

OPINION ON THE IMPACT OF ARTIFICIAL INTELLIGENCE ON FUNDAMENTAL RIGHTS

7 APRIL 2022



Opinion on the impact of artificial intelligence on fundamental rights

was adopted unanimously at the plenary session of 7 April 2022.

SUMMARY.

While research on artificial intelligence (AI) and the implementation of its practical applications are developing, current regulations remain incomplete when it comes to preventing possible major infringements of fundamental rights. In the context of the forthcoming adoption of the proposal for an EU Regulation on the subject, and the work being done in the Council of Europe, the National Consultative Commission on Human Rights (CNCDH) is calling on the public authorities to promote an ambitious legal framework in this area.

On the one hand, it recommends prohibiting certain uses of AI considered to be too harmful to fundamental rights, such as social scoring or remote biometric identification of people in publicly accessible spaces. On the other hand, it recommends placing on users of an AI system requirements that can guarantee respect for fundamental rights: an impact assessment, stakeholder consultation, and supervision of the system throughout its life cycle. The CNCDH finally calls for the recognition of rights for persons who have been the subject of a decision involving an algorithm, in particular the right to human intervention in the decision-making process, or a right to configure the operating criteria of the AI system.

TABLE OF CONTENTS

Summary.	2
Introduction.	4
1. Red lines.	8
1.1. The contributions and limitations of the prohibitions laid down by the proposal for an EU regulation.	9
1.2. The need to extend the ban to other areas.	12
2. A framework that guarantees respect for fundamental rights.	14
2.1. Control at all stages of AI system development.	14
2.2. Guaranteed respect for fundamental rights with regard to individual decisions.	22
Summary of recommendations.	29
List of people heard.	32

INTRODUCTION.

1. For some time now, what is now commonly called “artificial intelligence” or “AI” is at the heart of each individual’s daily life: suggested content on social networks or online platforms, access to applications or locations using biometric authentication, automated medical diagnoses, etc. It has however received a great deal of attention in recent years, by taking advantage of massive public and private investment. While some place it at the heart of a “*new industrial revolution*”, others are concerned about a new wave of automation of activities previously reserved for human beings, and more generally the downward spiral of new governance by data, as well as, more broadly, about the possible major infringements of fundamental rights, not to mention its growing impact on the environment.

2. As a starting point, the National Advisory Commission on Human Rights (CNC DH) wishes to express its reservations with regard to the terminology used in this area. In fact, it observes an excess of anthropomorphisation in the terms used, starting with “artificial intelligence”, but also “neural networks”, “deep learning”, etc. This creates confusion about the real possibilities offered by data processing systems, which are based on procedures coded in computer systems, above all a question of mathematics. All those involved, in both the public and private sectors, should therefore do away with this expression because of its psychological impact, a source of reluctance or, on the contrary, excessive confidence and acceptance. For this reason, the CNC DH recommends that public institutions and the media adopt more neutral expressions, such as “Algorithmic Decision Support System” (ADSS). Nevertheless, for reasons of editorial convenience, and because this is the current practice, the CNC DH will refer to “AI” in this opinion.

Recommendation 1: The CNC DH recommends favouring, in institutional communication, a more neutral and objective terminology than the term “artificial intelligence”, such as the term “Algorithmic Decision Support System” (ADSS).

3. This term covers, more specifically, IT technologies that are based on different operating logics: a distinction is mainly made between symbolic (or cognitive) AI and connectionist AI. The first involves programming a series of explicit and unambiguous instructions – in other words an algorithm – to give a result that is predictable because it presents itself as the logical processing of the data entered into the system. The second, more recently created, is based on another type of algorithm, no longer focused on a logical approach to information processing but on a probabilistic approach: programmers design a learning algorithm and submit to the computer a data set from which it will “learn” or, more precisely, infer rules. This learning can be supervised or unsupervised: in the former case, the data used for learning is labelled, while in the latter it is “raw”. In the latter case, machine learning establishes correlations between the information fed into the system.

4. This type of learning, machine learning, raises unprecedented challenges in regard to symbolic AI. While the instructions coded in conventional software can easily be communicated (although the system can struggle to understand when there are a lot of instructions and a lot of data to process, as is the case with the online higher education admissions platform, Parcoursup), the model that the machine reaches at the end of its learning can more or less easily be the subject of information, as the system designers cannot in some extreme cases (in particular in the case of deep learning) tell the operating model that the machine has reached to achieve its results.

5. While AI could, according to some, enable us to “*activate our fundamental rights*”¹, it nevertheless poses undeniable risks to them. At national level, the National Commission on Computer Technology and Freedom (CNIL) published a report on algorithms in 2017, the result of extensive consultation with stakeholders in the sector and citizens, faced with the need to “*enable man to keep the upper hand*”², and also focused its reflection on specific applications of AI³. In 2017, again, the Defender of Rights warned about the risks of discrimination caused by the use of algorithms in the fight against social security fraud, targeting categories of people to be checked as a priority⁴. Since then, the Defender of Rights has been pursuing a broader reflection on AI and discrimination⁵. At international level, many bodies have also warned about the impact of AI on fundamental rights⁶. In particular, the European Union Agency for Fundamental Rights (FRA) and the Ad Hoc Committee on Artificial Intelligence (CAHAI), a Council of Europe body responsible for examining the possibilities of establishing a legal framework on AI, have drawn up an inventory of fundamental rights likely to be threatened by AI: in particular respect for human dignity, respect for privacy and data protection, equality and non-discrimination, access to justice, access to social rights, etc.

6. The deployment of AI is all the more concerning as there is currently no comprehensive legal framework, both nationally and internationally, to stem its flow. The regulations in force provide only partial references, whether it concerns the protection of personal data – with, in particular within the European Union, the General

1 According to the formula in the Villani report: “*Donner un sens à l’intelligence artificielle : pour une stratégie nationale et européenne*”, 28 March 2018.

2 CNIL, “*Comment permettre à l’homme de garder la main ? Les enjeux éthiques des algorithmes et de l’intelligence artificielle*”, December 2017.

3 CNIL, “*Chatbots : des humains comme les autres*”, LINC, 3 February 2017; “*Reconnaissance faciale : pour un débat à la hauteur des enjeux*”, 2019.

4 Defender of Rights, “*Lutte contre la fraude aux prestations sociales : à quel prix pour les droits des usagers*”, September 2017.

5 Defender of Rights, in partnership with the CNIL, “*Algorithmes : prévenir l’automatisation des discriminations*”, 2020. More recently, the Defender of Rights published a report on biometric technologies: “*Technologies biométriques : l’impératif respect des droits fondamentaux*”, 2021.

6 See in particular: EU Agency for Fundamental Rights, “*Getting the future right: Artificial Intelligence and Fundamental Rights*”, 14 December 2020; UNESCO, Recommendation on the Ethics of Artificial Intelligence, November 2021; CAHAI, “*Feasibility Study*”, 17 December 2020; OECD, Recommendation of the Council on Artificial Intelligence, 22 May 2019.

Data Protection Regulation (GDPR) – or non-discrimination. However, this is still not enough since a large number of AI systems operate using non-identifying data and can have consequences on fundamental rights exceeding the protection of personal data and non-discrimination, as well as forms of discrimination likely to target groups not covered by the criteria of discrimination prohibited by law⁷.

7. For several years now, initiatives have come from the private sector. Recognising the need to offer trustworthy AI solutions to ensure commercial success, professionals offer ethics guides for designers and developers⁸. In addition, international institutions make recommendations to States along the same lines, such as those adopted by UNESCO on 24 November 2021 in order to “*make AI systems work for the good of humanity, individuals, societies and the environment and ecosystems, and to prevent harm*”⁹.

8. The concerns expressed through these texts, often formulated on the basis of a reference to “*ethical principles*”, largely coincide with human rights, particularly when it comes to autonomy or freedom, respect for human dignity, or non-discrimination. They are, however, of limited scope, relying on the self-regulation of stakeholders, the good will of manufacturers and companies, and do not impose any obligations on the States.

9. Given the significant impact of AI on fundamental rights, this approach does not seem sufficient. That is why the CNCDH has taken the initiative to review the issue. The magnitude of the issues raised with regard to fundamental rights through the design, deployment and use of AI systems calls for a binding legal framework to be put in place to ensure that these rights are respected. The CNCDH has carefully followed the discussions and work currently undertaken in this regard within the Council of Europe by CAHAI¹⁰, which could lead to the adoption of a legal framework for “*the development, design and application of artificial intelligence, based on Council of Europe’s standards on human rights, democracy and the rule of law*”¹¹. The CNCDH calls for the adoption of

7 CNIL, “*Comment permettre à l’homme de garder la main ? Les enjeux éthiques des algorithmes et de l’intelligence artificielle*”, December 2017, p. 49. Some algorithms are designed to establish correlations between different individual characteristics, from which they constitute groups and make predictions about behaviour at group level, for example “*dog owners living in the Paris region, aged 35 to 40, who do a sporting activity at least twice a week*”. Being identified as a member of this group can therefore lead to automated decisions that have adverse or beneficial effects for individual members, such as a differentiated health insurance tariff.

8 See in particular the practical guide “*Ethical AI*”, launched in September 2021 by Numeum, the leading professional association of digital companies in France.

9 Most recent text: UNESCO, Recommendation on the Ethics of Artificial Intelligence, November 2021.

10 Since the CAHAI fulfilled its mandate (2019-2021), it has been replaced by the Committee on Artificial Intelligence (CAI) for the development of an appropriate legal framework on the development, design and application of artificial intelligence, on the basis of the Council of Europe’s standards.

11 See in particular its latest publication: CAHAI, “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”, 2 December 2021.

an “AI Convention 108+”¹². With regard to the proposal for a European Union Regulation on AI (hereinafter referred to as the ‘proposal for a Regulation’), which establishes a first regional legal framework in order to “*foster the development, use and uptake of artificial intelligence in the internal market that at the same time meets a high level of protection of public interests, such as health and safety and the protection of fundamental rights, as recognised and protected by Union law*”¹³, it is a prerequisite for the CNCDH to take fundamental rights into account. Nevertheless, the Commission notes that there are insufficient guarantees to ensure effective compliance with the latter. Insofar as the proposal for a Regulation must meet the challenges of protecting these rights in the use of AI systems, the Commission recommends that this text ensure, to this end, the creation of a binding legal framework.

Recommendation 2: The Commission recommends strengthening, in the proposal for a European Union regulation on AI, the provisions to ensure the establishment of a binding legal framework guaranteeing effective respect for fundamental rights. In addition, the CNCDH recommends the adoption, within the framework of the Council of Europe, of a “Convention 108+ on AI”.

10. Thus, by drawing inspiration from the various steps already outlined by national bodies such as the Defender of Rights or the CNIL and European and international bodies, the CNCDH wishes to define the outline of a general framework, respectful of fundamental rights, for AI systems. In doing so, its opinion will help feed the necessary amendments to the EU AI Regulation. Further opinions will follow in the future in order to identify the risks to human rights, specific to the use of AI in certain sectors, as well as the guarantees that may be addressed. Moreover, the CNCDH has already expressed itself, in its opinion on the fight against online hate¹⁴, on the use of algorithms for moderation of content on social networks.

11. The CNCDH specifies that the use of a binding legal framework does not, of course, exclude flexible legal systems, with certifications and labels, under the supervision of the regulatory authority, in order to promote the development of good practices likely to accompany the implementation of this regulation.

¹² “Convention 108+”: The Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, available online:

https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/LIBE/DV/2018/09-10/Convention_108_EN.pdf. Opened for signature in the Council of Europe on 28 January 1981, Convention 108 was the first binding international legal instrument in the field of data protection. In order to respond to the challenges raised by digital media, on 10 October 2018, an amendment protocol was opened for signature by the States parties to the Convention. This new version is now referred to as “Convention 108+”.

¹³ Proposal for a Regulation on AI, cons. 5. Available online: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52021PC0206>.

¹⁴ CNCDH, *Opinion on the fight against online hate*, Plenary session of 8 July 2021, Official Journal of the French Republic No. 0170 of 24 July 2021, text No. 79.

12. Concerned with promoting an approach based on human rights¹⁵, the CNCDH will aim here both to highlight the need to include in the reflection and supervision of AI systems the most marginalised segments of the population and, above all, to stress the importance of establishing a legal framework to ensure respect for fundamental rights. The human rights approach involves putting humans in a position to define their needs and therefore support the development of an AI at the service of humans and their autonomy. This approach should further irrigate the ongoing reforms, as they aim to ensure respect for fundamental rights.

Recommendation 3: The CNCDH recommends that a human rights-based approach be taken into account in the ongoing reforms, as they aim to ensure respect for fundamental rights.

13. The observations and recommendations of the CNCDH will address the two components of an AI framework respectful of fundamental rights: a definition of “*red lines*”, in other words the uses of AI to be prohibited (1); guarantees to be promoted in order to ensure a framework of AI systems respectful of fundamental rights (2).

14. Within the framework of this opinion, and based on the terminology enshrined by the bodies of the European Union¹⁶, the CNCDH will designate as “designer” or “supplier” the natural or legal person who develops an AI system, as “user” any natural or legal person, public authority, agency or other body, including under private law, using an AI system under its own authority, and as “affected person” or “person targeted by an AI system” any natural person exposed to or impacted by an AI system.

1. RED LINES.

15. Some uses of AI cause too serious an infringement of fundamental rights to be permitted. It is the responsibility of the public authorities to prohibit their implementation. The proposal for an EU Regulation prohibits some use cases of AI, rightly considered as “*particularly harmful*”¹⁷. The CNCDH however points to certain limits in the definition of prohibited uses. In addition, other uses seem equally dangerous for fundamental rights and human dignity and, as such, would deserve to be banned as well.

15 CNCDH, *Opinion “Pour une approche fondée sur les droits de l’homme”*, Plenary session of 3 July 2018, Official Journal of the French Republic No. 0161 of 14 July 2018, text No. 104.

16 See the GDPR and the proposal for a Regulation on AI.

17 Proposal for a Regulation on AI, “Explanatory Memorandum”.

1.1. The contributions and limitations of the prohibitions laid down by the proposal for an EU regulation.

16. The proposal for an EU regulation on AI lists uses “*prohibited as contravening Union values*”¹⁸. This technology can be “*misused and provide novel and powerful tools for manipulative, exploitative and social control practices*”¹⁹. The CNCDH agrees with the idea that certain uses of AI should be purely and simply banned, given the extent of their impact on fundamental rights and freedoms.

17. The systems particularly problematic from this point of view, identified by the proposal for a Regulation, pose serious threats to the protection of fundamental rights and freedoms:

- systems based on subliminal components individuals cannot perceive, or exploiting vulnerabilities of children and people due to their age, physical or mental incapacities, and which, by distorting their behaviour, are likely to cause them psychological or physical harm;
- AI systems allowing the social scoring of natural persons, depending on their behaviour or personal characteristics, by or on behalf of the public authorities, for the purposes of the harmful or unfavourable treatment of certain natural persons or groups of persons;
- ‘real-time’ remote biometric identification in publicly accessible spaces for the purpose of law enforcement.

18. The CNCDH questions the definition and scope of the prohibitions thus imposed on the use of AI systems.

19. First of all, with regard to the first case, the terms used to define “malware” applications could cover a large number of situations, interfaces and online services, currently very much in vogue: “nudge” or “sludge” devices²⁰, set up by social media or online shopping sites, exploit cognitive biases to guide user behaviour, in order to capture their attention, encourage them to buy a product, etc. Terms such as “subliminal” or “substantial” are particularly complex and make the reach of this first limitation to the use of AI uncertain. This prohibition, however, has the merit of raising questions about the permissible risks of AI manipulation, in particular with regard to the processes of automating the processing of information resulting from

18 Proposal for a Regulation on AI, “Explanatory Memorandum”.

19 *Ibid.*, cons. 15.

20 “Nudges” correspond to techniques to guide the user of an online service to make a choice. For example, indicating on a site that an item or product is “popular” will encourage an internet user to view or buy it. See in particular: “La forme des choix”, CNIL, 2019. On the other hand, “sludges” have a deterrent purpose. For example, with regard to the management of cookies on a site, some sites simplify the “Accept All” option and make the “Configure Cookies” option more complex, so that users tend to favour the former.

neuromarketing research²¹. It is all the more essential, moreover, in the development of “*augmented dark patterns*” allowing, through algorithmic processing, dynamic action to be taken on users’ stimuli to exploit their vulnerabilities. In this respect, the CNCDH supports the will of European parliamentarians, expressed elsewhere²², to prohibit the use of manipulation techniques, which aim to encourage users to make certain choices, on very large online platforms.

Recommendation 4: The CNCDH recommends the prohibition of the use of choice interfaces insofar as they have the purpose or effect of manipulating users, to their detriment, by exploiting their vulnerabilities.

20. Secondly, and more worryingly, the proposal for the EU text prohibits social scoring, which aims to evaluate people based on their social behaviour or known or predicted personal or personality characteristics, only when it is practised by or on behalf of the authorities. However, private companies, for example social networks, can also process large amounts of personal data and perform scoring. Therefore, the CNCDH endorses the position of the European data protection authorities, in favour of a ban on any type of social scoring, regardless of the nature, public or private, of the entity that implements it²³.

Recommendation 5: The CNCDH recommends banning any type of social scoring set up by government authorities or by any company, public or private.

21. Finally, the prohibition on the real-time biometric identification of natural persons – i.e. the use of AI for automated recognition of human characteristics such as face, voice, gait, etc. – in publicly accessible spaces and for the purpose of law enforcement raises, in its current form, questions about its scope of application, but also concerns about the scope of derogations accepted by the proposal for a Regulation.

22. The CNCDH joins the European Commission in highlighting not only the risk to respect for privacy generated by such a system, but also the “*feeling of constant surveillance*” likely to be generated by this technology and the risk of “*indirectly dissuading the exercise of the freedom of assembly and other fundamental rights*”²⁴,

21 Neuromarketing looks at the functioning of the brain, particularly using brain imaging (MRI), to better understand how consumers react to advertising and sales devices.

22 See the debates on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act).

23 EDPS-EDPB Joint Opinion 05/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Available online: https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.

24 Proposal for a Regulation on AI, Cons. 18. The CNCDH also pointed out this “chilling effect” risk associated with the use of drones equipped with facial recognition software: CNCDH, *Opinion on the proposal for a law on global security*, Plenary session of 26 November 2020, Official Journal of the French Republic No. 0290 of 1 December 2020, text No. 83.

starting with freedom of movement. The CNCDH points out, however, that this dissuasive effect is equally valid for a post biometric identification system, i.e. one that is not in real time compared to when the images were collected²⁵. Furthermore, in the absence of an express prohibition, the use of this technology for preventive purposes is permitted; but the reservations expressed above with regard to its use for the purpose of law enforcement are also valid, if not more so, when it is used to prevent violations of public order.

23. Furthermore, the proposal for a Regulation provides for three exceptions to the prohibition of real-time biometric identification of people: the targeted search for specific potential victims of crime, including missing children; the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA²⁶. This third exception is particularly worrying since by accepting this technology for more than thirty offences, it in principle waters down the effectiveness of the prohibition.

24. In short, the CNCDH therefore recommends the prohibition of remote biometric identification of people in publicly accessible spaces, due to the risks of serious infringement of fundamental rights and freedoms linked to a real or alleged challenge of anonymity in the public space²⁷. However, it recognises the legitimacy of the first two types of exceptions envisaged by the proposal for a Regulation, while stressing the need to ensure, where appropriate, strict supervision of them. The exception should therefore be limited in particular to the targeted search for specific potential victims of crime or to the prevention of a specific, substantial and imminent threat to the life or physical safety of persons and that of structures, facilities and establishments of vital importance.

Recommendation 6: The CNCDH recommends prohibiting the remote biometric identification of persons in publicly accessible spaces, by way of exception permitting its use, insofar as it is strictly necessary, adapted and proportionate, for the prevention of a serious and imminent threat to the life or physical safety of persons and that of

²⁵ The proposal for a Regulation accepts the possibility of using this technology in this case, nevertheless categorising it as a high-risk AI system.

²⁶ In particular: participation in a criminal organisation, illicit trafficking in arms, ammunition and explosives, corruption, environmental crimes, including illicit trafficking in endangered animal species and illicit trafficking in endangered species and plant species, aid for irregular entry and residence, racism and xenophobia, organised or armed robbery, illicit trafficking in cultural assets, including antiques and works of art, fraud, racketeering and extortion, falsification of administrative documents and trafficking in forgeries, illicit trafficking in hormone substances and other growth drivers, illicit trafficking in nuclear and radioactive materials, trafficking in stolen vehicles.

²⁷ In the same vein: EDPB-EDPS Joint Opinion 05/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 18 June 2021.

structures, facilities and establishments of vital importance.

25. After reviewing the uses of AI deemed unacceptable by the proposal for a Regulation, the CNCDH questions the method used by the European Commission to achieve this identification, in the absence of sufficient indications in this regard. Based on the “triple test” derived from the case law of the European Court of Human Rights, endorsed by the case law of the Constitutional Council and the Council of State²⁸, the CNCDH considers that, in order to be considered legitimate, the infringement by an AI system of freedom must be “*appropriate, necessary and proportionate*”: appropriate, i.e. relevant for the legitimate objective pursued; necessary, provided that it does not exceed what is required to achieve this objective and other means less detrimental to freedom were not possible; proportionate, in that it should not, by the workload it creates, be disproportionate to the desired result²⁹.

1.2. The need to extend the ban to other areas.

26. The CNCDH has opted in this opinion for a comprehensive and cross-sectional approach to AI systems. It has not therefore carried out a detailed examination of the various applications likely to cause an excessive infringement of fundamental rights justifying their prohibition. This type of analysis may be carried out in future opinions devoted to new technologies or specific sectors. However, in addition to the observations on use cases prohibited by the proposal for an EU regulation on AI, the CNCDH would like to highlight two types of use of particular concern for the respect of human rights: predictive justice and the recognition of emotions in support of a selection process.

1.2.1. AI in court.

27. Applications such as those used in the United States to assess the risk of convicts reoffending³⁰ pose a serious threat to their fundamental rights. The lack of transparency of the software used (designed by private companies) calls into question its compatibility with the fundamental rights of individuals and the guarantee of the rights of defence. In fact, under the pretext of the right to business secrecy and intellectual property, software designers are not required to share the source code of their algorithms, from which their instructions derive. It is therefore impossible for both the judge and the parties to the trial to understand precisely the methodology used by the algorithm to produce its results³¹.

28 CC, Decision No. 2008-562 DC of 21 February 2008, Act pertaining to post-sentence preventive detention and diminished criminal responsibility due to mental deficiency.

29 See in particular: M. Guyomar, “*Le passeport biométrique au contrôle : empreintes et cliché*”, *Actualité Juridique Droit Administratif*, 2012, p. 35; J.-M. Sauvé, “*Le principe de proportionnalité, protecteur des libertés*”, Institut Portalis, Aix-en-Provence, 17 March 2017.

30 See in particular COMPAS, the re-offending risk analysis and assessment algorithm used in several US States.

31 See: Supreme Court of Wisconsin, July 13, 2016, *State v. Loomis*.

28. This type of recourse to AI is already prohibited in France since “no court decision involving an assessment of a person’s behaviour can have as a basis automated processing of personal data intended to assess certain aspects of the person’s personality”³². The formula adopted does not, however, preclude any possibility of providing magistrates with an AI application for other purposes, for example to automate the calculation of compensation for damage.

29. In view of the workload currently weighing on magistrates, the guarantees that could be provided in order to preserve the impartiality of the judge (explicability/intelligibility of how the algorithm works, reinforcement of the means of remedying automation bias) do not seem sufficient to prevent the risk of an almost systematic take-up of the machine’s results. In addition, the intervention of an AI system to provide assistance to the judge may raise doubts among litigants as to its impartiality. However, the judge must not only be independent and impartial, he/she must also appear to be it³³. It is therefore the right to a fair trial that is thus threatened by this type of software.

30. Furthermore, if the judge can be allowed a computer tool to facilitate his/her assessment of compensation for victims³⁴, it cannot be based on machine learning, which is too opaque to meet the requirement of explicability that litigants are entitled to expect³⁵. The CNCDH also notes that the Ministry of Justice prematurely terminated, in January, its experimentation with such software, by establishing the multiplicity of criteria to be taken into account to characterise the extent of bodily injury and the excessive importance of the means to be mobilised to study and prevent algorithmic biases in order to achieve a satisfactory level of performance³⁶.

Recommendation 7: The CNCDH recommends continuing and deepening reflection in order to identify the contributions and limits of the use of AI in the context of judicial proceedings.

1.2.2. AI and recognition of emotions.

31. Emotion recognition technologies are based on a premise that is not very scientific, namely that emotions can be detected by facial expressions or, more

32 French Data Protection Act No. 78-17 of 6 January 1978, Art. 47 amended by Order No. 2018-1125 of 12 December 2018. On this point, French law goes further than the GDPR, since the GDPR only prohibits decisions, including court decisions, based “exclusively” on automated processing.

33 According to appearance theory, enshrined by the European Court of Human Rights, “justice must not only be done, it must also be seen to be done”.

34 Automated analysis of court decisions is technically based on natural language processing and machine learning: they can be a useful information tool for legal professionals.

35 On this point, Law No. 2016-1321 of 7 October 2016 (known as the “Digital Republic” calls for the transparency of public algorithms, thus offering a guarantee against a possible “black box” phenomenon in terms of judicial use of artificial intelligence. See the response of the Minister of Justice published in the OJ Senate of 01/10/2020, page 4462, available online: <https://www.senat.fr/questions/base/2020/qSEQ200616942.html>.

36 E. Marzolf, “Le ministère de la Justice renonce à son algorithme DataJust”, *Acteurs publics*, 14 January 2022.

generally, by body language. However, as the CAHAI recalls, “*no solid scientific evidence corroborates the idea that it would be possible to ‘read’ the emotions or the state of mind of a person on his/her face or using other biometric data*”³⁷. In addition to the performance of the AI system, which progress in the design could remedy, its character may be inappropriate. Several national³⁸ and international³⁹ authorities have already expressed their concern in this regard, as European data protection bodies have even recently recommended banning the deduction of emotions through the use of AI, except in certain specific cases “*particularly for health and research purposes*”⁴⁰.

32. Sharing the same fears, the CNCDH therefore recommends applying a prohibition principle in this area, unless it can demonstrate that this biometric technology is able to reinforce the independence of people, or more broadly the effectiveness of their fundamental rights. To this end and despite its approximations, this AI technology can, for example, promote learning activities for people with disabilities or be useful for other human-machine interactions (such as robot companions for the elderly).

Recommendation 8: The CNCDH recommends that emotion recognition technologies should be banned, by way of exception permitting their use as long as they aim to reinforce the independence of people, or more broadly the effectiveness of their fundamental rights.

2. A FRAMEWORK THAT GUARANTEES RESPECT FOR FUNDAMENTAL RIGHTS.

33. Although many AI applications do not cause a disproportionate infringement of fundamental rights and freedoms that would justify their prohibition, the CNCDH calls on public authorities to enforce certain guarantees by public and private stakeholders when AI is being designed, developed and used. This must mainly involve control and supervision of the AI system, at all stages of its life cycle, in view of its impact on fundamental rights. In addition to the vigilance to be exercised with regard to the system envisaged as a whole (2.1), the decisions resulting from its implementation must be accompanied by safeguards that are able to protect individuals (2.2).

2.1. Control at all stages of AI system development.

34. Depending on the areas concerned (organisation of work, calculation of

37 CAHAI, Feasibility Study, p. 8.

38 Defender of Rights, “*Technologies biométriques : l’impératif respect des droits fondamentaux*”, 2021; CNIL, “*Reconnaissance faciale : pour un débat à la hauteur des enjeux*”, 2019.

39 EDPS, EDPB, EDPB-EDPS Joint Opinion 05/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 18 June 2021.

40 EDPS, EDPB, *ibid.*, p.14

social security benefits, aid to companies, etc.), opting for an AI solution can have repercussions on employees or on the persons targeted by the software decisions. A human rights-based AI approach means involving those affected by its deployment, in particular the most vulnerable⁴¹: 1. upstream, carrying out an impact assessment on fundamental rights is an essential step, and the result will affect the level of stakeholder involvement at this stage; 2. secondly, the monitoring phase of AI system implementation must include them in order to assess whether its functioning leads to infringements of fundamental rights; 3. more generally, the extent of this human control requires an improved offering of professional training and awareness-raising for individuals.

2.1.1. An impact assessment on fundamental rights: a necessary prerequisite.

35. Apart from certain uses of AI to be prohibited due to the seriousness of the risks they pose to fundamental rights and freedoms⁴², other uses may have a more or less adverse impact on fundamental rights. These adverse effects are most often mentioned on the basis of a reference to the “sensitive” sectors in which the system is deployed, such as police, justice or health, or an inventory of fundamental rights and freedoms that may be challenged by AI technology⁴³.

The approach of the EU AI Regulation: an a priori and centralised definition of risks for fundamental rights.

36. For its part, the proposal for a Regulation reserves special treatment for AI systems classified as “high-risk”, identified as such by the European Commission because of the “extent” of their adverse impact on health, safety or fundamental rights. These are AI systems that operate in “sensitive” areas: biometric identification, management and operation of critical infrastructure, education and training, employment, access to essential private services, public services and social benefits, police, justice, migration management. The assessment of the impact of AI systems on human rights is therefore centralised and carried out a priori: high-risk systems are included in a list⁴⁴ that the Commission may expand, under certain conditions and in accordance with a procedure that may take time⁴⁵. Of course, data controllers still have an obligation to carry out, in accordance with GDPR requirements, an impact assessment relating to data protection

⁴¹ See the opinion of the CNCDH, *Opinion “Pour une approche fondée sur les droits de l’Homme”*, Plenary session of 3 July 2018, Official Journal of the French Republic No. 0161 of 14 July 2018, text No. 104.

⁴² See above.

⁴³ See in particular the CAHAI feasibility study.

⁴⁴ Appendix 3 to the proposal for a Regulation.

⁴⁵ The Commission has the power to amend the list in Appendix 3 by “delegated acts”, the adoption procedure of which is set out in Article 73 of the proposal for a Regulation: it can take 6 months from notification of the proposed amendment to its entry into force, given that the European Parliament and the Council may also oppose it.

and, more generally, the rights and freedoms of natural persons⁴⁶.

37. The CNCDH has reservations about this system for several reasons. First, technological innovation is moving faster than regulations, and the list of high-risk AI systems used by the European institutions may not take into account current and future uses of particular concern for fundamental rights. Then, AI can, particularly with regard to machine learning in its most automated version (deep learning), evolve independently of the intentions of its designers, which is why risks ignored at the system's design stage may occur when the algorithm is developed or used⁴⁷. Above all, inclusion in this list gives rise to a series of obligations, mainly incumbent on suppliers, relating to risk management, data integrity, control and monitoring of the system, etc., while users of a high-risk AI system have few obligations⁴⁸.

An impact study on fundamental rights incumbent on users.

38. The CNCDH considers that it is necessary to go further by also requiring the user of the AI system to carry out a study of its impact on fundamental rights. For several reasons: firstly, because this study places the responsibility on the public or private body wishing to use this IT option; secondly, because it could possibly lead to dialogue between all stakeholders (employees, customers, users of a public service, etc.) on the appropriateness of its use; finally, because it will be a source of information for oversight of the AI system throughout its life cycle, or even for the person who is ultimately affected by an automated decision.

39. If there is no question, within the framework of this general opinion, of detailing the elements to be taken into account when carrying out this analysis, the CNCDH would like to formulate a certain number of recommendations relating to the broad lines that could guide its implementation.

40. It therefore recommends that users should assess the impact of the use of the AI system on fundamental rights and, if risks are identified, carry out an assessment taking into account the probability and severity of these risks. This would include, for example, tax fraud detection algorithms, social security benefits, personnel management support systems, voice control software in distribution platforms, etc.

41. This analysis may include the communication of elements provided by the designer, when they have an impact on fundamental rights: data sets used for machine

⁴⁶ In accordance with Article 35 of the GDPR. The proposal for a Regulation also takes care to reiterate it in Article 29.

⁴⁷ For example, Microsoft's chatbot, intended in 2016 to participate in social media exchanges, quickly made abusive and racist comments, reversing the intention of its programmers. This drift is the result of the interaction of the AI system with ill-intentioned people: M. Tual, "A peine lancée, une intelligence artificielle de Microsoft dérape sur Twitter", *Le Monde*, 24 March 2016.

⁴⁸ See Article 29 of the Proposal for a Regulation on AI and Article 52 for certain more specific AI systems.

learning, for example for suspicious behaviour detection software intended to equip surveillance cameras; the different types of settings possible, for example, for automated content moderation tools, etc. The user must therefore indicate and justify his/her configuration choices by putting in perspective the intended purpose and the risks of infringement of fundamental rights.

The content of the impact assessment.

42. As regards its content, any impact assessment should, as a minimum, consist initially of mentioning the purposes of using the envisaged AI system, and identifying the fundamental rights that may be affected.

43. Secondly, this study should reveal the answers provided by the user to the questions that usually feed into the system controlling fundamental rights and freedoms:

- Is an AI system necessary for the task at hand? To what extent does it bring added value to prior operation?
- Is the AI system appropriate for the intended purpose? To what extent will it be possible to accomplish the task?
- Is the AI system proportionate? To what extent is the potential infringement of human rights, including the environmental impact, justified in relation to the benefits expected of achieving the legitimate objective of the system?

44. Finally, on a more technical level, the impact study should include the procedures put in place to monitor the application, as well as the measures taken to mitigate the risks involved⁴⁹.

Stakeholder consultation procedures subject to the conclusions of the impact assessment.

45. Depending on the level of risk for fundamental rights identified at the end of the impact assessment, stakeholder consultation should potentially be planned in order to discuss whether or not to use the AI solution envisaged by management. Three risk levels could be selected to determine the terms and conditions of this consultation:

- a high level (for all processing affecting the rights of individuals): consultation of all stakeholders, including staff representatives and associations of public service users or consumers, ensuring the inclusion of associations of disadvantaged persons;
- a moderate level (for processing that does not directly affect people's rights, such as leave management software): the consultation referred to in the case of high level

⁴⁹ The High-Level Expert Group on Artificial Intelligence (EU) refers to this type of analysis, upstream of the self-assessment it recommends and for which it relies on a multi-criteria checklist for "Trustworthy AI": High-Level Expert Group on Artificial Intelligence, The Assessment List for Trustworthy Artificial Intelligence (ALTAI), June 2020.

would be optional, and could be requested for example by the affected persons or their representatives;

- a low level: the impact assessment would be communicated at the request of the persons affected by the AI system or their representatives.

46. An impact assessment by the user is all the more necessary as it will facilitate supervision of the AI system once it is set up.

Recommendation 9: The CNCDH recommends that the user of an AI system assess the impact of using this system on fundamental rights and, if risks are identified, carry out an assessment taking into account the probability and severity of these risks. The impact assessment should include at least:

- a statement of the purpose(s) attached to the use of the envisaged AI system;
- identification of the fundamental rights likely to be affected by the system;
- a review of the envisaged AI system, based on an assessment of its necessity, its suitability and the proportionality to the infringements of fundamental rights in relation to the intended purpose;
- the procedures put in place to monitor the application, and the mitigation measures with regard to the risks incurred.

Recommendation 10: Depending on the risks posed by an AI system to fundamental rights in a particular context of use, the CNCDH recommends ensuring, prior to the decision to use it, a stakeholder consultation, according to appropriate procedures, including for example staff representatives and, more broadly, the persons affected by the AI system.

2.1.2 Supervision of the system throughout the life cycle.

47. The quality and relevance of the data selected to design algorithms and the proportionality of possible infringements of fundamental rights by the AI system may be subject to control prior to the use of the AI system. However, infringements of fundamental rights may occur after the system is taken over by the user. This is why continuous vigilance over the functioning of the AI system must be ensured. This supervision must be organised, as has been explained above, with regard to stakeholder consultation at the end of the impact study, according to procedures that are more or less demanding on the user depending on the risks for the fundamental rights identified by this study.

Continuous vigilance over the effects of the AI system on fundamental rights.

48. Periodic control must be carried out at the various stages of use of an AI system. It must be based on the human rights impact assessment by ensuring that the risks identified upstream have not materialised or, where required, identify the measures to

be taken to neutralise these risks. Infringements of fundamental rights not identified by the impact assessment may also be identified on this occasion.

49. In this regard, particular attention should be paid to the risks of discrimination caused by AI systems. These risks have already been widely documented⁵⁰. While the automation of decision-making processes may give the impression at first glance of being free from the prejudices inherent to human subjectivity, algorithms may reproduce, strengthen or generate biases, particularly systemic ones, likely to aggravate discriminatory situations. As the Commissioner for Human Rights of the Council of Europe points out, there is a major risk of “essentialisation” and strengthening “stereotypes” because the predictive nature of the algorithm is based on the homogenised behaviour or characteristics of groups.

50. These biases can be caused by the data used to feed into the machine. The data sets used to train the algorithmic model may indeed include discrimination, for example if an AI is expected to select the best applications for a given position only from the files of previously recruited people, the model may reproduce discriminatory biases as long as they characterise these recruitments (racist or sexist for example). These data sets may also not be sufficiently representative of the diversity of the population, for example if facial recognition software is populated primarily by photographs of Caucasian people. The software will tend to make identification errors on black people, leading to unjustified questioning.

51. Algorithm classifications are therefore likely to generate discrimination against individuals, because of their membership of a group, which can indirectly correspond to a group protected by non-discrimination law (by proxy, *i.e.* a variable linked to a prohibited discrimination criterion, for example a dietary habit that would be evidence of religious beliefs).

52. In addition, the discriminatory effects of algorithms are not noticeable on an individual scale, given the opacity of the functioning of algorithms, but also because these discriminations are much easier to observe at group level than at individual level⁵¹. For this reason, continuous vigilance must be exercised over AI systems and algorithm classification when these systems produce results that have, even indirectly, effects on the rights and freedoms of individuals⁵².

⁵⁰ See in particular the reports of the Defender of Rights: see above, footnote no. 4.

⁵¹ Defender of Rights, in partnership with the CNIL, “*Algorithmes : prévenir l’automatisation des discriminations*”, 2020, p. 6.

⁵² In the same vein, the Defender of Rights and the CNIL recommend regular monitoring of the effects of algorithms after their deployment, along the same lines as control of adverse drug reactions: Defender of Rights and the CNIL, “*Algorithmes : prévenir l’automatisation des discriminations*”, p. 10.

Conditions for supervision by the user.

53. The proposal for a Regulation on AI requires suppliers of high-risk AI systems to establish a quality management system, which should gather information about all the procedures and instructions they put in place to comply with the requirements of the Regulation⁵³. These instructions must include, among other things, the terms of the monitoring established by the supplier after marketing of an AI system, in order for it to comply with the regulation in the long term⁵⁴. Control over how the supplier performs this monitoring is entrusted by the proposal for a Regulation to approved conformity assessment bodies⁵⁵, or even to a supplier internal assessment⁵⁶.

54. The CNCDH considers this approach necessary but insufficient where AI systems are concerned for which the impact assessment reveals a significant risk of infringement of fundamental rights. In this case, supervision should be based in addition on tests and surveys carried out by the stakeholders, the user and the affected persons, at intervals to be determined depending on the context of use. From this point of view, the “human oversight” provided for in Article 14 of the proposal for a Regulation, in particular to “*fully understand the capacities and limitations of the AI system*”, should be clarified and supplemented by a reference to the collegial nature of the process and the opening up of this college to the various stakeholders.

55. Oversight relating to the impact of a recruitment or management AI system within a company or administration must, for example, include the involvement of staff representatives. These representatives are in fact key players, particularly within the company’s social and economic committee (CSE), in the assessment of the psychosocial risks caused by an AI system within their organisation but also its impact on the organisation of work.

56. If the impact assessment does not reveal any significant risks to fundamental rights, supervision of the AI system may be the sole responsibility of the user. The CNCDH indeed insists on the need to highlight its role in the vigilance to be maintained throughout deployment of the system, beyond the mere communication of “*relevant data*” to the supplier in order to comply with the monitoring obligation incumbent upon the supplier, in accordance with the proposal for a Regulation⁵⁷.

Recommendation 11: The CNCDH recommends setting up oversight of the AI system, according to a procedure likely to vary according to the risks of infringements of fundamental rights as identified by the impact assessment, in order to maintain

53 Proposal for a Regulation on AI, Art. 17.

54 Proposal for a Regulation on AI, Art. 61.

55 “Notified bodies”.

56 Proposal for a Regulation on AI, Art. 43.

57 According to Article 61 of the proposal for a Regulation on AI.

continuous vigilance on the part of the user with regard to the effects of the system, including its discriminatory effects.

2.1.3 Training and awareness of AI issues.

57. In view of the place to be assigned to stakeholders in the control and supervision of AI systems, particularly employees and the individuals affected by the system's decisions, the CNCDH recommends the implementation of training and awareness campaigns on AI technologies. Training modules could be widely disseminated to employees and everyone, for example in the form of MOOCs. The CNCDH therefore recommends public investment in the design of training and information tools accessible to as many people as possible.

Recommendation 12: The CNCDH recommends encouraging public investment in the design of training and information tools accessible to as many people as possible.

58. In addition, public authorities should organise public debates on this issue. Based on the model of the États Généraux de la Bioéthique (Bioethics Forum) organised by the National Advisory Committee on Ethics (CCNE), these consultations would have a dual purpose: on the one hand, to inform citizens about how these systems operate and their purpose, and, on the other hand, to enable them to position themselves on national guidelines on this subject. In doing so, it is a question of promoting the expression of a diversity of views on a number of uses of AI. Special attention should be paid to the poorest people, in order to ensure that they are able to participate.

Recommendation 13: The CNCDH recommends organising national consultations along the same lines as the États généraux de la bioéthique (Bioethics Forum) organised by the National Advisory Committee on Ethics.

59. The proportion of algorithms in daily life and the functioning of society calls for the acquisition of a true computer culture from a very young age. The CNCDH thus emphasises the need for National Education to strengthen the training of students in the technical, political and societal challenges of artificial intelligence and to propose, to this end, educational materials for teachers.

Recommendation 14: The CNCDH recommends that the Ministry of Education strengthens the training of students in the technical, political and societal issues surrounding artificial intelligence and proposes, to this end, educational materials for teachers.

2.2. Guaranteed respect for fundamental rights with regard to individual decisions

60. The weight given to the algorithm in decision-making varies from one use to another: human intervention may be anticipated, or even required, but in some cases decision-making can also be fully automated. The need for and the conditions of this intervention depend on the level of risk of infringement of fundamental rights. Moreover, the persons targeted by an AI system must be informed, and also have intelligible information on how the algorithm works and on the weight given to the algorithm in the individual decision.

2.2.1 Human intervention guaranteeing consideration of individual specificities.

61. To “enable humans to keep the upper hand”⁵⁸ over AI, the CNCDH calls for the reintroduction of humans at the end of the automated decision-making process: either at user level, responsible for checking the result produced by the algorithm, or at the level of the person affected.

Human intervention at user level: checking the algorithmic result.

62. In some cases, human oversight of the general functioning of the AI system, as mentioned above, requires in addition human intervention in relation to individual decisions based on the results of this system. The GDPR has enshrined a right “*not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*”⁵⁹. The legislator has taken care to include it in the 1978 French Data Protection Act, specifying, in light of the scope of the GDPR, that the prohibition concerns any “*automated processing of personal data*”⁶⁰.

63. However, the current regulation has two limitations: on the one hand, it does not cover algorithmic processing using anonymised data and, on the other hand, it envisages many derogations from its prohibition in principle, starting with the ability of EU Member States to authorise such processing, on condition, however, that it provides for “*suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests*”.

64. The need for human intervention has already been highlighted by the Constitutional Council for individual decisions adopted on the basis of an algorithm whose operating procedures cannot be communicated (because their communication

⁵⁸ According to the formula enshrined by the CNIL in its report of 2017. See above.

⁵⁹ Proposal for a Regulation on AI, Art. 22.

⁶⁰ French Data Protection Act 78-17 of 6 January 1978, Art. 47(2)

would undermine defence secrecy, State security, etc.)⁶¹. More recently, with regard to the French system for detecting connections likely to reveal a terrorist threat, the CJEU argued that *“any positive result obtained following automated processing must be subject to an individual re-examination by non-automated means before an individual measure adversely affecting the persons concerned is adopted”*⁶².

65. The CNCDH considers that the need for human intervention should be imposed more generally, to varying degrees, depending on the field in question, for all algorithmic processing having effects on the rights of individuals.

66. In order to ensure effective human control in the context of use of the AI system, the conditions of human intervention may vary depending on the impact of the AI system on fundamental rights, in terms of:

- the composition of the supervisory body (individual or college);
- the extent of the information made available to stakeholders, knowing that, in some cases, it will be necessary to provide additional data to those processed by the system;
- the type of training to be provided to persons in charge of intervention;
- the appropriate moment to intervene (at the end of the result obtained by the machine, therefore upstream of the decision, or ex post at the request of the person).

67. Ensuring effective human intervention means informing the stakeholder about the characteristics of the algorithm used: the technology at the origin of its design, the type of data used for its modelling, the operating parameters and the weighting of the criteria used by the algorithm designer, reliability, etc. This information on the *“inner workings of the machine”* is required to encourage standing back from the AI system used, thereby reducing the cognitive automation bias of placing excessive reliance on automated decision-making processes. This is why the CNCDH reiterates the need for professionals assisted by an AI application in the performance of their duties (doctors, magistrates, administrative agents, recruiters, etc.) to have clear, complete and comprehensible information on these aspects.

68. In order to neutralise the automation bias, the CNCDH also recommends training any stakeholder called upon to oversee the individual results produced by the AI system, concentrating in particular on its limits (biases derived from the data, the probabilistic nature of the results obtained, etc.). In addition, the Commission recommends that no particular constraints should be imposed, for example by additional formalities, on these persons when they depart from the algorithmic indication.

69. Finally, and particularly with regard to citizens, automation of decisions can

61 Constitutional Council, Decision No. 2018-765 DC, Law related to the protection of personal data, § 70.

62 CJEU, 6 October 2020, Case C 511/18, La Quadrature du Net, § 182.

only aggravate the alienation felt, faced with the growing dematerialisation of public services and the difficulty of asserting their rights⁶³. The CNCDH therefore endorses the recommendation of the Defender of Rights when it advocates “*the systematic maintenance of alternative access and the possibility of sufficiently close, competent and available support*”⁶⁴.

Recommendation 15: The CNCDH recommends:

- ensuring human intervention to oversee individual decisions resulting from an AI system in accordance with the latter’s risk level;
- ensuring effectiveness thereof through appropriate training and information for personnel on the characteristics of the system, without imposing any particular constraint on them when they deviate from the AI system’s recommendation;
- ensuring the systematic maintenance of alternative access to a human agent for public service users.

Human intervention at the level of the person affected: the right to configuration.

70. In some cases, the person affected by the functioning of AI is directly exposed to the results of an algorithm designed, according to the accompanying marketing message, to “*meet its needs*”. From this perspective, the algorithm presents itself as a tool shaped by an operator and used by the people who resort to its services. In this respect, the CNCDH is particularly concerned about the lack of control by the user of the system’s operating parameters.

71. As the CNCDH has already noted in its opinion on the fight against online hate, such control seems particularly necessary with regard to content selection algorithms on social networks. Freedom of conscience in fact requires the autonomy of users to be strengthened and their control over the content offered to them to be increased. The CNCDH therefore renews its recommendation to recognise a right to configure the criteria for determining the content received, as regards both their selection and their presentation⁶⁵.

72. More generally, this right to adapt the parameters of the algorithm would result in a new manifestation of the role that should be recognised for the user the moment a genuine “*human-centred AI*” is advocated for, an intention shared by the French, European and international authorities. The user could thus be given the right to adapt the parameters of AI systems when they feed into interpersonal or human-machine interactions, which in particular will be developed with the growth of chatbots.

63 Defender of Rights, “*Dématérialisation des services publics : trois ans après, où en est-on ?*”, Report, 2022.

64 *Ibid.*, p. 5.

65 For a more detailed presentation of this right and its implications, see: CNCDH, *Opinion on the fight against online hate*, Plenary session of 8 July 2021, Official Journal of the French Republic No. 0170 of 24 July 2021, text No. 79.

Recommendation 16: The CNCDH recommends that users of AI systems be given the right to configure their criteria, in particular in order to determine the selection and presentation of the content received, and more generally in the event of human-machine interactions.

2.2.2. Information guaranteeing human dignity.

73. Information on the characteristics of the AI system used is necessary for monitoring individual decisions. It must also allow the affected person to understand the reasons and, possibly, to challenge it. To do so, the person must be informed that the decision to which he or she is subject is based, in part or in full, on an automated process. He/she must then have the information enabling him/her to understand how the algorithm used works.

Information on the intervention of an AI system in the decision.

74. The person exposed to an AI system must be informed about it. The CNCDH considers that this is a prerequisite for which there should be no exceptions.

75. The current regulations already recognise a right for users of the administration concerned by the use of an AI system to be informed that an individual administrative decision has been made on the basis of an algorithm⁶⁶. However, there are a large number of exceptions to this right. Indeed, the law provides for the exclusion of this information in the event that it has an impact on: the secrecy of the deliberations of the Government and of the responsible authorities under the executive power; national defence; the conduct of France's foreign policy; State security, public security, the security of persons or the security of the administrations' information systems; currency and public credit; the conduct of proceedings before the courts or preliminary operations in such proceedings, unless authorised by the competent authority; the investigation and prevention, by the competent authorities, of offences of any kind; other secrets protected under Article L.124-4 of the French Environmental Code⁶⁷.

76. The CNCDH regrets the extent of these grounds, especially since they were initially intended to justify the exclusion of a communication or a consultation of an administrative document. There are two different types of information to distinguish: on the one hand, information on the nature of the process at the origin of the decision, in this case the use of an automated decision-making process and, on the other hand, information on how the algorithm used works. If it is acceptable that the logical rules making up the algorithm can escape communication due to a certain number of public imperatives, this cannot justify there being no mention that the individual decision is based on the use of an algorithm, even though this piece of information will be useful

⁶⁶ Art. L.311-3-1 of the French Code of Relations between the Public and the Administration.

⁶⁷ Art. L.311-5 of the French Code of Relations between the Public and the Administration.

to the interested party in the event of a judicial appeal.

77. The CNCDH believes that this obligation of information should be extended to private individuals, while being designed more broadly with regard to public entities.

78. While the proposal for an EU Regulation mentions a “*transparency obligation*”, it does so restrictively by reserving it for certain AI systems. First, it imposes an obligation on providers of AI systems intended to interact with humans, forcing them to design them in such a way that the persons affected are warned about them. Secondly, it imposes an obligation on users of emotion recognition systems, or a biometric categorisation system, to inform the persons exposed to it. Finally, the manipulation of images or audio or video content, such as deepfakes, must be accompanied by a warning message.

Recommendation 17: The CNCDH recommends systematically informing people when they are exposed to or required to interact with an AI system and, when they are the subject of a decision, that this decision is based, where applicable, in part or in full on algorithmic processing.

79. Those affected by a decision resulting from an AI system must not only have a right to be informed of it, but also have a right to contest this decision with a human being. This human being must be able to review the individual’s file. In order to make this right to review effective, easily accessible channels must be made available to those concerned.

Recommendation 18: The CNCDH recommends guaranteeing the affected person the right to review, by a human being, of any individual decision based totally or partly on algorithmic processing, provided that it has significant consequences for him/her.

Information on the conditions of automated decision-making.

80. The requirement for AI system explicability is found in most of the reference texts on governance of AI⁶⁸. Sometimes understood to mean “transparency”, in any case likely to have diverse acceptations, explicability essentially refers to the need for the designer and/or the user of an AI system to be able to provide affected persons with comprehensible information on the functioning of the algorithm. The right to have the means to understand the reasons why AI has achieved its outcome should be guaranteed. This right to information should extend to communicating the conditions of possible human intervention in the decision-making process.

68 See in particular: OECD, Council Recommendation on AI, 13; UNESCO, Recommendation on the Ethics of AI, §§ 37 et seq.

81. Currently, the regulations impose this requirement only with respect to the administration. In fact, it must communicate “*in an intelligible form*” to persons who have been the subject of an individual decision taken on the basis of algorithmic processing, if they so request, the following information⁶⁹:

- The degree and method of contribution of the algorithmic processing to decision-making;
- The data processed and its sources;
- The processing parameters and, where applicable, their weighting, applied to the situation of the affected person;
- The operations carried out by the processing.

82. The CNCDH recommends considering the extension of this obligation to private organisations whose activity may affect the rights of individuals (social networks, banks, insurance, etc.). In this regard, the protection of intellectual property cannot constitute an insurmountable obstacle: it would not mean making the source code of a software public, but rather communicating to the person requesting it information on the elements taken into account by the machine (including the main criteria relating to his or her individual situation), in a language that is easy to understand, in order to explain the process that led to the decision.

83. In subsequent opinions, the CNCDH will examine the proportionality of the restrictions on the requirement of disclosure of such information by the administration and private bodies, when they are likely to infringe a secret protected by law⁷⁰, depending on the fields in question.

84. The explicability requirement applies more broadly to ensure the effectiveness of the right to appeal against individual decisions based on an algorithm.

85. Keeping the person who is the subject of a decision fully informed, by the administration or by a private body, a banking institution for example, means providing him/her with explanations about how the algorithm applied to his/her personal situation works and, more importantly, returning, where applicable, the conditions for human intervention.

Recommendation 19: The CNCDH recommends that administrations communicate in an intelligible form information on the functioning of the algorithm, as well as on the possible part played by human intervention in the decision-making process. It also recommends thinking about extending this obligation to private organisations.

86. In view of the growing importance of the deployment of artificial intelligence systems, the resulting major challenges for the respect of fundamental rights, as well

⁶⁹ R.3111-3-1-2 of the Code of Relations between the Public and the Administration.

⁷⁰ In particular “defence secrecy” and business secrecy.

as the preservation of the rule of law and democracy, not forgetting the environment, the National Consultative Commission on Human Rights intends to continue its work on artificial intelligence in the future, especially to examine its impacts, particularly in the areas of health, education, employment and justice.

SUMMARY OF RECOMMENDATIONS.

Recommendation 1: The CNCDH recommends favouring, in institutional communication, a more neutral and objective terminology than the term “artificial intelligence”, such as “Algorithmic Decision Support System” (ADSS).

Recommendation 2: The Commission recommends strengthening, in the proposal for a European Union regulation on AI, the provisions to ensure the establishment of a binding legal framework guaranteeing effective respect for fundamental rights. In addition, the CNCDH recommends the adoption, within the framework of the Council of Europe, of a “Convention 108+ on AI”.

Recommendation 3: The CNCDH recommends that a human rights-based approach be taken into account in the ongoing reforms, as they aim to ensure respect for fundamental rights.

Recommendation 4: The CNCDH recommends the prohibition of the use of choice interfaces insofar as they have the purpose or effect of manipulating users, to their detriment, by exploiting their vulnerabilities.

Recommendation 5: The CNCDH recommends banning any type of social scoring set up by government authorities or by any company, public or private.

Recommendation 6: The CNCDH recommends prohibiting the remote biometric identification of persons in publicly accessible spaces, by way of exception permitting its use, insofar as it is strictly necessary, adapted and proportionate, for the prevention of a serious and imminent threat to the life or physical safety of persons and that of structures, facilities and establishments of vital importance.

Recommendation 7: The CNCDH recommends continuing and deepening reflection in order to identify the contributions and limits of the use of AI in the context of judicial proceedings.

Recommendation 8: The CNCDH recommends that emotion recognition technologies should be banned, by way of exception permitting their use as long as they aim to reinforce the independence of people, or more broadly the effectiveness of their fundamental rights.

Recommendation 9: The CNCDH recommends that the user of an AI system assess the impact of using this system on fundamental rights and, if risks are identified, carry out an assessment taking into account the probability and severity of these risks. The

impact assessment should include at least:

- a statement of the purpose(s) attached to the use of the envisaged AI system;
- identification of the fundamental rights likely to be affected by the system;
- a review of the envisaged AI system, based on an assessment of its necessity, its suitability and the proportionality to the infringements of fundamental rights in relation to the intended purpose;
- the procedures put in place to monitor the application, and the mitigation measures with regard to the risks incurred.

Recommendation 10: Depending on the risks posed by an AI system to fundamental rights in a particular context of use, the CNCDH recommends ensuring, prior to the decision to use it, a stakeholder consultation, according to appropriate procedures, including for example staff representatives and, more broadly, the persons affected by the AI system.

Recommendation 11: The CNCDH recommends setting up oversight of the AI system, according to a procedure likely to vary according to the risks of infringements of fundamental rights as identified by the impact assessment, in order to maintain continuous vigilance on the part of the user with regard to the effects of the system, including its discriminatory effects.

Recommendation 12: The CNCDH recommends encouraging public investment in the design of training and information tools accessible to as many people as possible.

Recommendation 13: The CNCDH recommends organising national consultations along the same lines as the États généraux de la bioéthique (Bioethics Forum) organised by the National Advisory Committee on Ethics.

Recommendation 14: The CNCDH recommends that the Ministry of Education strengthens the training of students in the technical, political and societal issues surrounding artificial intelligence and proposes, to this end, educational materials for teachers.

Recommendation 15: The CNCDH recommends:

- ensuring human intervention to oversee individual decisions resulting from an AI system in accordance with the latter's risk level;
- ensuring effectiveness thereof through appropriate training and information for personnel on the characteristics of the system, without imposing any particular constraint on them when they deviate from the AI system's recommendation;
- ensuring the systematic maintenance of alternative access to a human agent for public service users.

Recommendation 16: The CNCDH recommends that users of AI systems be given the right to configure their criteria, in particular in order to determine the selection and presentation of the content received, and more generally in the event of human-machine interactions.

Recommendation 17: The CNCDH recommends systematically informing people when they are exposed to or required to interact with an AI system and, when they are the subject of a decision, that this decision is based, where applicable, in part or in full on algorithmic processing.

Recommendation 18: The CNCDH recommends guaranteeing the affected person the right to review, by a human being, of any individual decision based totally or partly on algorithmic processing, provided that it has significant consequences for him/her.

Recommendation 19: The CNCDH recommends that administrations communicate in an intelligible form information on the functioning of the algorithm, as well as on the possible part played by human intervention in the decision-making process. It also recommends thinking about extending this obligation to private organisations.

LIST OF PEOPLE HEARD

Celine Castets-Renard, Professor of Private Law, University of Ottawa Research Chair on Accountable Artificial Intelligence in a Global Context

For the CGT-UGICT: Sylvain Delaitre (CGT Thalès), Jean-Luc Molins (National Secretary), Matthieu Trubert (joint organiser of the UGICT-CGT Digital Collective)

Mona Caroline Chammas, Attorney & Integrity Director GOVERN&LAW

Régis Chatellier, Innovation, Research and Foresight Project Manager at the French Data Protection Authority (CNIL)

Raja Chatilla, INRIA Research Director, Institute of Intelligent Systems and Robotics, Sorbonne University

Defender of Rights, represented by Sarah Bénichou, Assistant to the Director of “Promotion of Equality and Access to Rights” and Gaëtan Goldberg, Head of Digital, Rights and Freedoms

Sonia Desmoulin-Canselier, CNRS Research Officer at the University of Nantes Laboratory of Law and Social Change

Laurence Devillers, Professor of Computer Science Applied to Social Sciences at Paris IV University and IT Researcher at CNRS’s Interdisciplinary Laboratory of Digital Sciences in Saclay

Tim Engelhardt, Head of Human Rights in the Rule of Law and Democracy Section (United Nations High Commissioner for Human Rights)

Jean-Gabriel Ganascia, Professor of Computer Science at the Faculty of Science at Sorbonne University and Researcher in Artificial Intelligence at LIP6

Alexei Grinbaum, Director of Research at the Laboratory for Philosophy of Science (LARSIM) at CEA-Saclay

David Gruson, Member of the Management Board for the Po Paris Sciences Health Programme

Joanne Kirkham, Researcher at the University of Paris II Panthéon-Assas

Claude Kirchner, Director of the National Digital Ethics Pilot Committee

Michel Lansard, ATD Fourth World

Karine Lefevre, Vice-Chair of the National Ethics Advisory Committee

Daniel Le Métayer, Director of Research at the National Institute for Research in Digital Science and Technology (INRIA)

Grégoire Loiseau, Professor of Private Law, University of Paris 1 Panthéon-Sorbonne

Frédéric Marty, CNRS Research Officer at GREDEG, Université Côte d'Azur

Winston Maxwell, Director of Law and Digital Studies at Telecom Paris, Institut Polytechnique de Paris

Yannick Meneceur, Magistrate seconded to the Council of Europe, adviser in digital transformation and artificial intelligence and research associated with IHEJ

Arthur Messaud, La Quadrature du Net

For Numeum: Valentin Hueber, ESN/ICT Delegate; Anissa Kemiche, Head of European Affairs; Katya Lainé, Chairman of the Numeum AI Committee; François Lhemery, Deputy Director of Public Affairs and Communication

Louis Perez, Researcher at the University of Paris II Panthéon-Assas

Patrick Perrot, Gendarmerie Officer, AI Coordinator at the Gendarmerie

Fabien Tarissan, Researcher in Computer Science at CNRS (Section 06), Professor attached to ENS Paris-Saclay

Romain Tinière, Professor of Public Law, Grenoble-Alpes University

Renaud Vedel, Prefect, Coordinator of the National Strategy for Artificial Intelligence

Cédric Villani, Deputy

Serena Villata, Research Officer at CNRS, Laboratory of Computer Science, Signals and Systems (I3S), Sophia Antipolis University



COMMISSION NATIONALE
CONSULTATIVE
DES DROITS DE L'HOMME

RÉPUBLIQUE FRANÇAISE

Created in 1947 at the instigation of René Cassin, **the National Consultative Commission on Human Rights (CNC DH) is the French national institution responsible for promoting and protecting human rights with level 'A' accreditation from the United Nations.**

The CNC DH performs a three-pronged role that involves the following:

- enlightening the public decision-making process with regards to human rights;
- monitoring the effectiveness in France of rights protected by international human rights conventions;
- overseeing France's implementation of recommendations made by international committees.

The CNC DH is independent and operates based on the principle of the pluralism of ideas. This being the case, as the only institution that maintains continuous dialogue between civil society and French experts in the field of human rights, the Committee comprises 64 qualified individuals and representatives of non-governmental organisations with their roots in civil society.

The CNC DH has been an independent National Rapporteur on the fight against all forms of racism since 1990, on the fight against the trafficking and exploitation of human beings since 2014, on "Business and Human rights" since 2017, on the fight against homophobia since 2018 and on the right of persons with disabilities since 2020.

20 Avenue Ségur - TSA 40 720 - 75334 PARIS Cedex 07

Tel : 01.42.75.77.09

Mail : cncdh@cncdh.fr

www.cncdh.fr

 @CNC DH

 @cncdh.france