Provisional version
23 September 2020

Need for democratic governance of Artificial Intelligence

Report[1]
Committee on Political Affairs and Democracy
Rapporteur: Ms Deborah BERGAMINI, Italy, Group of the European People's Party

Summary

Artificial Intelligence (AI) is a part of new reality. Its broad use will increasingly influence various aspects of our lives and transform our societies. This influence may be both beneficial and harmful.

The report focuses on the impact of AI on democracy. It provides an overview of the various ways in which the use of AI-based technology may, and already does, affect the functioning of democratic institutions and processes and the social and political behaviour of citizens. It concludes that the use of AI, and its potential for abuse by States and State agencies, as well as by big private corporations, poses a real threat to the institutions, processes and norms of our rights-based democracies.

To prevent this threat, there is a need for a framework to ensure that this technology is developed and used in full respect of our values, fundamental rights, rule of law and democracy. The report argues that the Council of Europe should play a pioneering role in designing ways and formats to ensure that AI-based technologies are used to enhance, and not to damage democracy.

The Draft resolution welcomes the setting up of an Ad hoc Committee on Artificial Intelligence (CAHAI) and calls upon member States to work together towards a legally binding instrument aimed at ensuring democratic governance of AI and, where necessary, complement it by sectoral legal instruments. It sets out a series of elements that should be taken into account in this process. The Draft recommendation invites the Committee of Ministers to express support to the elaboration of such instrument, possibly in the form of a convention.

---

[1] Reference to Committee: Doc 14868, Ref. 4445 of 12 April 2019.

A.   Draft resolution[2]

1.   Technology has always had a strong impact on the course of human history. Yet, the pace of technological progress has never been as swift, and its effects on humans as direct, tangible and wide-ranging as they are now, at the edge of the Fourth Industrial Revolution. Artificial Intelligence (AI) which is the key driver of it, is broadly considered to be a determining factor for the future of humanity as it will substantially transform individual lives and impact on human communities.

2.   AI-powered devices are already widely present in our daily lives and carry out multiple tasks previously fulfilled by human individuals, both in personal and official capacity. Predictive algorithms, inherent to AI, are frequently deployed for important decisions, such as university admissions, loan decisions and human resources management but also for border control (including at airports) and crime prevention (through predictive policing practices and the use, within the criminal justice system, of risk-assessment instruments in repeat offending). As all our societies are struggling to fight the ongoing Covid-19 pandemic, AI is also used to enhance pharmaceutical research and help analyse medical data.

3.   However, the long-term effects of AI on humans and society are still far from being clear. While AI may generate great opportunities in advancing economic and social progress, it also presents a series of complex challenges. On the one hand, AI is hoped to bring about a substantial increase in productivity and economic growth, scientific breakthroughs, improving health care, higher life expectancy, security and ever-rising convenience. On the other hand, there are fears that AI might severely disrupt labour markets around the globe, lead to an increased income, wealth and social inequality, and jeopardise social and political stability, as well as international security.

4.   AI-based technologies have an impact on the functioning of democratic institutions and processes, as well as on social and political behaviour of citizens. Its use may produce both beneficial and damaging impact on democracy. Indeed, the rapid integration of AI technologies into modern communication tools and social media platforms provides unique opportunities for targeted, personalized and often unnoticed influence on individuals and social groups, which different political actors may be tempted to use to their own benefit.

5.   On the positive side, AI can be used to improve government accountability and transparency, help fight corruption and produce many benefits for democratic action, participation and pluralism, making democracy more direct, efficient and responsive to citizens' needs. AI-based technologies can broaden the space for democratic representation by decentralizing information systems and communication platforms. AI can strengthen informational autonomy for citizens, improve the way they collect information about political processes and help them participate in these processes remotely by facilitating political expression and providing feedback channels with political actors. It can also help establish greater trust between the state and the society and between citizens themselves.

6.   However, AI can be – and reportedly is – used to disrupt democracy through interference in electoral processes, personalised political targeting, shaping voters' behaviours, and manipulating public opinion. Furthermore, AI has seemingly been used to amplifying the spread of misinformation, "echo chambers", propaganda and hate speech, eroding critical thinking, contributing to rising populism and the polarisation of democratic societies.

7.   Moreover, the broad use by States and private actors of AI-based technologies to control individuals such as automated filtering of information amounting to censorship, mass surveillance using smartphones, gathering of personal data and tracking one's activity on- and offline may lead to the erosion of citizens' psychological integrity, civil rights and political freedoms and the emergence of digital authoritarianism – a new social order competing with democracy.

8.   Concentration of data, information, power and influence in the hands of a few big private actors involved in developing and providing AI-based technologies and services, and growing dependence of individuals, institutions and society as a whole on these services, are also a cause for concern. These big companies no longer serve as simple channels of communication between individuals and institutions but play an increasingly prominent role on their own, controlling and filtering information flows, exercising automated censorship of content published on social media, setting the agenda and shaping and transforming social and political models. Acting on the basis of business models prioritising profits of shareholders over common good, these actors may be a threat to democratic order and should not stay beyond democratic oversight.

---

9.   The Assembly notes that, in the last years, governments, civil society, international institutions and companies have been engaged in extensive discussions with a view to identifying a set of commonly accepted principles on how to respond to the concerns related to AI use. It welcomes the fact that the Council of Europe, as a leading human rights organisation, has been actively involved in these discussions on the future of AI and its governance, and in particular the contribution to this process by the Committee of Ministers, the Commissioner for Human Rights and the intergovernmental co-operation bodies.

10.   The Assembly considers that self-regulatory ethical principles and policies voluntarily introduced by private actors are not adequate and sufficient tools to regulate AI as they do not necessarily lead to democratic oversight and accountability. Europe needs to ensure that the power of AI is regulated and used for common good.

11.   Therefore, the Assembly strongly believes that there is a need to create a cross-cutting regulatory framework for AI, with specific principles based on the protection of human rights, democracy and rule of law. Any work in this area needs to involve all stakeholders, including in particular citizens and major private companies involved in developing and providing AI-based technologies and services.

12.   The Council of Europe, as a leading international standard-setting organisation in the field of democracy, must play a pioneering role in designing ways and formats to ensure that AI-based technologies are used to enhance, and not to damage democracy.

13.   In this context, it welcomes the setting up, by the Committee of Ministers, of an Ad hoc Committee on Artificial Intelligence (CAHAI), to examine the feasibility and potential elements based on broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence. It calls upon the Council of Europe member States and other observer States participating in CAHAI to work together towards a legally binding instrument aimed at ensuring democratic governance of AI and where necessary complement it by sectoral legal instruments.

14.   The Assembly deems that such instrument should:

   14.1. guarantee that AI-based technologies are developed, designed and operated in full compliance with, and in support of, Council of Europe's standards on human rights, democracy and the rule of law;

   14.2. promote a common understanding and provide for the respect of key ethical principles and concepts and the implementation of the above-mentioned standards, including:

      14.2.1.   transparency, including accessibility and explicability;
      14.2.2.   justice and fairness, including non-discrimination;
      14.2.3.   human responsibility for decisions, including liability and the availability of remedies;
      14.2.4.   safety and security;
      14.2.5.   privacy and data protection.

   14.3. seek to maximise possible positive impact of AI on the functioning of democratic institutions and processes, including, inter alia:

      14.3.1.   improve government accountability;
      14.3.2.   help fight corruption and economic crime;
      14.3.3.   facilitate democratic action, participation and pluralism;
      14.3.4.   make democracy more direct, efficient and responsive to citizens' needs;
      14.3.5.   broaden the space for democratic representation by decentralizing information systems and communication platforms;
      14.3.6.   strengthen informational autonomy for citizens, improve the way they collect information about political processes and help them participate in these processes remotely by facilitating political expression and providing feedback channels with political actors;
      14.3.7.   improve transparency in public life and help establish greater trust between the state and the society and between citizens themselves;

14.4. contain provisions to prevent and/or limit possibilities that AI is misused to damage and disrupt democracy, including, inter alia, through:

14.4.1. interference in electoral processes, personalised political targeting, shaping voters' political behaviours, and manipulating public opinion;
14.4.2. amplifying the spread of misinformation, "echo chambers" and propaganda;
14.4.3. eroding individual and societal critical thinking;
14.4.4. contributing to rising populism and the polarisation of democratic societies.

14.5. contain provisions to limit the risks of the use of AI-based technologies by States and private actors to control people, which may lead to an erosion of citizens' psychological integrity, civil rights and political freedoms;

14.6. contain safeguards to prevent the threat to democratic order resulting from concentration of data, information, power and influence in the hands of a few big private actors involved in developing and providing AI-based technologies and services, and growing dependence of individuals, institutions and society as a whole on these services, and provisions that the activity of such actors is subject to democratic oversight.

15. Furthermore, the Assembly believes that, in order to ensure accountability, the legal framework to be put in place should provide for an independent and proactive oversight mechanism, involving all relevant stakeholders, that would guarantee effective compliance with its provisions. Such mechanism would require a highly competent body (*inter alia* in technical, legal and ethical terms), capable of following the new developments on digital technology and evaluating accurately and authoritatively its risks and consequences.

16. When it comes to algorithms and social media platforms, the Assembly deems it necessary to:

16.1. make more transparent the decision-making factors behind algorithmically generated content;
16.2. give users more flexibility to decide how algorithms shape their online experience;
16.3. urge platforms to conduct more systematic human rights due diligence in order to understand the social impact of their algorithms;
16.4. consider establishing an independent expert body to provide oversight over tech platforms and the operation of their algorithms;
16.5. tighten privacy controls on user data so that algorithms have less ability to exploit data in the first place.

B. Draft recommendation[3]

1. The Parliamentary Assembly refers to its Resolution XXX (2020) on *Need for democratic governance of Artificial Intelligence* and welcomes the ongoing efforts made so far at national, European and international levels, including at the level of the Council of Europe, to create a regulatory framework for Artificial Intelligence. In the absence of such framework, the Assembly notes that private companies developing and using AI-based technologies have so far opted for a self-regulation policy through soft-law instruments in this field.

2. The Assembly welcomes the work done by the Ad hoc Committee on Artificial Intelligence (CAHAI), which is mandated to examine the feasibility and potential elements based on broad multi-stakeholder consultations, 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.

3. The Assembly is convinced that strong and swift action is needed on the part of the Council of Europe to address the challenges arising from the use of AI-based technologies which may interfere, amongst others, with the functioning of democratic institutions and processes. Soft-law instruments and self-regulation have proven so far not sufficient in addressing these challenges and in protecting human rights, democracy and rule of law. The Council of Europe is in a strategic position to provide the necessary guidance and support, in close co-operation and co-ordination with other European and international institutions and organisations, towards creating a global regulatory framework for Artificial Intelligence.

4. In light of the above, the Assembly recommends that the Committee of Ministers

    4.1. expresses support to the elaboration of a legally binding instrument governing Artificial Intelligence, possibly in the form of a convention, taking into account, *inter alia*, the elements contained in Resolution XXX (2020);

    4.2. ensures that such a legally binding instrument is based on a comprehensive approach, deals with the whole life cycle of AI-based systems, is addressed to all stakeholders, and includes mechanisms to ensure the implementation of this instrument.

---

[3] Draft recommendation adopted unanimously by the committee on 23 September 2020.

C.   Explanatory memorandum by Ms Bergamini, Rapporteur

1.   Introduction

1.1.   Procedure

1.   On 9 April 2019, the Committee on Political Affairs and Democracy initiated a motion for a resolution on the *Need for democratic governance of Artificial Intelligence.*[4] Noting a growing consensus that artificial intelligence (AI) will be a determining factor for the future of humanity, the motion stresses how AI is already influencing the functioning of democracy (e.g. interference in electoral processes, personalised political targeting, shaping voters' behaviours and spreading misinformation to manipulate public opinion). The motion also notes that concentration of power in the hands of a few big private actors, beyond democratic oversight, is a cause for concern. It thus calls for setting up national and international regulatory frameworks to ensure democratic governance of AI and prevent its misuse. The motion was referred to our Committee for report on 12 April 2019 and I was appointed Rapporteur on 25 June 2019.

2.   On 2 October 2019, the Committee held a hearing with the participation of Ms Birgit Schippers, Senior lecturer in Politics at St Mary's University College, Belfast; Mr Paul Nemitz, Principal Advisor, Directorate-General for Justice and Consumers, European Commission; and Mr Yannick Meneceur, Information Society and Action against Crime Directorate, Directorate General Human Rights and Rule of Law, Council of Europe. On 27 January 2020, the Committee held an exchange of views with Mr Dario Fumagalli, legal expert in the field of privacy protection.

3.   In my capacity as Rapporteur, I had the opportunity to represent the Parliamentary Assembly in AI-related events, including the OECD Global Parliamentary Network (Paris, 11 October 2019) where I spoke on the session "How are countries approaching their AI strategies and policies". In May and June 2020, I held exchange of views (by videoconference) with Mr Steven Feldstein, Non-resident Fellow, Democracy, Conflict and Governance, Carnegie Endowment for International Peace; as well as three representatives from Facebook European headquarters in Brussels: Ms Marisa Jimenez Martin, Deputy Director of EU Affairs and Public Policy; Mr Janne Elvelid, Policy Manager EU affairs; and Ms Michela Palladino, EU Public Policy officer. These meetings focused on the ethical use of AI, the role of algorithms in social media platforms and their possible implications for democracy, the need for a regulatory framework which would establish a value- and principles-based system, and for more co-operation between private companies and international organisations.

1.2.   Rationale of the report

4.   As the key driver of the Fourth Industrial Revolution, AI's effect can be seen everywhere and in every aspect of our lives. In its embodied form of robots, it will soon be driving cars, stocking warehouses and caring for the young and elderly. Predictive algorithms, inherent to AI, surround us, whether it is the auto play function on YouTube, a movie recommendation on Netflix or an advertisement on Google search. These algorithms are frequently deployed for loan decisions, university admissions and recruitment but also for police work, at airports, borders or in judicial decisions.[5] As all our societies are struggling to fight the ongoing Covid-19 pandemic, AI is also used to enhance pharmaceutical research and help analyse medical data.[6]

5.   AI holds the promise of solving some of society's most pressing issues, but also presents challenges such as inscrutable "black box" algorithms,[7] potential bias and discrimination in the modelling and outcome of data analysis-based tools, perpetuation of biases through the use of historical data, potential job displacement, quasi-absence of women in tech careers and unethical use of data. Having led to unprecedented access to and exchange of information, AI has also amplified some negative trends contributing to rising populism and the polarisation of democratic societies.

---

[4] Doc. 14868.
[5] *Deutsche Bank Research*, "Digital Politics. AI, big data and the future of democracy", August 2019.
[6] The Assembly President, the Rapporteur on "AI in health care: medical, legal and ethical challenges ahead", Ms Selin Sayek Böke, and I made statements to stress that the use of AI in the context of the Covid-19 pandemic, including for tracing and tracking, should be responsible, ethical and comply with human rights standards upheld by the Council of Europe. The issue of privacy and data protection in the context of tracking of patients and tracing of their contacts is dealt with by the report on "The impact of the Covid-19 pandemic on human rights and the rule of law", adopted by the Committee on Legal Affairs and Human Rights on 29 June 2020. Issues related to health care and AI are examined in Ms Sayek Böke's report to be adopted by the Committee on Social Affairs, Health and Sustainable Development.
[7] See the Council of Europe glossary on Artificial Intelligence.

6.   In the last couple of years, governments, civil society, international institutions and companies have been engaged in extensive talks with a view to identifying a set of commonly accepted principles on how to respond to the complex challenges posed by AI. The Council of Europe, as a leading human rights organisation, has been actively involved in these discussions on the future of AI and its governance. In his last report as Secretary General, Thorbjørn Jagland called for a strategic, transversal approach on AI, developed and applied in line with European standards on human rights, democracy and the rule of law.[8] The Committee of Ministers has on several occasions stressed the potential need to set up a regulatory framework for AI.[9] Similarly, the Parliamentary Assembly has stated in a number of texts the importance of adopting a holistic approach and considering challenges and opportunities related to AI in diversity.[10] In that same spirit, the Assembly initiated a number of reports relating to AI's impact in various fields.

7.   This report focuses on how AI influences and impacts the functioning of democracy, and how multiple stakeholders can engage in and contribute to the dialogue on AI. Above all, it makes a case for creating a common ground where institutions and private companies can establish an open, clear and willing co-operation to build a common democratic AI governance framework.

2.   Artificial Intelligence, definition and ethical principles

8.   Discussions on AI have created a certain amount of unease by those who fear that it will evolve from being a benefit to humanity, to taking it over. However, not everybody is operating from the same definition of the term and while the basic elements are generally the same, the focus of AI shifts depending on the entity that provides the definition. On the other hand, ethical issues associated with AI are proliferating and rising to popular attention as intelligent machines become omnipresent. For example, AI can and do model aspects essential to moral agency, i.e. the ability of individuals or collective entities to make moral judgments, to take ethical decisions based on notions of right or wrong, and to be held accountable for these actions. Therefore, it is essential to be aware of the danger of using AI in order to replace human intelligence in decision-making processes. In fact, machine learning generates algorithms that seem to be very good at making previsions but not in understanding why things happen. The lack of causal thinking is one of the main issues that should lead us to bind AI with strong ethical principles.[11]

9.   Several international organisations have attempted to define the concept. According to the Council of Europe glossary, AI is "a set of sciences, theories and techniques whose purpose is to reproduce by a machine the cognitive abilities of a human being".[12] European Commission defines it as "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals".[13] As for the Organisation for Economic Co-operation and Development (OECD), "an AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy".[14] Finally, UNESCO notes that "while there is no definition of AI, debate tends to focus on machines capable of imitating certain functionalities of human intelligence, including such features as perception, learning, reasoning, problem solving, language interaction, and even producing creative work".[15]

10.   Similarly, while there is an apparent agreement that AI should be ethical, there is a continuing debate about what constitutes ethical AI. In the past five years, private companies, research institutions and public sector organisations have issued numerous sets of principles and guidelines for ethical AI.[16] These guidelines

---

[8] Ready for future challenges, reinforcing the Council of Europe, April 2019.

[9] *Inter alia*, Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling, and Conclusions of the High Level Conference 'Governing the game changer-Impacts of artificial intelligence development on human rights, democracy and the rule of law' (26-27 February 2019). For more information on Committee of Ministers' work on AI, see chapter 5 of this explanatory memorandum.

[10] Recommendation 2102 (2017) "Technological convergence, artificial intelligence and human rights", Resolution 2255 (2019) "Public service media in the context of disinformation and propaganda"; Resolution 2281 (2019) "Social media: social threads or threats to human rights?".

[11] "What AI still can't do", Brian Bergstein, MIT Technology Review, 19 February 2020.

[12] See the Council of Europe glossary on Artificial Intelligence.

[13] Communication from the Commission to the European parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Coordinated plan on Artificial Intelligence, 7 December 2018, COM(2018) 795 final.

[14] Recommendation of the Council on Artificial Intelligence adopted on 22 May 2019, OECD/LEGAL/0449.

[15] Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of artificial intelligence, UNESCO Executive Board, 206th session, October 2019.

[16] According to a recent study conducted by Harvard University, seemingly every organisation with a connection to

tend to agree on some generic principles like transparency, justice, responsibility and privacy, but they seem to sharply disagree over details of what should be done in practice. Also, other principles such as sustainability, dignity and solidarity are significantly underrepresented, suggesting that these issues are currently flying under the radar of the mainstream ethics debate on AI.

11. For the purposes of all Assembly reports on AI, the concept of AI and ethical principles that should apply to AI systems should be understood as described in Appendix I to this document.

3. The impact of Artificial Intelligence on Democracy

12. In this section, largely drawing from studies by Kevin Körner (Deutsche Bank Research, quoted in footnote 5) and Catelijne Muller,[17] as well as from contributions made by experts and Committee members, I would like to provide a mapping of the various ways in which the use of AI-based technologies may, and already does, affect the functioning of democratic institutions and processes, and the social and political behaviour of citizens.

13. Democracy is government of the people by the people for the people. It provides checks against the concentration of power in the hands of a few and can function properly only if based on sound institutions which enjoy confidence of an active, committed and informed citizenry and are able to provide for dynamic balance of interests of constituents. The crisis of modern democracies touches almost all elements of democratic order, including erosion of, and loss of confidence in institutions, mis- and disinformation of the public, break-up of cohesion and polarisation of society. Modern technologies, including AI-based systems may both help resolve and aggravate this crisis.

14. The use of AI by humans is not neutral. It can be used to strengthen government accountability and can produce many benefits for democratic action, participation and pluralism, making democracy more direct and responsive. However, it can also be used to strengthen repressive capabilities and for manipulation purposes. Indeed, the rapid integration of AI technologies into modern communication tools and social media platforms provides unique opportunities for targeted, personalised and often unnoticed influence on individuals and social groups, which different political actors may be tempted to use to their own benefit.

15. The experience of the last few years helps to identify some key areas where the use of AI-based technology can threaten to undermine and destabilise democracy, including, inter alia:

. access to information (misinformation, "echo chambers" and erosion of critical thinking);
. targeted manipulation of citizens;
. interference in electoral processes;
. erosion of civil rights;
. shifts of financial and political power in the data economy.

16. Moreover, the broad use by States of AI-based technologies to control citizens such as automated filtering of information amounting to censorship, and mass surveillance using smartphones and CCTV coupled with vast integrated databases, may lead to the erosion of political freedoms and the emergence of digital authoritarianism – a new social order competing with democracy.

3.1. Access to information - Misinformation, "echo chambers" and erosion of critical thinking

17. A well-functioning democracy requires a well-informed citizenry and implies that people with different views come together to discuss in order to find common solutions through dialogue. By determining which information is shown and consumed (a website algorithm can selectively guess what information a user would like to see based on information about the user), AI-based technologies used in online media can contribute to advancing misinformation and hate speech, create "echo chambers"[18] and "filter bubbles" which lead individuals into a state of intellectual isolation where there is no place for dialogue, thus eroding critical thinking and disrupting democracy. Also, by prioritising the news and information which users like, algorithms tend to reinforce their opinions, tastes and habits, and limit access to diverging views, thus reducing users' free choice.

---

technology policy has authored or endorsed a set of principles for AI. "Principled artificial intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI", Berkman Klein Center for Internet Society, Harvard University, 2020.
[17] "The impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law", Report by Catelijne Muller, LLM, CAHAI (2020)06-fin.
[18] Echo chamber can be defined as an environment in which individuals encounter only beliefs or opinions that coincide with their own, so that their existing views are reinforced, and alternative ideas are not considered.

18.   When it comes to the role of algorithms in advancing misinformation and hateful speech, most of the focus has been on content moderation, i.e. to what extent algorithms are able to identify and suppress posts that break community standards and cross the line when it comes to spreading bad/false information. However, an equally important and more troubling use of algorithms by companies like Facebook are "content shaping" algorithms. Actually, several AI-based platforms exercise automated censorship (by algorithms defined by owners) of content published on social media by private persons, political actors and even State institutions, and deny, or take off-line, information and views that the owners of platforms dislike, thus restricting freedom of expression.

19.   In fact, content shaping algorithms determine what individual users "see online, including user- generated or organic posts and paid advertisements. Some of the most visible examples of content-shaping algorithms include Facebook's News Feed, Twitter's Timeline, and YouTube's recommendation engine. Algorithms also determine which users should be shown a given advertisement. The advertiser usually sets the targeting parameters (such as demographics and presumed interests), but the platform's algorithmic systems pick the specific individuals who will see the advertisement and determine its placement within the platform."[19]

20.   Facebook's internal research reinforces this view. Their team concluded that "64% of all extremist group joins are due to [their] recommendation tools" and that the majority came from Facebook's "Groups You Should Join" and "Discover" algorithms, thus recognising that their "recommendation systems grow the problem."[20] In other words, despite many technology platforms arguing that they are pursuing a hands-off policy regarding content by simply allowing users to say what they would like and not interfering with their free speech rights, in reality they are silently putting their hands on the scale to determine which posts will be viewed and read by millions, i.e. which posts will go viral. Thus, their algorithms are very much shaping what users see and what users react to. At present, considering that the overriding incentive that Facebook and other platforms follow is revenue and profit, it can be assumed that even when content spreads misinformation, the algorithm will bump up its visibility, as long as it increases user engagement on the site.

3.2.   Targeted manipulation of citizens and interference in electoral processes

21.   Although propaganda and manipulation of information are not new instruments in the political toolbox, AI-based communication technologies have tremendously amplified their scale and outreach. Thanks to AI- based technology, online and social media play an increasingly important role in the political process in order to influence people and to favour or reject/deny partisan interests. Some trends reported by political experts include large scale co-ordinated misinformation, including through "deep fakes;[21] micro-targeting of voters; polarisation of public debate; undermining confidence in democratic institutions, political parties and politicians, as well as public trust in the reliability of information; control of information flow and public opinion.

22.   During elections, AI can be effectively used to engage the voters on an individual level along the entire election process. Chatbots and discussion forums on social media platforms encouraging people to leave comments/feedback/brickbats at the end are all various ways in which the public mood can be gauged. Moreover, AI can help collect all this data in real time and enable party campaigners to alter their campaigns, accordingly, depending on what the public feels about them. In addition, AI can be used to manipulate individual voters. By analysing the unique psychographic and behavioural user profiles of voters, AI is being used to persuade people to vote for a specific candidate or even create a bias against that candidate's opponent, and to strengthen the certainty about their choice.

23.   While micro-targeting for political campaigns may simply be seen as commercial advertising, it may threaten democracy, public debate and voters' choices substantially when the related practices rely on the collection and manipulation of users' data (big data analytics) to anticipate and influence their political opinions and election results (computational propaganda).

24.   The most significant cases of alleged AI-based interference in democratic process relate to the 2016 presidential elections in the United States of America. While the political consulting firm Cambridge Analytica (now defunct), was accused of helping Donald Trump win the election by promoting anti-Hillary Clinton content

---

[19] "It's not just the content, it's the business model: Democracy's online speech challenge", Nathalie Marechal and Ellery Roberts Biddle,17 March 2020, New America.

[20] "Facebook executives shut down efforts to make the site less divisive", Jeff Horwitz and Deepa Seetharaman, The Wall Street Journal, 26 May 2020.

[21] The term deep fake is typically used to refer to a video that has been edited using an algorithm to replace the person in the original video with someone else (especially a public figure) in a way that makes the video look authentic.

among voters, some major news aggregators and "mainstream media" outlets were reported to favour news and video positively covering Clinton and negatively portraying Trump.

3.3. Erosion of civil rights

25.   Data availability and rapid progress in AI systems will see an increased use of predictive analytics, not only by companies, banks and recruiters, but by also government institutions and authorities. If the related shortcomings and risks are not addressed adequately, the technology-based amplification of bias and prejudice, as well as statistical flaws and errors, could lead to an entrenchment of historical inequity. This could undermine protection from discrimination and guarantees of equal treatment, which are enshrined in the constitutions of modern democratic societies as well as the European Convention on Human Rights (Article 14) and other Council of Europe instruments.[22]

26.   AI systems' use to profile, track and identify people and screen, sort and even nudge their behaviour can have a chilling effect on the freedom of expression and the freedom of assembly and association (guaranteed by Articles 10 and 11 of European Convention on Human Rights). Using facial recognition in public areas may interfere with a person's freedom of opinion and expression, simply because of the fact that the protection of 'group anonymity' no longer exists. This could discourage people to attend demonstrations and join in peaceful assembly, which is one of the most important elements of democratic society. Individuals may also prefer to refrain from expressing certain points of view and accessing some sources of information if they fear that the data collected on their activities may be used by AI-powered tools designed to take decisions on them (e.g. recruitment or promotion to a new position).

3.4. Concentration of power in the hands of digital companies

27.   One of the more general concerns about AI technologies in terms of democracy is an unprecedented and un-checked concentration of data, information and power in the hands of a small group of major digital companies which develop and own the algorithms, as well as the centralisation of the Internet itself. These big companies no longer serve as simple channels of communication between individuals and institutions but play an increasingly prominent role on their own, setting the agenda and shaping and transforming social and political models. If too much political power is concentrated in a few private hands which prioritise shareholder value over the common good, this can threaten the authority of democratic states. Thus, there is a clear need to reduce the influence of major private companies on democratic decision-making. Moreover, public-private collaborations in AI and its use in sensitive fields, such as public order; security and intelligence; border control, but also in research and development, blur the boundaries between the responsibilities, processes and institutions of democratic states, and the interests of private corporations[23].

3.5. Mass surveillance and the strengthening of authoritarianism

28.   AI may facilitate abuses of power by States and state agencies: as a dual-usage technology, it can be deployed to undermine important human rights that are integral to the functioning of democracies.[24] Advances in AI-based surveillance technology, such as facial, voice and motion recognition, together with a web of surveillance cameras in public places, allow the tracking of individuals in the real world. These AI capacities have come to the forefront during the Covid-19 pandemic (see footnote 6). As with progress in other technologies, tools for surveillance together with predictive analytics can both be used to increase security, safety or traffic control, as well as enable governments to control large crowds and predict the formation of protest and riots.[25] Thus, AI-driven blanket surveillance measures threaten our right to privacy and to freedom of expression.

3.6. AI and political decision-making

29.   Over the last decades, one has witnessed a certain degree of de-politisation of decision-making. A 2019 survey on Europeans' attitudes towards technology found that a quarter of people would prefer it if policy decisions were made by AI instead of politicians, regardless of the fact that AI decisions are based on statistical

---

[22] Principled artificial intelligence: "Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI", Berkman Klein Center for Internet Society, Harvard University, 2020.
[23] Contribution by Ms Birgit Schippers, Senior Lecturer in Politics, St Mary's University College (Belfast), to the hearing in the Committee, 2 October 2019, AS/Pol/Inf (2019)10.
[24] Idem.
[25] Algorithms and human rights, "Study on the human rights dimensions of automated data processing techniques and possible regulatory implications", Council of Europe, 2017.

correlation of available data and not on a causal relation between an event and a decision.[26] This mindset probably reflects the growing mistrust of citizens towards governments and politicians, and underlines a questioning of the Western model of representative democracy.

30.   This approach can engender passivity amongst voters, rather than encouraging them to question the reasons for the choices made and to be aware of the fact that such choices are rooted in interests (in the noble sense) or values which need not necessarily be unobjectionable, absolute or "scientific" in order to be considered to be valid. This phenomenon is often taken to extremes and ends up exploiting, on a rhetorical level, a form of contemporary *ipse dixit*. Machine-generated decision-making is difficult, even impossible for humans to trace or reconstruct. When unaccountable, black-boxed algorithms take decisions that affect people's lives, especially in sensitive areas, there is a serious danger to the democratic values of transparency, accountability and equality, and to the principle of democratic legitimacy. Automation bias, which is the acceptance of machine-generated decisions, either without or with limited human control, undermines transparency and accountability. AI systems can produce profoundly unjust, unfair and even discriminatory outcomes that undermine democratic processes and institutions, and that impact negatively on individuals, especially on individuals from vulnerable communities.[27]

31.   Accustoming society to accepting choices not on the basis of critical reasoning but according to the dictates of authority is extremely unjust, and therefore harmful, given that it is impossible to establish incontrovertibly who should be regarded by public opinion as an authoritative source. AI-assisted technologies may make people believe that they are making their own choices, whereas in reality they are merely following patterns. In this way, AI may be used as an instrument to abuse direct democracy. More broadly, AI-assisted political decision-making may ultimately lead to establishing a form of automated democracy and depriving humans of autonomy over political processes. Defining political goals may not be left to algorithms and must remain with humans enjoying democratic legitimacy and assuming political and legal responsibility.

32.   Summing up, AI-based technology provides the tools to interfere with the procedures and processes of democracies and undermine democratic institutions. The use of AI, and its potential for abuse by States and state agencies, and by private corporations, poses a real threat to the institutions, processes, and norms of our rights-based democracies. [28] In order to prevent this threat, "We need a framework which ensures that this technology is developed and deployed in full respect not only of our values but also of our written law, fundamental rights, rule of law, democracy and the full body of secondary law <…> The principle must be that nothing can be legal if carried out by AI as an automation process if it would be illegal if it is carried out by human."[29]

4.   Ongoing efforts to create a regulatory framework for AI

33.   As mentioned above, national and international organisations are trying to respond to the concerns related to AI use. This section presents an overview of the actions taken by major international organisations, as well as some national initiatives, aimed at setting up a regulatory framework for AI.

4.1.   United Nations

34.   UNESCO[30] is preparing the first global standard-setting instrument on ethics of artificial intelligence, following the decision of UNESCO's General Conference at its 40th session in November 2019. This inclusive and multidisciplinary process is expected to include consultations with a wide range of stakeholders, including the scientific community, people of different cultural backgrounds and ethical perspectives, minority groups, civil society, government and the private sector. The first version of the draft text of the recommendation has been published online and open for consultation. Inclusiveness, trustworthiness, the protection of environment and privacy are amongst the principles included in this Recommendation.

---

[26] Center for the Governance of Change, "European Tech Insights 2019. Mapping European attitudes to technological change and its governance".

[27] Contribution by Ms Birgit Schippers, Senior Lecturer in Politics, St Mary's University College (Belfast), to the hearing in the Committee, 2 October 2019, AS/Pol/Inf (2019)10.

[28] Idem.

[29] Contribution by Mr Paul Nemitz, Principal Advisor, "Strategies for cross-cutting justice policies and legal actions", Directorate-General for Justice and Consumers, European Commission, to the hearing in the Committee, 2 October 2019, AS/Pol/Inf (2019)10.

[30] *https://en.unesco.org/artificial-intelligence*

35. As it is described in the European Union AI strategy, the European Commission is taking a three-step approach: setting-out the key requirements for trustworthy AI; launching a large-scale pilot phase for feedback from stakeholders; and working on international consensus building for human-centric AI. The Commission has introduced seven key requirements for a trustworthy AI:

- human agency and oversight: AI systems should enable equitable societies by supporting human agency and fundamental rights, and not decrease, limit or misguide human autonomy;
- robustness and safety: trustworthy AI requires algorithms to be secure, reliable and robust enough to deal with errors or inconsistencies during all life cycle phases of AI systems;
- privacy and data governance: citizens should have full control over their own data, while data concerning them will not be used to harm or discriminate against them;
- transparency: the traceability of AI systems should be ensured;
- diversity, non-discrimination and fairness: AI systems should consider the whole range of human abilities, skills and requirements, and ensure accessibility;
- societal and environmental well-being: AI systems should be used to enhance positive social change and enhance sustainability and ecological responsibility;
- accountability: mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes.

36. Finally, in its White Paper[31] presented on 19 February 2020, the Commission envisages a framework for trustworthy AI, based on excellence and trust. In partnership with the private and the public sector, the aim is to mobilise resources along the entire value chain and to create the right incentives to accelerate deployment of AI, including by smaller and medium-sized enterprises. This includes working with member States and the research community, to attract and keep talent. As AI systems can be complex and bear significant risks in certain contexts, building trust is essential. Clear rules need to address high-risk AI systems without putting too much burden on less risky ones.

4.3. OECD

37. In May 2019, the OECD adopted the Recommendation on Artificial Intelligence.[32] The Recommendation identifies five complementary values-based principles for the responsible stewardship of trustworthy AI:

- AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being;
- AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society;
- There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them;
- AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed;
- Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

4.4. Actions at national level

38. National authorities in many countries have also developed strategies and policies to promote and regulate AI. I will quote only a few examples:

- Finland: a dedicated governmental steering group published a national AI strategy[33] in 2017 (the first country in the EU to do so), containing, inter alia, a section on AI ethics;
- Germany published an AI strategy in 2018,[34] focusing on the need to boost research and

---

[31] White Paper on Artificial Intelligence: a European approach to excellence and trust. A broad process of consultation on the White Paper has been launched by the European Commission after its publication, and a review of the different contributions is ongoing with a view to finetuning policy and legislative options which will be undertaken by the Commission in the future.
[32] OECD/LEGAL/0449, adopted on 22/05/2019.
[33] Finland's Age of Artificial Intelligence, Turning Finland into a leading country in the application of artificial intelligence Objective and recommendations for measures, Publications of Ministry of Economic Affairs and Employment 47/2017.
[34] Strategie Künstliche Intelligenz der Bundesregierung, November 2018.

development while ensuring that AI development is socially responsible;

- Russian Federation: The Duma has recently adopted the national strategy for the development of AI until 2030. It refers, *inter alia*, to the principles on which the development and use of AI is based upon (protection of human rights and freedoms, security and transparency). Emphasis is also given to raising public awareness, as well as creating an integrated system for regulating social relations with the development and use of AI technologies;[35]
- United Kingdom: the AI Sector Deal[36] and the House of Lords' report on AI[37] (both published in April 2018) emphasise the importance of robust thinking and policy around AI ethics.

5.   The work of the Council of Europe

39.   The Council of Europe, as a leading human rights organisation, plays a significant role in promoting human rights compliant AI. On 13 February 2019, the Committee of Ministers adopted an important Declaration on the manipulative capabilities of algorithmic processes.[38] The Ministers called on member States to tackle the risk that individuals may not be able to form their opinions and take decisions independently of automated systems, and that they may even be subjected to manipulation due to the use of advanced digital technologies, micro-targeting techniques. Noting that machine learning tools have the growing capacity not only to predict choices but also to influence emotions and thoughts, sometimes subliminally, the Committee of Ministers encouraged states to assume their responsibility to address this growing threat by taking appropriate and proportionate legislative measures against illegitimate interferences, and empowering users by promoting critical digital literacy skills.

40.   On 26 and 27 February 2019, the Helsinki conference on Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law, organised by the Council of Europe and the Finnish Presidency of the Committee of Ministers, stressed that "effective supervisory mechanisms and democratic oversight structures regarding the design, development and deployment of AI must be in place", and that "the functioning democratic processes require an independently informed public, and the encouragement of open and inclusive debates. Public awareness of the potential risks and benefits of AI must be enhanced, and necessary new competencies and skills developed. Due public trust in the information environment and AI applications must be fostered".[39]

41.   On 11 September 2019, the Committee of Ministers set up an Ad hoc Committee on Artificial Intelligence (CAHAI), to examine the feasibility and potential elements based on broad multi-stakeholder consultations, 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.

42.   The Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, adopted on 8 April 2020 in the context of the Covid-19 pandemic, issues a set of guidelines calling on governments to ensure that they do not breach human rights through their own use, development or procurement of algorithmic systems. In addition, as regulators, they should establish effective and predictable legislative, regulatory and supervisory frameworks that prevent, detect, prohibit and remedy human rights violations, whether stemming from public or private actors.

6.   Artificial Intelligence and big companies: accountability and ethics

43.   Ethical initiatives help develop a shared language to discuss and debate social and political concerns. They provide developers, company employees, and other stakeholders a set of high-level value statements or objectives against which actions can be later judged. They are also educational, often doing the work of raising awareness of particular risks of AI both within a given institution and externally, amongst the broader concerned public. But they are not appropriate tools to ensure accountability.

44.   As stressed in a report published by AI Now Institute[40] at New York University in 2019, companies creating AI-based solutions to everything, from grading students to assessing immigrants for criminality, are bound by little more than a few ethical statements they decided on themselves. Although there have been

---

[35] Information provided by Mr Piotr Tolstoi, member of the delegation of the Russian Federation, as an information note to the Rapporteur dated 10 June 2020.
[36] Policy Paper, AI Sector Deal, Updated 21 May 2019.
[37] AI in the UK: ready, willing and able?, House of Lords, Select Committee on Artificial Intelligence, Report of Session 2017–19.
[38] Decl (13/02/2019)1.
[39] Conclusions of the High-Level Conference 'Governing the game changer-Impacts of artificial intelligence development on human rights, democracy and the rule of law' (26-27 February 2019).
[40] https://ainowinstitute.org/AI_Now_2019_Report.pdf.

some efforts by the companies, the authors argue that "The frameworks presently governing AI are not capable of ensuring accountability". "As the pervasiveness, complexity, and scale of these systems grow, the lack of meaningful accountability and oversight – including basic safeguards of responsibility, liability, and due process – is an increasingly urgent concern."

45.   In 2018, Google's CEO Sundar Pichai released a public set of seven "guiding principles" designed to ensure that the company's work on AI will be socially responsible. These ethical principles include the commitment to "be socially beneficial" and to "avoid creating or reinforcing unfair bias." In an article for Financial Times, Mr Pichai called for AI to be regulated, but argued for a more sensible approach, supporting that individual areas of AI development, like self-driving cars and health tech, required tailored rules. [41] It is also worth mentioning that Google launched its own independent ethics board in 2019 but shut it down less than two weeks later following controversy about who had been appointed to it.

46.   Other companies, including Microsoft, Facebook, and police body camera maker Axon, also assembled ethics boards, advisors, and teams. Such developments are encouraging, and it is noteworthy that those at the heart of AI development have declared they are taking ethics seriously. However, as stated above, they do not provide a solid basis to engage companies' responsibility**.**

7.   Conclusions

47.   Democracy implies that people with different views come together to find common solutions through dialogue. Instead of creating a public common space and a common agenda, AI-based communication platforms seem to favour individualistic and polarised attitudes and lead to the emergence of closed Internet communities sharing the same views, thus undermining social cohesion and democratic debate and, in contrast, contributing to proliferation of hate speech, compartmentation and segmentation of society. The fact that full segments of the population are not using platforms due to various gaps in the usage of ICTs (ie. based on gender, age, social origin) also needs to be factored in this reflection. In the European Union, for example, there is a gender gap in digital skills of 11%, with a higher gap for above basic skills and especially for those above 55 years.[42] Private companies which apply the rules of the market and not those of democracy are taking no responsibility for allowing the fuelling hate speech and distributing violent content.

48.   AI-based technologies interfere in the functioning of democratic institutions and processes, and have an impact on social and political behaviour of citizens. All machine learning algorithms work on a classifier structure in which the machine learns to make a set of assumptions about different strands of data. Like all iterative learning processes, machine learning too can suffer from false negative or positive reports. While these reports are common errors in any such study, once such error margin is transferred to political decisions, it can lead to the systematic repression of specific ethnic or social groups, the wrongful implication of suspects or unnecessary systematic profiling of citizens.[43]

49.   There is an obvious gap between the pace of technological development and the regulatory framework. Self-regulatory principles and policies cannot be the only tools to regulate AI as they do not lead to accountability. Europe needs to ensure that the power of AI is regulated and used for common good. Therefore, there is a need to create a regulatory framework for AI, with specific principles based on the protection of human rights, democracy and rule of law. Any work in this area needs to involve all stakeholders, including in particular citizens and private companies. The work of CAHAI, which should eventually lead to setting up a legal framework for democratic governance of artificial intelligence, based on Council of Europe's standards, needs to be fully supported and encouraged.

50.   In order to ensure accountability, the legal framework to be put in place should provide for an independent oversight mechanism that would guarantee effective compliance with its provisions. Without such a mechanism, big ICT companies would simply continue business as usual, and against the huge power and transnational nature of these companies, most States would continue to close their eyes to uncompliant behaviour as a necessary and acceptable cost for pursuing a vital interest.

51.   However, such an oversight mechanism can only be effective if it can be proactive and engaged *ex ante*. Indeed, while it would be important to introduce sanctions for incompliant behaviour, a mechanism that would limit itself to *ex post* penalties and fines - which are usually easily affordable by big private companies no matter the amount - would not achieve the desired outcome. That is because it is often very difficult, if not

[41] "Why Google thinks we need to regulate AI", Sundar Pichai, 20 January 2020.
[42] European Commission, Women in digital scoreboard (2019).
[43] Hamid Akın Ünver, "Artificial Intelligence, Authoritarianism and the Future of Political Systems" (The Centre for Economics and Foreign Policy Studies (EDAM), July 2018).

impossible, to restore the previous situation or "erase the damage" after a given AI technology has been introduced and used, as unethical and/or incompliant with human rights, democracy and rule of law as it may be.

52.   A proactive oversight mechanism requires a highly competent body (*inter alia* in technical, legal and ethical terms), capable of following the new developments on digital technology and evaluating accurately and authoritatively its risks and consequences. It goes without saying that such a body should involve all relevant stakeholders.

53.   More critically, the role of AI in changing the power balance between institutions, political actors, and executive organs needs more structured research. Given the scale of legitimacy and sovereignty problems relating to outsourcing political decisions to algorithms, the role of constitutions, parliaments and the political elites in relation to AI needs to be studied in-depth with a specific focus on how political authority should be situated in the age of automated decisions.[44]

54.   This does not mean that AI cannot be a force for good, or render politics more efficient, or more responsive to citizens' needs. If used well, AI can broaden the space for democratic representation by decentralising information systems and communication platforms. It can bolster informational autonomy for citizens and improve the way they collect information about political processes and help them participate in these processes remotely. Just as AI can be used to strengthen opaqueness and unaccountability, it can also improve transparency and help establish greater trust between the state and the society and between citizens themselves.[45]

55.   For its part, the Council of Europe, as a leading international standard-setting organisation in the field of democracy, should play a pioneering role in designing ways and formats to ensure that AI-based technologies are used to enhance democracy through citizens assemblies, electronic agoras and other deliberative and participatory forms of people's involvement in democratic processes.

---

[44] Hamid Akın Ünver, Op. cit.
[45] Idem.

APPENDIX I

Artificial Intelligence – description and ethical principles

There have been many attempts to define the term "artificial intelligence" since it was first used in 1955. These efforts are intensifying as standard-setting bodies, including the Council of Europe, respond to the increasing power and ubiquity of AI by working towards its legal regulation. Nevertheless, there is still no single, universally accepted 'technical' or 'legal' definition.[46] For the purposes of this report, however, it will be necessary to describe the concept.

The term "artificial intelligence" is generally used nowadays to describe computer-based systems that can perceive and derive data from their environment, and then use statistical algorithms to process that data in order to produce results intended to achieve pre-determined goals. The algorithms consist of rules that may be established by human input, or set by the computer itself, which "trains" the algorithm by analysing massive datasets and continues to refine the rules as new data is received. The latter approach is known as "machine learning" (or "statistical learning") and is currently the technique most widely used for complex applications, having only become possible in recent years thanks to increases in computer processing power and the availability of sufficient data. "Deep learning" is a particularly advanced form of machine learning, using multiple layers of "artificial neural networks" to process data. The algorithms developed by these systems may not be entirely susceptible to human analysis or comprehension, which is why they are sometimes described as "black boxes" (a term that is also, but for a different reason, sometimes used to describe proprietary AI systems protected by intellectual property rights).

All current forms of AI are "narrow", meaning they are dedicated to a single, defined task. "Narrow" AI is also sometimes described as "weak", even if modern facial recognition, natural language processing, autonomous driving and medical diagnostic systems, for example, are incredibly sophisticated and perform certain complex tasks with astonishing speed and accuracy. "Artificial general intelligence", sometimes known as "strong" AI, able to perform all functions of the human brain, still lies in the future. "Artificial super-intelligence" refers to a system whose capabilities exceed those of the human brain.

As the number of areas in which artificial intelligence systems are being applied grows, spreading into fields with significant potential impact on individual rights and freedoms and on systems of democracy and the rule of law, increasing and increasingly urgent attention has been paid to the ethical dimension.

Numerous proposals have been made by a wide range of actors for sets of ethical principles that should be applied to AI systems. These proposals are rarely identical, differing both in the principles that they include and the ways in which those principles are defined. Research has shown that there is nevertheless extensive agreement on the core content of ethical principles that should be applied to AI systems, notably the following:[47]

- *Transparency.* The principle of transparency can be interpreted widely to include accessibility, explainability and explicability of an AI system, in other words the possibilities for an individual to understand how the system works and how it produces its results.
- *Justice and fairness.* This principle includes non-discrimination, impartiality, consistency and respect for diversity and plurality. It further implies the possibility for the subject of an AI system's operation to challenge the results, with the possibility of remedy and redress.
- *Responsibility.* This principle encompasses the requirement that a human being should be responsible for any decision affecting individual rights and freedoms, with defined accountability and legal liability for those decisions. This principle is thus closely related to that of justice and fairness.
- *Safety and security.* This implies that AI systems should be robust, secure against outside interference and safe against performing unintended actions, in accordance with the precautionary principle.
- *Privacy.* Whilst respect for human rights generally might be considered inherent in the principles of justice and fairness and of safety and security, the right to privacy is particularly important wherever an AI system is processing personal or private data. AI systems must therefore respect the binding standards of the EU General Data Protection Regulation (GDPR) and the Council of Europe's data protection convention 108 (and the 'modernised' convention 108+), as applicable.

---

[46] For a wide-ranging overview of attempts to define 'artificial intelligence', see *AI Watch: Defining Artificial Intelligence – Towards an operational definition and taxonomy of artificial intelligence,* Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., and Delipetrev, B., European Commission Joint Research Centre, 2020.

[47] See *AI Ethics Guidelines: European and Global Perspectives,* Draft Report commissioned by the Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI), Ienca & Vayena, March 2020.

The effective implementation of ethical principles in relation to AI systems requires an 'ethics by design' approach, including a human rights impact assessment so as to ensure compliance with established standards. It is not sufficient for systems to be designed on the basis of technical standards only and for elements to be added at later stages in an attempt to evince respect for ethical principles.

The extent to which respect for these principles should be built into particular AI systems depends on the intended and foreseeable uses to which those systems may be put: the greater the potential impact on public interests and individual rights and freedoms, the more stringent the safeguards that are needed. Ethical regulation can thus be implemented in various ways, from voluntary internal charters for the least sensitive areas to binding legal standards for the most sensitive. In all cases, it should include independent oversight mechanisms, as appropriate to the level of regulation.

These core principles focus on the AI system and its immediate context. They are not intended to be exhaustive or to exclude wider ethical concerns, such as democracy (pluralistic public involvement in the preparation of ethical and regulatory standards), solidarity (recognising the differing perspectives of diverse groups) or sustainability (preserving the planetary environment).