration, also drew attention to the nexus between gender and artificial intelligence.

In the preambular section of the Agreed Conclusions, the Commission notes the continuity of the gender discriminatory approach and the creation of biases by moving from the real to the virtual context: biases, specifically, are produced by the use of algorithms in artificial intelligence-based programs and systems.

While it is true that such programs and systems have had an important and positive impact in the configuration and activation of new public services, in economic pro-

[18] See UN, Commission on the Status of Women, Sixty-seventh session. Agreed Conclusions. 20 March 2023. Indeed the Agreed Conclusions could be considered as a legal starting point official document to promote the compilation of other relevant UN legal documents, which have a recommendatory relevance or could be retained as a key-tool to encourage the negotiation of legally binding legal instruments.

gress and for collective and extensive social welfare, and in the readjustment of work settings especially during and after the pandemic emergency, at the same time there have been and will continue to be negative consequences on the personal and professional lives of women and girls.

To counter this trend, appropriate emphasis must be placed on efforts to overcome structural and systemic cultural and societal stereotypes and biases that slow down women's and girls' access to STEM disciplines and digital careers. Their presence as female researchers, innovators and entrepreneurs, members and leaders of teams created in the public and private sectors in the digital field is an essential precondition to bring their contributions, in terms of knowledge and experience in the sub-sectors of artificial intelligence, software programming, cloud computing platforms, and data management. The design, development, and application of digital technologies that rest on the neutral and objective collection, storage, and management of data needs such input in order to prevent technical malfunctions and biases inherent in algorithms since the setting of collecting and analyzing data is not really neutral along the gender perspective.

These considerations have been translated into recommendations to the attention of all stakeholders, both public and private, in the operational section of the Agreed Conclusions. In this regard, the adoption of public policies aimed at promoting gender equality and equal opportunity in STEM disciplines is essential to encourage women and girls on their path to employment and professional growth. At the same time, private actors are recommended to adopt technological investment methodologies that give due consideration to preventing and countering discriminatory behaviors and conducts produced by programs and systems based on artificial intelligence, predictive algorithms, and robotics. Finally, a common recommendation addressed to both public and private actors is aimed at adopting measures to regulate and evaluate basic requirements underlying the aforementioned programs and systems for the better prevention and management of gender biases.

Indeed, these considerations are echoed in the aforementioned Recommendation adopted at UNESCO on the Ethics of Artificial Intelligence[19], which can thus be considered a relevant tool for the compilation of useful standards to favour the adoption of a gender-responsive approach to artificial intelligence as a key-precondition towards the elaboration and compilation of dedicated hard standards. In fact in principle automated mechanisms have a potentially discriminatory impact if they are not designed technologically in an appropriate way: the resulting consequences pervade the social sphere and the female component both in positive terms - for example increasing educational knowledge in digital matters, flexible work solutions, acquisition of knowledge for access to financial resources – as well as in negative terms as for labour and wage management and for the highest risk of exposure to all forms of real and virtual violence and harassment over women and girls as key-victims.

The gender dimension is addressed in Policy Area 6 of the document, whose recommendations to member states focus on the need to ensure that digital technologies in

[19] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 24 November 2021; see also UNESCO, *Artificial intelligence and gender equality*, 2020.

a broad sense, and artificial intelligence specifically, in the duty cycle of programs and systems that rely on it, are instrumental in promoting gender equality.
The governmental authorities are first and foremost required to allocate adequate financial resources in favor of the female population, within the framework of appropriate strategic plans dedicated to the digital topic and the use of related technologies, in support of interventions in the educational and professional sectors.
From a social point of view, in a participatory view by the female component in the definition of policies to overcoming gender gaps, it is important to incentivize the presence and contribution of women and girls in digital teams, both as components and leaders, so that ethics by digital design incorporates the gender factor in an appropriate way for the functioning of programs and systems based on artificial intelligence. Such participatory involvement is instrumental for the prevention and management of gender biases, thus relying on technological knowledge and skills to overcome educational and professional stereotypes and biases effectively and systematically.
Public-private collaboration is also central in the Recommendation, with the goal of identifying and technologically removing gender biases produced by artificial intelligence-based programs and systems when they not only alter the recognition and acceptance of gender diversity but also when they seriously endanger women and girls in the virtual context. This echoes the need, identified at the regional level in the Council of Europe reasoning towards the compilation of a legally binding instrument over AI and algorithms, for an enhanced dialogue and collaboration between public institutions – as rulers – and private companies – as digital technicians – to prevent discriminatory and gender biases through targeted policies and related positive actions.
Finally, member states are urged to encourage the female component in the private sector engaged in the development of digital technologies, facilitating her entry into this male-dominated professional career and subsequent advancement to top positions, and supporting her in accessing financial incentives for this purpose.

4. Towards an AI hard standard setting to preventing and countering discrimination and fostering gender equality

In a complementary view to the UN framework, where the digital topics been translated into potential voluntary rules and standards, even when formulated as recommendations addressed to member states, at the regional intergovernmental level main steps have been promoted towards the compilation of norms specifically devoted to artificial intelligence, appropriately recalling the relevance of the gender component and digital gender biases.
As above recalled, in the Council of Europe, the topic has been debated since May 2019 when the Commissioner for Human Rights adopted a document of a recommendatory nature to the attention of the membership to define the nexus between artificial intelligence and human rights[20]: this document has had the aim not only of

[20] Council of Europe, Commissioner for Human Rights, *Unboxing artificial intelligence: 10 steps to protect*

incentivizing the development of artificial intelligence-based programs and systems but also of preventing or mitigating the negative impact they may have on individual and collective life and rights and freedoms.
In order for this complex goal to be satisfactorily achieved, public stakeholders acquiring, developing, and applying artificial intelligence-based programs and systems are urged to introduce procedures to assess their impact on human rights, to ensure broad and transparent information on the ways through which programs and systems are designed, and to regulate in a timely manner the legislative system that enables them to carry out independent and effective control over the development, dissemination, and use of programs and systems by both public and private actors in terms of their impact on the protection of human rights.
The relevance of the topic led to an important step, namely the creation of a special committee (Ad Hoc Committee on Artificial Intelligence - CAHAI) that exercised its mandate from 2019 to 2021 with the main purpose of testing the possibility of compiling a dedicated binding legal instrument, carrying out a series of multi-stakeholder consultations and producing interesting background papers.
CAHAI has reasoned on the basis of multiple legal standards: first, those of binding and non-binding legal scope adopted in the Council of Europe and appropriately related to the process of design, development and application of digital technologies with respect to the protection of human rights, democracy and the rule of law as fundamental pillars of the Organization since its establishment in 1949; then additional legal instruments of binding and non-binding scope adopted in other intergovernmental global and regional systems. In exercising its mandate, again, CAHAI has paid special attention to the gender dimension.
In an early study paper[21], the Ad Hoc Committee placed the nexus between gender and artificial intelligence in the broader context of promoting principles of equality, non-discrimination and solidarity, specifically emphasizing the scope of the provisions of the European Convention for the Protection of Human Rights and Fundamental Freedoms with reference to Article 14 and its Protocol No. 12, focused on the principle of non-discrimination, albeit gender-based.
Artificial intelligence-based programs and systems have perpetuated and increased discriminatory conducts and behaviours from the real to the virtual context, limiting its monitoring and control and incentivizing biases, regardless of whether produced by mere technological error or consciously and intentionally by the designers of artificial intelligence systems, thus confirming that AI-design is neither neutral nor ethical. The presence and technical contribution of the female component in the design teams of these systems is essential to prevent and manage biases, in the form of multiple discriminations up to violence and harassment against women and girls.
Complementing these remarks of CAHAI, in order to give an overview of binding and non-binding international and regional instruments in force, the opportunity to

human rights, 2019.

[21] Council of Europe, *Towards regulation of AI systems. Global perspectives on the development of a legal framework on Artificial Intelligence (AI) systems based on the Council of Europe's standards on human rights, democracy and the rule of law,* Compilation of contributions DGI (2020) 16.

draft a convention to regulate artificial intelligence-based mechanisms beyond soft standards was explored by some experts who compiled and presented an interesting study paper in 2021, thus offering interesting insights to the attention of the same Ad Hoc Committee in charge for the preparation the feasibility study towards a legal framework on artificial intelligence[22].

Drawing from the extensive production of guidelines and principles for the ethical configuration and application of artificial intelligence-based programs and systems, the experts pointed out that the voluntary or self-regulatory nature was necessary at the outset for a generic, flexible and adaptable as well as reviewable language and yet led to criticisms related to a rhetorical approach and the practical difficulty of introducing ethical foresight into technological mechanisms, especially when automated. Moreover, experts have differentiated the nature of soft laws adopted by Council of Europe bodies - such as recommendations and declarations or of guidelines with input from all stakeholders - or implemented by member states often on the basis of a public-private partnership (guidelines, codes of conduct).

However, given the extreme dynamism of technological knowledge and the high risk of biases produced by limited ethics by design, the chance of either revising existing binding legal instruments in the digital domain or, otherwise, compiling a dedicated binding legal instrument on artificial intelligence were both considered as necessary and viable options in the Council of Europe.

In the former case, several alternatives were envisaged: the compilation of a Protocol to the aforementioned European Convention, binding on the States Parties and potentially impacting on the backlog of the European Court of Human Rights; and the revision of existing binding legal instruments, such as the Budapest Convention on Cybercrime or Convention 108+ on the Management of Personal Data, although the specificity of artificial intelligence-based programs and systems would require an adaptation of the monitoring and verification mechanisms to automated mechanisms that already operate within the framework of such legal frameworks.

In the latter case, two options have been proposed as to the drafting process.

With respect to the option for the drafting of a framework convention that would lay out basic principles and areas of implementation inherent in the design and functioning of artificial intelligence-based programs and systems, it has been appreciated for the rapid evolution of digital knowledge and technological tools and ethical challenges; a framework convention, however, could leave room for the states parties in terms of obligations for its implementation, possibly framed through additional protocols.

The option for the drafting of a convention would undoubtedly ensure the elaboration of a detailed legal discipline, formulating rights and obligations that could be perceived as too inflexible with respect to technological and digital dynamics and also in terms of domestic legal compliance.

The latter option has been accommodated by the membership of the Council of Europe, promoting this process through the creation, in place of the CAHAI, of the Committee on Artificial Intelligence (CAI), tasked with the compilation of the afore-

[22] D. Leslie - C. Burr - M. Aitken - J. Cowls - M. Katell - M. Briggs, *Artificial intelligence, human rights, democracy, and the rule of law. A primer*, The Alan Turing Institute, 2021.

mentioned binding legal instrument by November 2023.
Indeed, the CAI has worked in line with the document produced by CAHAI at the end of its mandate[23] in which a number of elements have been proposed to underpin the drafting of this new convention.
Indeed, in the document adopted by CAI in 2023, to be considered as the preliminary text of the future Artificial Intelligence Convention[24], an analysis focused on the gender dimension is proposed.
CAHAI had already planned to include a gender reference in the provision concerning the guarantee of equal treatment and respect for the principle of non-discrimination with regard to the design, development and application of systems based on artificial intelligence; for a specific management of gender biases, the Ad Hoc Committee had also noted the opportunity to draft additional provisions concerning specific categories of subjects, including women and girls, who are directly affected by artificial intelligence mechanisms and who, for this reason, must be able to participate in the elaboration of monitoring and control procedures regarding their proper functioning.
The CAI incorporated these indications and formulated the principle of non-discrimination (as of today, Art. 3), introducing factors that artificial intelligence-based systems could entail to operate with discriminatory impact: «sex, gender, sexual orientation, race, color, language, age, religion, political or any other opinion, national or social origin, association with a national minority, property, birth, state of health, disability or other status, or based on a combination of one or more of these grounds».
An additional reference of this principle is included in Art. 12, concerning equality and anti-discrimination, which provides for an obligation on States Parties to ensure that «the design, development and application of artificial intelligence systems respect the principle of equality, including gender equality and rights related to discriminated groups and individuals in vulnerable situations».
In conclusion, apart from a different but complementary approach in dealing with this topic in intergovernmental global and regional systems, it is quite clear the relevance of the issue of gender biases produced by artificial intelligence-based programs and systems and the need to seize this opportunity for drafting legally binding instruments beside soft laws as a further step to ensure the adoption and implementation of hard regulations in this matter.

[23] As mentioned above Council of Europe, *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law*, 2021.

[24] Council of Europe, *Revised zero draft [framework] convention on artificial intelligence, human rights, democracy and the rule of law*, 2023.