



Provisional version

22 September 2020

Committee on Social Affairs, Health and Sustainable Development

Artificial intelligence in health care: medical, legal and ethical challenges ahead

Rapporteur: Ms Selin Sayek Böke, Turkey, SOC

Report¹

A. Draft recommendation²

1. Good health is a precondition for more fulfilling individual lives and progress of society as a whole. Artificial Intelligence (AI) is the latest technological innovation fast impacting health care. Like many technological innovations in health care, it harbours major potential to improve both individual and public health, but also presents risks to individual rights and to public health. Furthermore, the speed of the development and deployment of these technological developments is much faster than that of the legal framework regulating them, which requires close attention of policy makers and politicians.

2. AI applications in health care represent the paradigm shift that is shaping up in health care generally, by moving focus from disease and therapy to self-managed health / well-being / prevention, and away from 'one-size-fits-all' treatment protocols to precision medicine tailored to the individual. In this developing environment, full respect for human rights, including social rights, needs to underpin public policy making for health care, and guide the further technological progress. This is required to ensure that more mature AI mechanisms can be deployed safely from a human rights perspective, and that benefits from innovation are spread fairly and equitably across society.

3. The Parliamentary Assembly notes that the scientific community has urged public debate on the implications of AI applications in health care and highlighted the need for all stakeholders to be more accountable. Policy makers, including parliamentarians, at national, European and international levels must better understand the wide-ranging risks, socio-economic impacts and opportunities inherent in the design, development and deployment of AI technologies in health care so as to seek pragmatic improvements and propose adequate regulatory options that ensure full respect for human dignity and rights through legal and ethical frameworks - as far as possible with a global reach. This requires a collaborative, multidisciplinary approach to defining AI-related risks and challenges in health care.

4. The Covid-19 outbreak has focused attention on the role AI can play through real-time surveillance, assessment and management of disease data. It has also revived the much-needed debate on acknowledging the right to health as a fundamental human right which should be secured through legal instruments and appropriate health-care systems that are publicly provided and ensure universal access.

5. The Assembly heeds with concern the warning from the World Health Organization (WHO) that the existing digital divide and inequalities (within and between countries, as well as societal groups) coupled with the spread of AI might exacerbate the unequal spread of health care and problems of effective access to health care, reduce the number and skills of health professionals, accentuate bias and increase "disrespectful clinical interactions", thus *de facto* worsening health-care inequalities and outcomes. The Assembly recalls that there is a broad global consensus around the essential ethical principles of AI. It strongly supports WHO work on developing ethics guidance on AI in health care on the basis of the shared perspectives of various stakeholders, including the Council of Europe and this Assembly.

6. Given that to date the private sector has driven most of the research and development of AI applications for health care, national public health-care authorities should adopt a strategic approach to coordinating digitalisation policies, research and investment, as well as management and exploitation of

¹ Reference to Committee: Reference no. 4473 of 30 June 2019.

² Draft recommendation adopted unanimously by the Committee on 22 September 2020.

personal data, with a view to ensuring full protection of fundamental rights and striking a healthy balance between individual, business and public interests. In this context, the Assembly reaffirms its call, in Recommendation 2166 (2019) on human rights and business, for the revision of the Recommendation CM/Rec(2016)3 in order to reflect modern challenges and member States' obligations under the European Social Charter (ETS Nos. 35 and 163), including the right to health.

7. The Assembly stresses the pertinence of existing Council of Europe legal instruments, in particular the European Convention on Human Rights (ETS No. 005), the Convention on Human Rights and Biomedicine (Oviedo Convention, ETS No. 164) and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108+) in relation to AI-driven transformations in health care. However, it believes that the scope and depth of these transformations and the undeniable impact of AI technology on human dignity and fundamental rights is such that the Council of Europe as the guardian of human rights should develop a dedicated legal instrument on AI. It thus strongly supports the work of CAHAI (Ad hoc Committee on Artificial Intelligence) towards preparing such a dedicated legal instrument.

8. The Assembly notes that privacy, confidentiality of personal data and informed consent are the cornerstones of patient rights worldwide. At the same time, certain restrictions on the use of personal health data may disable essential data linkages and induce distortions, even errors, in AI-driven analysis. It is debatable whether the anonymisation or pseudonymisation of personal health data are appropriate solutions.

9. The Assembly welcomes the intention of the Council of Europe Committee on Bioethics (DH-BIO) to work on trust, safety and transparency in the application of AI in health care. The Assembly encourages it to take a comprehensive approach, to proceed with this work as a matter of priority, and to seek synergies with other Council of Europe bodies working in this field.

10. Moreover, the Assembly considers that cyber-safety requirements for AI-enabled medical devices (including implantable and wearable health-care products) should further be explored in the framework of the Convention on Cybercrime (CETS No.185), whereas the Expert Committee on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT) could complement DH-BIO work by helping define the liability of stakeholders – from developers to regulatory authorities, intermediaries and users (including public authorities, health-care professionals, patients and the general public) – with regard to the development, maintenance, use of and any damage from medical AI applications.

11. The Assembly therefore recommends that the Committee of Ministers:

- 11.1. instruct CAHAI to prepare a dedicated legal instrument on AI, preferably a binding instrument with a global reach, such as a convention open to non-member states, with an emphasis on human rights implications of AI in general and on the right to health in particular;
- 11.2. involve other relevant Council of Europe bodies in CAHAI work, with a view to adequately covering health-related challenges, notably in terms of privacy, confidentiality and cyber-safety of sensitive personal health data, informed consent and liability of stakeholders;
- 11.3. mandate the Council of Europe Committee on Bioethics (DH-BIO) and the Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data to seek synergies in their work towards guiding member States on good governance of health data, with a view to preventing any sovereign or commercial misuse of personal data through medical AI applications;
- 11.4. update Recommendation CM/Rec(2016)3 on human rights and business in order to reflect modern challenges and member States' obligations under the European Social Charter (ETS Nos. 35 and 163), including the right to health.

12. The Assembly, furthermore, recommends that the Committee of Ministers encourage member States to:

- 12.1. build a holistic national approach, involving national parliaments, to the use of AI technology in health-care services based on multi-stakeholder involvement and accountability, as well as adequate evaluation of socio-economic and human rights impacts, with a view to consolidating their population's full access to public health-care services and giving effect to everyone's right to health as set out in the European Social Charter;

- 12.2. participate more actively in the development and deployment of AI applications for health-care services at national level, and provide for sovereign evaluation and screening of such applications by independent institutions, as well as an exhaustive authorisation-process for their deployment, in particular in public health services, to counter risks to individual rights and public health, in accordance with the precautionary principle;
- 12.3. examine legal and technical options for certification and validation of both publicly and commercially developed AI applications (covering the end-product and every stage of the AI design process) for health at both national and European levels;
- 12.4. strengthen their national human rights impact assessment framework for all health-related AI applications;
- 12.5. guarantee that AI-driven health applications do not replace human judgement completely and that thus enabled decisions in professional health care are always validated by adequately trained health professionals;
- 12.6. elaborate a legal framework for clarifying the liability of stakeholders for the design, deployment, maintenance and use of health-related AI applications (including implantable and wearable medical devices) in the national and pan-European context, redefine stakeholder responsibility for risks and harms from such applications and ensure that governance structures and law enforcement mechanisms are in place to guarantee the implementation of this legal framework;
- 12.7. discuss how to bridge the requirement of strong protection of personal data and the need to use certain types of personal health data for the public good in the context of AI-powered enhancements in public health, in a human-rights compliant fashion, including as regards better preparedness of governance structures to anticipate and manage the pandemic response;
- 12.8. accelerate their accession, if they have not yet done so, to the Oviedo Convention (ETS No. 164) and its protocols, and to the amending Protocol (CETS No. 223) of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data;
- 12.9. adapt their education and training systems towards integrating AI-literacy into the curricula of schools and medical training institutions, with an emphasis on the ethical principles of AI and responsible uses of AI applications;
- 12.10. enhance investments in building the necessary digital infrastructure to overcome the prevailing digital divides and to ensure that AI-driven innovations do not exacerbate existing health inequalities;
- 12.11. engage a national debate on AI for health in order to raise the population's and health professionals' awareness of both the risks and the benefits inherent in the use of AI applications for wellness / health care, in particular with regard to certain commercially developed applications already on the market profiting from current legal voids;
- 12.12. consider options for harmonising interconnectivity of national health data networks and databases so as to enable human-rights compliant data linkages for AI-powered analysis and build "learning health systems".

B. Explanatory memorandum by Ms Sayek Böke, rapporteur

“The interests and welfare of the human being shall prevail over the sole interest of society or science.”

Article 2 of the Convention on Human Rights and Biomedicine (Oviedo Convention)

1. Introduction

1. On 4 July 2019, I tabled a motion for a recommendation on “Artificial intelligence in health care: medical, legal and ethical challenges ahead” (Doc. 14948)³ and on 2 October 2019 I was appointed rapporteur. The motion points to the increasing use of artificial intelligence (AI) applications in health care, such as for health monitoring, drug development, virtual health assistance and physicians’ clinical decision support (notably in diagnostics and choice of optimal treatment). Whilst technological developments in this field are advancing fast, legal and ethical frameworks are lagging behind.

2. The scientific community is urging public debate on the implications of AI applications in health care and the need for all stakeholders to be more accountable. Policy makers at both national and European levels need to better understand risks and opportunities inherent in the design, development and deployment of AI technologies so as to seek pragmatic improvements and propose adequate regulatory options that ensure full respect for human dignity and rights. Moreover, given that to date the private sector has driven most of the research and development of robotic applications for health care, public health care authorities should adopt a strategic approach to coordinating digitalisation policies, research and investment, with a view to full protection of fundamental rights.

3. In this context, the Parliamentary Assembly should examine how the Council of Europe’s standard-setting role could be fully exploited and, if necessary, enhanced in order to guide national decisions. As rapporteur I reviewed the Organisation’s existing human rights framework in order to evaluate the coverage of issues related to AI applications in health care, in particular the European Convention on Human Rights (ETS No. 005), the Convention on Human Rights and Biomedicine (Oviedo Convention, ETS No. 164), the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108+) and the Recommendation “Unboxing Artificial Intelligence: 10 steps to protect Human Rights” by the Council of Europe Commissioner for Human Rights, as well as the more recent Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems⁴. In doing so, I will also refer to substantial work on AI in general that is going on within the Council of Europe (see the appendix for a summary of common definitions and ethical principles) and in the international arena with multiple institutional actors involved.

4. We need a holistic approach of the whole health care cycle – from pre-clinical to clinical and from individualised to epidemic needs. We need a framework where the powers of machines and humans are used to complement each other. As AI applications move from narrow to general in nature, the power is shifting from humans to machines. This is why we must elaborate a framework that keeps the human in the centre of the process. For the purposes of this report I carried out a fact-finding visit to the World Health Organisation (WHO) and the International Labour Office (together with the Committee’s rapporteur on AI and labour markets) in Geneva on 16-17 January 2020; participated in the OECD Global Parliamentary Network meeting (10-11 October 2019) and the international conference on new technologies in health held in Thessaloniki on 21-22 November 2019. I also benefited from the Committee’s exchange of views with Ms Corinna Engelhardt-Nowitzki, Head of the Industrial Engineering department at the University of Applied Sciences in Vienna, on 3 December 2019 in Paris, and held a discussion with representatives of the Council of Europe Bioethics Unit on 30 January 2020 in Strasbourg. Finally, our Committee held an online exchange of views with Ms Effy Vayenna, Professor and Chair of Bioethics in the Department of Health Sciences and Technology of the Swiss Federal Institute of Technology, on 2 June 2020.

2. The promise: AI optimising and enhancing health systems

5. Digitalisation, ultra-fast data processing and new types of network connectivity have been driving medical and technological progress in the last few decades. Algorithmic problem-solving programmes and deep learning of recent years have enabled machines to emulate human analytical capacity and to approximate human decision making to unprecedented levels. Today, medical AI applications are smart enough to help detect disease at an early stage, deliver preventative health services and telemedicine,

³ Reference No. 4473 of 30 September 2019 to Committee on Social Affairs, Health and Sustainable Development.

⁴ Adopted by the Committee of Ministers on 8 April 2020.

optimise diagnostic and treatment decisions, ensure personalised health care and precision medicine, build genomic sequencing databases and discover new treatments or medications. Some medical algorithms already equal or outperform human specialists in narrowly defined tasks (such as in analysing medical imaging for the early detection of some cancers, stroke, pneumonia, heart disease) allowing for faster diagnoses, and some argue that AI-machines outrank humans since they do not get tired, nor do they let emotions interfere in their “decisions”.

6. The move from ordinary computers to machine- and deep-learning algorithms means predictions will be based less on rigid rules set by humans and more on the autonomous learning mechanisms of machines. While this might lead to improved analytical and prediction power through discovery of data patterns that people might miss, it also means that the human intervention and understanding of the algorithms are much more restricted. Any regulatory framework has to take this fast pace of change into account.

7. Alas, medical and technological advancements do not automatically translate into better and more equitable health outcomes even when health expenditure is rising in many countries.⁵ A greater use of AI applications in health promises to increase access of the population to medical services with enhanced quality, safety and efficiency, although each of these aspects is debatable and remains to be proven in real life. We have to ensure that AI applications do not just focus on improving the current standard of care but also spread affordable health care. Indeed, the picture is moving all the time as different stakeholders are testing new approaches and trying to push the limits in the uncharted domains.

8. Here, I would like to give a few illustrations of what AI can do for us to help enhance our health. According to OECD studies, various countries are increasingly using electronic systems and mobile services to underpin medical practice and support public health. This particularly concerns the patients’ files storage and records of medical imaging, which enables medical staff to reduce medication errors and also better coordinate care. Furthermore, interconnection of medical databases with other supporting systems (such as insurance coverage) and AI applications already allows for the detection of fraud and excessive prescription, or to project future health care needs and to better allocate resources towards a “learning health system”. In “learning health systems”, data is collected routinely throughout the process of care and is analysed to improve care, eroding the boundaries between clinical research and care, and having significant regulatory implications.

9. The OECD notes that standards and interoperability of networks remain key challenges to be addressed in order to tap the full potential of such AI-supported systems.⁶ Clearly, we need to standardise the electronic health records and their management, if we want to have an unbiased coverage of data, preserve data integrity (such as through blockchain technology) and secure the comparability of health data across countries, while always ensuring that legal and regulatory frameworks (in particular conventions on data protection and privacy) are taken into account.

10. Private companies and researchers have been exploring the capacity of algorithmic applications to identify the most effective antidepressant medication (considering patients’ specific characteristics), detecting depression and predicting suicide, or managing early stage dementia. Moreover, AI has proven to be highly effective in screening medical imaging to help detect and diagnose pathological health conditions such as pneumonia, breast and skin cancers and eye disease. Robotic tools controlled by AI have also been tested for surgery. AI can be used to link clinical data, research and professional guidelines to assist in making informed treatment decisions. The nexus of laboratories, medical molecules, patients, clinical centres and diagnostic centres enables the development of new treatments for chronic diseases more efficiently. Multinational pharmaceutical companies are racing for new frontline treatments and drugs that can reach patients in need ever faster and ensure personalised medical care.

11. On the side of patient-doctor relations, mobile AI applications (such as virtual health assistants) can ensure real-time monitoring for diagnostics, treatment and observation purposes and help uncover risk factors affecting personal health and wellbeing, as well as recommend healthier behaviour to prevent ill-health and to issue alerts for pathologies in the making. Specialists believe that such technologies combining sensors and analysis can be particularly useful in optimising medical care for the elderly and persons with disabilities or enhancing health care through tele-medicine for remote and isolated locations. Importantly, AI could relieve physicians from certain time-consuming clerical tasks and could increase their time for caregiving practices. But – it also bears the risk of technology being used to replace the human caregivers.

12. In terms of public health management, AI is helping to detect infectious disease breakouts and sources of epidemics early, to survey epidemics, to identify unforeseen adverse effects of both new medicines and

⁵ “Tackling Wasteful Spending on Health”, OECD, 10 January 2017.

⁶ “Artificial intelligence in society”, OECD, published on 11 June 2019.

external factors on human health, to better understand and tackle multiple health risk-factors (such as those stemming from chemical molecules used in food production, industry and households) and to plan investments in health care for public research, medical infrastructure and the timely training of specialists. AI could also help in achieving health-related SDG goals.

3. New risks and challenges from medical, legal and ethical angles

13. Although today's AI applications in health are still mostly narrow in scope and limited to specific problem-solving tasks, technological developments are advancing very fast. To perform in an optimal manner, all AI algorithms require huge amounts of data; in health care, a substantial part of such data derives from individuals, and is of a particularly sensitive nature. This raises issues of the adequate protection of personal data and risks to privacy. Moreover, health-related information is highly sensitive in that any bias in the functioning of an algorithm could lead to inadequate prescriptions of treatment and subject entire population groups to unwarranted risks that may threaten not only rights but also lives. At the same time, certain restrictions on the use of personal health data may disable essential data linkages and induce distortions, if not errors, in AI-driven analysis. It is debatable whether the anonymisation of personal data could be an appropriate solution.

14. Is the current personal data protection framework sufficient to deal with the threats that AI bears concerning the use of such data and privacy? This points to the need to examine how some existing legal instruments – such as the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) and its amending Protocol (CETS No. 223) - apply in the context of growing use of AI in health care. As it were, the amending Protocol (opened to signature on 10 October 2018 and signed by 36 countries and ratified by five countries at the time of writing these lines)⁷ foresees “new rights for the persons in an algorithmic decision-making context, which are particularly relevant in connection with the development of artificial intelligence”.

15. In what concerns fundamental rights in the AI decision-making context, the population's awareness of the use of algorithmic applications in the field of health and understanding of implications of such usage is highly important to make the health care system more transparent, to build trust of all users and to ensure informed user consent. This may require the establishment of a national health-data governance framework which could build on proposals from the international institutions. The latter include the Recommendation “Unboxing Artificial Intelligence: 10 steps to protect Human Rights” by the Council of Europe Commissioner for Human Rights (May 2019), the Ethics Guidelines for Trustworthy AI put forward by the EU (April 2019), the OECD Recommendation and Principles on AI (of May 2019) and the G20 Principles on Human-centred Artificial Intelligence (of June 2019). Such a health-data governance structure has to be an integral part of democratic governance structures and should be independent from any political pressure by States and any interference by big firms. Moreover, “algorithm literacy” of the population and health care professionals should be planned already now.

16. AI is not limited to big data; it also encompasses algorithms, computing power and health-related products, rendering both personal data protection and product safety critical. One major concern with AI in health is safety risks for the users of implantable and wearable health care devices. These could be affected by either a commercial misuse or a malicious takeover, inducing real bodily harm to patients just as IT hackers and viruses do to computers and their networks. Moreover, there is a question of legal liability with such devices and the use of data-evidence from such devices in courts. I obtained more information in this respect through my participation in the International Conference on “New technologies in health: medical, legal and ethical issues” (held in Thessaloniki on 21-22 November 2019) and include further comments on this matter in the chapter devoted to legal issues.

17. There is also a big question of trust regarding some AI applications in health. For instance, AI applications that were developed by commercial entities, or even some States which have a different understanding of personal freedoms and rights, may have a built-in bias (indeed, all AI applications are as biased as their creators, which does not bode well for women living in still largely patriarchal societies, nor for all other vulnerable or disadvantaged population groups). To avoid this kind of interference and to protect the population from potential harm, Council of Europe member States should participate more actively in the development of AI applications for health care services, or at least provide some sort of sovereign screening and authorisations for their deployment. States' involvement would also help to ensure that such applications are fed with sufficient, unbiased and well protected data. I therefore welcome the intention of the Council of Europe Committee on Bioethics to work on trust, safety and transparency in the application of AI in health

⁷ The amending Protocol will enter into force when all Parties to Treaty ETS 108 ratify it, or on 11 October 2023, if there are 38 Parties to the Protocol on this date.

care, and would like to encourage it to take a comprehensive approach and to proceed with this work as a matter of priority without delay.

18. Health is a fundamental human right. The potential benefits of AI in improving health conditions imply that AI promises to advance human rights. However, AI also bears the risk of challenging human rights by perpetuating existing societal biases through biases in data and algorithms, and the opacity of increasingly complex AI processes. As such an ethical debate regarding AI in health care becomes critical.

19. Setting ethical boundaries for AI usage in general and in health care in particular will not be easy. This requires public debate and involvement of specialists in several domains. As my fellow colleague observed during a recent debate at the OECD, businesses are speeding ahead with the development and use of commercial AI applications, whereas lawmakers are only contemplating possible legal safeguards. It is time to seek to close the gap.

20. Due diligence and quality control are necessary for any innovation and new technology, as is the case for AI processes. Several features of AI inherently create challenges and require due diligence and quality control. First of all, all algorithms are as good as the data used; this requires due diligence in the quality and the nature of the data. Second, algorithms are inherently biased, reflecting the biases existing in the data as well as those of algorithm designers; due diligence is needed in testing for biases rather than just throwing more data at the problem. Third, the digital nature of AI and the inherent characteristics of machine-learning can lead to blind spots in algorithms, which require due diligence towards these possible adversaries that originate from cybersecurity, computer security and other forms of computing malfunctioning. Finally, AI automates decision-making processes and alters the need to accumulate decision-making skills: due diligence is needed for all stakeholders to avoid the risk of de-skilling and a hollowing out of decision-making, and to keep the human at the centre of AI.

21. AI bears the potential of contributing to the well-being of the people, and to inclusive and sustainable development. Moreover, it can contribute to achieving the SDGs pertaining to health among others. As noted in OECD reports, this requires both public and private investment in AI-related research and development, with an interdisciplinary focus not only on technical issues but also with a social, legal and ethical perspective. The public sector should play a critical role in providing a regulatory framework that ensures respect of privacy and data protection, making data-sharing safe, fair, legal and ethical, minimising or eliminating the biases in databases and building mechanisms to enhance trustworthiness of datasets, algorithms and AI processes. Governments should also establish and support public and private sector oversight mechanisms of AI systems, ranging from compliance reviews, audits, conformity assessment to certification schemes.

4. Current state of policy frameworks

22. Many information and reference tools on AI, including for health care, are in the making. However, as we learned during the OECD Global Parliamentary Network meeting (10-11 October 2019), the United Kingdom is already implementing the Project ExplAIIn to create practical guidance explaining AI decisions to the ordinary public; Japan has AI Utilisation Guidelines for enhanced explainability of AI systems and their outcomes; Denmark has partnered with industry to develop a data ethics seal, Malta has launched a voluntary certification system for AI and Canada has an Algorithmic Impact Assessment tool to evaluate the potential impact of algorithms on citizens, including when they relate to health. WHO and the International Telecommunication Union have started developing a benchmarking process for AI health models aiming to provide global actors with independent and standardised evaluation frameworks.

23. In the last few years, various stakeholders at sectoral, national and international levels have produced a series of soft-law guidelines for a more ethical and rights-based approach to using AI technology in different fields – including for health. “The global landscape of AI ethics guidelines” study⁸ identified, in 2019, 84 major sets of such guidelines, with 5 of them dealing specifically with health and health care. The analysis shows that about 23% of guidelines were developed by private companies and 21% by governmental agencies; they are followed by academic research institutions (11%), international organisations (9.5%), as well as non-profit entities and professional associations (8% each) amongst others.

24. Geographically speaking, the USA and the United Kingdom have been leading the way, accounting respectively for 25% and 16%, or more than a third of all guidelines; the other more active countries include Japan (5%), Germany, France and Finland (some 4% each). Importantly, the British stakeholders – the Royal

⁸ “The global landscape of AI ethics guidelines”, Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland; published online on 2 September 2019.

College of Physicians, the UK Department of Health and Social Care, and the Future Advocacy⁹ – have elaborated 3 out of 5 existing guidelines for the health sector.¹⁰ The same meta-study identified the leading ethical principles (in the order of importance) as follows: transparency / understandability / disclosure, justice and fairness, non-maleficence / security / safety, responsibility, privacy, beneficence, freedom and autonomy / consent, trust, sustainability, dignity and solidarity. Although there is an apparent convergence around the key ethical issues, “the devil is in the detail” due to the interpretation of, the importance attached to, the pertinence in specific fields and the implementation-oversight challenges concerning the underlying principles.

25. Another major study on “Principled AI: mapping consensus in ethical and rights-based approaches to principles for AI”¹¹, based on 36 reference documents sourced worldwide, has identified eight key areas of convergence ranked in the order of importance and seen as a ‘normative core’: trust, fairness and non-discrimination, privacy, accountability, transparency and explainability, safety and security, professional responsibility, human control of technology, and promotion of human values / well-being. It is understood, however, that the core principles are only the beginning: they should be embedded in a general governance framework and reflect their “cultural, linguistic, geographic, and organisational context”, something that was also highlighted in this Committee’s earlier discussions on AI in health. WHO is currently working on a guidance document on the “Ethics and governance of artificial intelligence for health”, which should be finalised by the end of 2020.

26. The Council of Europe also has a number of reference texts – mainly studies, guidelines and recommendations – dealing with the impacts of AI and algorithmic processes on human rights, democracy and the rule of law. Moreover, some of its conventions (such as the European Convention on Human Rights, the Oviedo Convention, Convention 108+ and the Cybercrime Convention) are particularly relevant for assessing the potential impacts of AI in health and identifying regulatory voids. I appreciate also the ongoing dialogue and partnership with a number of private sector internet and telecommunications companies towards mapping regulatory needs more accurately and realistically.

27. The European Union as one of the Council of Europe’s major institutional partners published a white paper on AI and a strategy for data on 19 February 2020,¹² which is to be followed by a public consultation and the revised Coordinated Plan on AI (due for adoption by end 2020). The White Paper advocates a risk-based approach to regulation while leaving room for further developments and urges the adoption of AI by the public sector, including public administrations, hospitals and “other areas of public interest”, with a focus on segments of health care where “technology is mature for large-scale deployment”. So, on the one hand, health care is listed among the high-risk sectors (together with transport, energy, judiciary and some other services) and, on the other hand, enhanced regulation is called for only those uses that would pose heightened risks (of “injury, death or significant material or immaterial damage”) to individuals and legal entities. Moreover, the EU’s GDPR regulation is regularly evoked in relation to personal data processing in a world of ‘big data’. In terms of standardisation, we should note that the European Commission published the Recommendation on a European Electronic Health Record exchange format (2019/243 of 6 February 2019).

28. WHO as a leading global reference point for health has been piloting consultations towards a global strategy on digital health and is working on guidelines on the ethics and governance of AI in the health field which should be published in end-2020. From preliminary studies it appears that quite a lot of AI applications in health are already deployed in high-income countries, and the potential is deemed considerable for extending health care coverage, tapping the needs of the elderly, improving diagnostics, clinical decision making and prevention, developing precision medicine and research, tracking disease outbreaks and surveying public health, as well as reducing health care costs. Both in relation to the Council of Europe and WHO, the private sector companies are showing keen interest and support for the elaboration of soft-law instruments that could pave the way for hard-law regulatory frameworks.

29. In the global discussion on AI ethics, Brent Mittelstadt provides a pertinent viewpoint of comparison between AI-centred ethics initiatives and the four classic principles of medical ethics.¹³ The OECD and the

⁹ A non-partisan think-tank.

¹⁰ The other two codes emanate from the USA and a group of European and North American medical imaging and physicists associations.

¹¹ Research publication No. 2020-1 on “Principled AI: mapping consensus in ethical and rights-based approaches to principles for AI” by the Berkman Klein Center for Internet & Society at Harvard University, Massachusetts, USA; published on 15 January 2020.

¹² “White Paper on Artificial Intelligence - A European approach to excellence and trust” (COM(2020) 65 final) and the Communication on “A European strategy for data” (COM(2020) 66 final) by the European Commission, Brussels, 19 February 2020.

¹³ “Principles alone cannot guarantee ethical AI”, published online in ‘Nature – machine intelligence’ on 4 November 2019 by Brent Mittelstadt, Oxford Internet Institute, University of Oxford, Oxford, UK.

European Commission's High-Level Expert Group on AI seem to endorse this position, centring their guidance for the development of trustworthy AI around the principles of human autonomy, prevention of harm, fairness and explicability. At the same time, four major concerns seem to suggest that this 'principled' approach may have only limited impact on the design and governance of general AI ethics. This is because in comparison to medicine, the current development of AI lacks "(1) common aims [...], (2) professional history and norms, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms". Indeed, the very understanding and interpretation of grand principles in AI is lacking coherence, not to mention the absence of a commonly agreed definition of AI itself.

5. AI for health in real life

30. The potential of AI for health is shaping up through AI-powered applications that reveal a very different degree of maturity in different fields of health care. Private sector investment is driving this process worldwide. Although the role of start-ups is picking up, much of this investment is undertaken by large and mostly multinational firms, who co-operate with medical facilities via accessing patient data. The applications, for example, include training processes in image analysis and patient diagnosis aimed at cancer treatments, among many other applications.

31. Some such applications have led to mixed overall user experience, ranging from successes in correlating data from multiple sources to a lack of added value and failures to advance quality health care based on AI due to lack of accuracy. Critics point to the large firms rushing to capture market shares, and hastily launching personalised medicine without having sufficiently comprehensive data and with promises of results that could not be delivered.

32. This is symptomatic in the digital sector as a whole, where marketing virtual promises may prevail over reality and a quality service offer. This approach is particularly problematic in the area of health services where excessive commercialisation might exacerbate inequalities in access to health care rather than enhancing access and could undermine solidarity as the underlying principle in most European health care systems. Moreover, some experts warn about the risk of 'deskilling' among health professionals if they rely on algorithmic systems more and more to the detriment of critical multifactor analysis and acquisition of experience for making medical judgements.

33. Almost all national and international frameworks of AI emphasise the need to ensure AI is equitable and inclusive. This objective is critical given both the heavy concentration of AI resources among a few firms and a few nations, and the exacerbating role AI could play in the already existing extent of health inequalities both within and between countries. Given the need for a digital infrastructure and digital and algorithmic literacy that would allow for digital connectivity, this could require addressing the existing digital divides. With a special focus on vulnerable groups who have difficulty in accessing health systems, overcoming the digital divides may allow for AI to be conducive in reducing the existing inequalities.

34. As it were, the current trends in AI for health show that machine learning is mainly used in managing chronic diseases (such as for diabetes, with integrated sensors and automated insulin injection), medical imaging analysis and the Internet of Things (with smart wearable devices communicating in real time with professional monitoring). Huge opportunities are seen with AI in medical research for developing new drugs and treatments, provided that limitations of both data sampling and algorithms can be overcome. Indeed, the health data of the poorer population who rarely see health practitioners or of those living in isolated communities ('health deserts') tend to be invisible in health databases, and even health professionals do not consistently record and code correctly all the relevant information on their patients' health condition. Moreover, certain population groups are significantly underrepresented (racial, ethnic, misogynous bias) in randomised clinical trials, which skews both the information collected and conclusions reached, and has particularly dire consequences for children.¹⁴ This 'systemic blindness' unfortunately affects both private sector and public sector research. As policy makers we need to seek ways in which future smarter algorithms and automated data collection (such as via wearable devices and online data queries, should they become accessible to all) could better reflect society as it is.

35. There are clearly positive examples of AI in healthcare provision, where AI supports clinical decision making by physicians and helps patients to better understand and actively manage their health. Ordinary users have the opportunity of entering their symptoms into a smartphone application and see explanations of possible health issues; they can also track those symptoms over time and share this data with physicians.

¹⁴ Article on "Machine learning for clinical decision making: pay attention to what you don't see" by Sherri Rose, Ph.D., associate professor of health care policy at Harvard Medical School and co-author of the first book on machine learning for causal inference – "Targeted Learning". Article published by STAT on 12 December 2019.

Some of these applications have been extended in 140 countries and reached over 15 million users in only a couple of years.

36. A recent example also provides positive applications. In December 2019, AI application developed by the Canadian company Blue Dot was reportedly the first to identify COVID-19 outbreak (which has now turned into a pandemic), while a Chinese AI company has proposed an AI tool for diagnosing COVID-19 in just 10 seconds (compared to about 15 minutes with a manual read) based on lung CT-scan images.¹⁵

6. AI and health care in light of Covid-19

37. The Covid-19 outbreak has changed the focus of all policy-making discussions. The direct health implications of the virus, and the indirect social and economic effects in dealing with the virus create a multifaceted urgency. Since the epicentre of this urgent situation is health related, naturally there is significant debate around how to monitor, predict and manage the Covid-19 outbreak as well as avoid and better manage future such health pandemics. In this debate AI plays a central role, alongside the much-needed debate of acknowledging that the right to health is among the very basic human rights and should be mapped into an appropriate healthcare system that is publicly provided and ensures universal access.

38. AI played a critical role in the initial detection of the pandemic. It has been used in tracking hospital capacity as well as the spread of the disease, in identifying high-risk patients, in developing treatment options and vaccine, and in capacity building for the next pandemic. The demand from politicians in “modelling and tracking” data is probably the most visible application of AI in this area which highlighted both the potential of AI and associated risks. Clearly, the pandemic has been a reminder of both the promise of AI and also the urgent need of striking a balance between protecting the collective interest and individual rights. The crisis has starkly reminded of issues regarding data access, sharing, liability, data and algorithm quality, complementarity of technology and the human, and finally the need for interdisciplinary cooperation and collaboration. It has also urged us to tap the full potential of the already existing frameworks homed by the Council of Europe, such as the European Social Charter that clearly states the “right to health” in Article 11, the Oviedo Convention and the Convention 108+ that ensures protection of personal data and privacy, among others.

39. Despite these existing frameworks, however, a clear need to put all of these into perspective with a focus on AI and health care has become evident. Indeed, had there been a trusted and well-defined regulatory framework, maybe AI could have had a much larger positive impact on the managing of this pandemic; the public’s concerns regarding the misuse and abuse of data by States as well as the private sector would have been mitigated. This experience points to the need to speed up the work, both to contribute to optimising solutions to the current pandemic and to being ready for such events in future. The Covid-19 outbreak has shed light on the most critical aspects of this much needed regulatory framework which should define the extent of public-private dialogue and the respective liabilities, and put in place the conditions and guarantees so that seeking the collective interest does not override the human rights. It should ensure that data and algorithm quality is guaranteed to prevent deepening the existing inequalities, and that technology for monitoring and tracking is only used temporarily - not as a permanent fixture.

40. The pandemic has brought the world to a critical juncture: will surveillance for the sake of health purposes lead to a totalitarian shift or will it be governed through citizen empowerment? Will isolationist reflexes deepen, or will multilateralism, cooperation and solidarity rise to the task? Both questions are relevant in any discussion of AI and healthcare; the former relates to a regulatory framework for the protection of human rights, the latter relates to whether AI in healthcare services will be driven by co-operation and solidarity or by profit-seeking objectives. Clearly, health and privacy can never be alternatives to each other, they can only go hand in hand. A regulatory framework must provide for both and ensure that technology is used for the better. Public trust in both the State and the private sector can only be built up if all agents adequately protect and guarantee the very basic values of human rights in developing and using AI. Given the urgency in using AI as an instrument to assist the fight against the pandemic, it is of utmost importance to agree on at least a workable basic regulatory framework that will enhance trust and make AI operational for the better, for the empowerment of the citizens in making better informed decisions and also by providing information to hold governments accountable for their decisions.

41. The current pandemic is a stark reminder of the inequalities due to over-marketization and the need for regulating markets and governing the potential conflicts between ethical principles and market forces. These are questions also relevant in the debate regarding AI and healthcare. Who owns medical data? Who is allowed to profit from it? Who is liable if AI causes damage? Who will ensure that the use of AI seeks equity above

¹⁵ “Covid-19 and artificial intelligence: protecting health care workers and curbing the spread” by Becky McCall, published on 20 February 2020 in The Lancet; see [https://doi.org/10.1016/S2589-7500\(20\)30054-6](https://doi.org/10.1016/S2589-7500(20)30054-6).

profits? These questions are that much more relevant when we realize that the pandemic has surfaced the need to seek equity and inclusivity in all policies – a critical aspect of AI in healthcare as well. As such, the pandemic shows how important is data transparency and how important it is to ensure unbiased data use.

42. The pandemic has also reminded us that any real progress through the use of technology and AI is only possible if the human takes the leading role. People who understand biomedicine, biology and population models, people who know of infections and virology, people who understand computing – all are needed in unison. Indeed, data and algorithms are only as good as the quality of the data, the knowledge and the expertise of the interdisciplinary teams that develop and service AI. Finally, the global crisis also reminds us how important are multilateralism and collaboration in dealing with global-scale events.

7. Ethical, legal and medical roadmaps

7.1. *Managing sensitive personal data and privacy in health*

43. Most current AI systems and processes rely on huge amounts of data sourced from individuals more or less directly. The spread of social media has already accustomed many of its users to voluntarily surrendering personal data into the global cyberspace in exchange for “free” services. Commercial AI applications in health are now targeting various categories of population in order to build their databases – for free; however, the ambition of many operators is to offer paid services and to turn the potential users of their services into regular consumers. To put it simply, your health is a business opportunity for the private sector, and AI offers new ways of doing business. At the same time, digital interconnections worldwide increasingly render national borders irrelevant and challenge the traditional models of law enforcement, including as regards protection of personal data and privacy.

44. With the development of AI for health care, data is the “fuel” of algorithms and becomes the key source of knowledge, know-how and progress. AI means data, algorithm and computing power. With AI applications in health care, the health data includes health care data as well as health-related lifestyle data, ranging from clinical and genetic data to behavioural and environmental data. As such, health care-related data that is relevant for AI processes comes from multiple sources: electronic health records (EHR), insurance claims, information on prescriptions and laboratory tests, research, wearable fitness devices as well as phones, Internet of Things (IoT) that monitor patients, social media. The General Data Protection Regulation (GDPR) defines health data to cover “personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.”

45. Health privacy can be evaluated from two perspectives: consequentialist concerns or deontological concerns. The former of these is concerned about health-related privacy given the possible tangible negative consequences for individuals if there are violations of privacy. These tangible consequences can range from physical harm to mental pain, embarrassment or paranoia. The deontological concerns relate to privacy even if there are no negative consequences or even if the individual is unaware of such a violation. Both require a well-defined privacy framework when it comes to health data and AI.

46. From a European perspective, the EU's GDPR (General Data Protection Regulation) and the Council of Europe's Convention 108+ aim to increase transparency in data processing and to enhance the protection of sensitive data. Whereas the Convention 108+ defends individual's “human rights and fundamental freedoms, and in particular the right to privacy” with regard to automatic processing of personal data relating to him / her¹⁶, the GDPR shifts emphasis to data protection and omits earlier references to privacy or private life¹⁷, including in provisions related to health research. GDPR thus defines data protection as an individual right that is limited by public interest reasons and needs to be balanced against other fundamental rights¹⁸.

47. This being said, the EU Charter of Fundamental Rights (2000) refers to respect for private and family life in Article 7. Moreover, this Charter's Article 8 spells out the essential aspects of the right to personal data protection, stressing that “such data must be processed fairly for specified purposes and on the basis of the

¹⁶ Article 1 of the Convention 108+ for the Protection of Individuals with regard to Automatic Processing of Personal Data (2018).

¹⁷ EU's Data Protection Directive (1995) referred to privacy in its Article 1(1): ““In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data”.

¹⁸ Joaquin Cayon-De Las Cuevas, Associate professor of health law at the University of Cantabria and Director of a research group on health law and bioethics (GRIDES) at IDIVAL, presentation on “Data protection in the age of the Big Data analysis: impact on the healthcare sector” during the International Conference on “New technologies in health: medical, legal and ethical issues” in Thessaloniki on 21-22 November 2019.

consent of the person concerned or some other legitimate basis laid down by law” and that “compliance with these rules shall be subject to control by an independent authority”.

48. Considering the ethical benchmarks set out in the Oviedo Convention and the Convention 108+ that affirm primacy of privacy and autonomy of data-subject, explicit (“free and informed”) consent and require anonymisation of personal data, the GDPR model¹⁹ represents a clear shift towards the primacy of public interest (such as for scientific research purposes), a broad consent and the pseudonymisation of personal data. Whilst anonymisation of data is in principle irreversible, with current digital tracing tools it may be not complete; pseudonymisation of data is reversible by activating additional information which is kept separately and must be subject to special protective measures (safeguards which include a separation between those responsible for coding data and the users, opinion of an ethics body and the duty of legal secrecy). Pseudonymisation could be a good solution that would allow individuals to benefit from research breakthroughs and new therapies enabled by their data.

49. Patients do not, most of the time, have sufficient levels of awareness to give free and informed consent. As the complexity of machine learning increases and the interaction of data sources that feeds these complex algorithms and neural networks also becomes more complex, consent becomes a difficult task. Increasing combination of AI and IoT technologies means more data, but this also means the consent the user gives might be surpassed through these dataset interactions and algorithm complexities. Therefore, the question of whether specific or broad consent is more optimal is an issue to be resolved.

50. So far though, privacy notices in line with the GDPR requirements seem to be hardly understandable to the general public. There is even an AI-powered tool – Polisis – which helps users to visualise privacy policies and extract a readable summary on what kind of data is collected, where the data could be sent, and what are a user’s options for opting out of data collection or sharing.

51. As often in research it is not possible to identify the exact purpose of personal data processing at the moment of data collection, individuals should be given the possibility to express their consent to the specific areas of (bio-medical) research. Paradoxically, the consent requirement *per se* does not reduce risks for individuals and might result in ineffective protection for them, whilst also hampering research opportunities opened up by the use of ‘big data’. Council of Europe member States are bound by the Council of Europe conventions and those belonging to the EU also have to abide by the treaties and regulations of the European Community. In the light of “Guidelines on artificial intelligence and data protection” (2019)²⁰ that insist on “privacy-by-design and by default”, it remains to be clarified how this position could be reconciled with the more open-ended approach contained in the GDPR.

52. Given the significant implication of the US-based companies in the commercial exploitation of AI and ‘big data’, it is important to bear in mind the EU-US Data Protection Umbrella Agreement of December 2016 which introduced “high privacy safeguards for transatlantic law enforcement cooperation”. Although its primary aim is to combat serious crime and terrorism, it also seeks to enhance the protection of Europeans’ data in line with the EU rules. However, there is a serious issue with the protection of personal data and privacy when European private AI companies are acquired by global giants. For instance, in the case of Google’s acquisition of British DeepMind Health, personal data of 1.6 million British patients-users of DeepMind was transferred to the US on an “inappropriate legal basis”, according to the ruling of the UK Information Commissioner’s Office in July 2017.

7.2. Defining liability of stakeholders

53. The current opacity of algorithms raises multiple questions with regard to the liability of stakeholders – from developers to regulatory authorities, intermediaries and users (including public authorities, health care professionals, patients and the ordinary public). If AI is to help improve our health, health care and even save lives, the responsibilities of all stakeholders need to be clearly delineated in order to prevent damage and to repair / compensate for harm in the worst-case scenario.

¹⁹ The Report on “Big Data and Health” (SHS/YES/IBC-24/17/3 REV.2, 15 September 2017) of the International Bioethics Committee of UNESCO (United Nations Educational, Scientific and Cultural Organisation) considered that “Big Data can be framed as a common good of humankind”.

²⁰ Guidelines issued by the Consultative Committee of the Convention for the protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108+).

54. The Council of Europe expert study on “Responsibility and AI”²¹ considers possible adverse implications from AI use such as malicious attacks on software, unethical system design or unintended system failure, loss of human control and the “exercise of digital power without responsibility” that can lead to tangible harm to human health, property and the environment. The study argues that voluntary commitments by the high-tech industry “typically lack any enforcement and sanctioning mechanisms and cannot therefore be relied upon to provide effective protection” and points to a ‘responsibility gap’ between the developers of AI applications and their potentially harmful outputs. The study also explains the complexities arising from the ‘many hands’ problem (arising from the involvement of many individuals, organisations, machines/technologies used, software / algorithms and end-users in the conception or operation of AI systems), human-computer interaction and the unpredictable nature of algorithmic systems in generating “potentially catastrophic risks” at unprecedented speed and scale.

55. The study puts forward four main findings: (1) a preventative approach may lead to both the development of collective complaints mechanisms and the strengthening of existing protections; (2) human-rights based legal responsibility means ‘strict responsibility’ that needs no proof of fault and is based on a policy choice of balance between fundamental rights and freedoms; (3) the existing legal structure (‘historical responsibility’) should facilitate the development of effective protection mechanisms and meaningful ‘algorithmic auditing’ via a multidisciplinary engagement of stakeholders; (4) effective protection in the digital era requires adequate governance mechanisms, instruments and institutions to monitor, constrain, oversee and investigate AI systems and, if necessary, sanction faults. States therefore must secure that governance and law enforcement mechanisms duly allocate “prospective and historic responsibility” for the risks and harms arising from AI-type digital technologies and hyper-connectivity.

56. Since AI has a self-learning feature, the concepts of “product”, “damage”, “defect” among many other relevant keywords, require revisiting the European Union’s Product Liability Directive (Directive 85/374/EEC). This directive establishes the principle of “liability without fault” or “strict liability”. However, given the changing nature of AI, a revision of the directive and the liability framework is needed. The legal framework could either be fault-based liabilities (where intention matters) or strict liability (regardless of intent or consent). The choice of legal framework should be pursued by a debate on what the insurance law should cover and what the right incentive structure would be.

57. This question also pertains to M-Health (mobile health) applications. M-health apps are useful to improve the efficiency of the system, for the empowerment of the patients and personalisation of medication and treatment. However, the legal framework has to be worked out. There are two categories of health-related applications, although the distinction is not always very clear: applications for the purpose of prevention, diagnosis and treatment of diseases (medical applications) versus applications relevant to lifestyle, fitness and well-being (non-medical applications). When an application is classified as M-health (m-app), it falls under the directive of medical devices; when it is classified as “wellness / fitness” application, it falls under the General Product Safety Rule. A legal clarification is needed given the implications on data protection and privacy, as well as liability.

58. At the same time, corporate responsibility has to be strengthened from the point of view of business-human rights-AI. As pointed out in PACE Recommendation 2166 (2019) on “Human rights and business – what follow-up to Committee of Ministers Recommendation CM/Rec(2016)3?”, there are good reasons for the Council of Europe to “engage in the work of the United Nations open-ended intergovernmental working group on transnational corporations and other business enterprises with respect to human rights (OEIGWG) on a legally binding instrument on business activities and human rights” – also in the context of the growing digital power of multinational companies, including in the field of health care. I believe that Recommendation CM/Rec(2016)3 also needs to be revised so as to take into account the massive implications of AI deployment by businesses.

59. Specifically in the health care field, according to current laws, we can mainly distinguish between the manufacturer’s liability (notably applying for AI-enabled medical devices or software applications that can be considered as devices), professional liability (where due diligence requirements apply to medical practitioners – also when they use AI tools), insurers’ liability and user liability (by individuals, but also relevant authorities such as for managing digital medical records, systems and data quality). There have been some voices proposing to consider defining a new type of legal term – “e-person” in relation to algorithms or constructing a “divided liability” notion. The argument goes as follows: if the agreed premise is that AI is as intelligent as humans, then AI should be equally responsible and liable as humans. This could require defining an e-person through a registered and identifiable AI which itself is backed up with assets.

²¹ Document DGI(2019)05 prepared by the Expert Committee on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT), rapporteur: Karen Yeung.

60. In this context, another delicate question arises on the means to protect health care professionals and patients from the potential conflict of interests from a hidden bias built-in through certain AI applications that may unduly promote certain medical treatments or pharmaceuticals. We should beware automated decision making that may involve patient profiling and discrimination, and push medical practitioners into validating AI-proposed treatments just because their insurers believe that the use of AI applications can reduce the risk of medical errors (as well as related litigation costs) and the overall costs of health care services.

7.3. *What about informed consent?*

61. For the users of AI in a health care setting it is essential to understand what the sophisticated new software can offer in addition to the existing tools, to be able to trust AI applications and to be properly informed when AI-enabled tools are used. Patients need clear explanations, and doctors need to know which AI applications have been properly conceived, are secure and supported by accurate data. Basically, all users need to have a choice; and to be able to exercise their choice, they need adequate information: AI as a “black box” is not acceptable in view of the risks involved to human health and life.

62. In health care in general, informed consent enables patients to make decisions together with their health care providers in a collaborative manner; it is an ethical and legal obligation for health care professionals. Informed consent implies ability to make a voluntary decision based on explanation of relevant medical information (such as diagnosis, purpose, risks and benefits of the proposed treatment and the possible alternative solutions). With AI-powered health care services, non-professionals are increasingly involved in mainstream and paramedical care, and the explanation of medical information can become complicated by both patient’s and physician’s fears, overconfidence or confusion, as well as the opacity of AI systems.

63. As referred to in the chapter 6.1. above, the very notion of consent in the medical field may be shifting with the spread of AI. It is becoming increasingly difficult for people as individuals to make meaningful decisions on the use(s) of their personal data (especially in the digital environment of ‘click to accept all’) and this loss of control needs to be compensated through better governance (including the enforcement and monitoring of safeguards, sanctioning of breaches, reparation of / compensation for damages). In the medical research field, the emphasis is clearly shifting towards securing broad (and informed) consent whereby individuals are asked to agree to a range of possible research sectors where his / her data could be used, an approach already tested for biobanks. This approach is sometimes replaced by ‘opt-out consent’ which may be less protective for personal data, or a ‘dynamic consent’ (implying regular updating of personal data that may allow different uses over time) when this is possible.

64. The European Parliament resolution of 12 February 2020 on “Automated decision-making processes: ensuring consumer protection and free movement of goods and services” (2019/2915(RSP)) welcomes “the potential of automated decision-making to deliver innovative and improved services to consumers, including new digital services such as virtual assistants and chatbots” and notes that when interacting with a system that automates decision-making, one should be “properly informed about how it functions, about how to reach a human with decision-making powers, and about how the system’s decisions can be checked and corrected”. This is important in general and specifically in the medical field, such as when algorithmic applications are used for public health management and medical research so that a healthy balance be found between individual and collective interests. There is not only the need for the patient to be informed when asked for consent, but also the need to ensure that the health care professionals are also fully informed on the limitations and nature of the AI processes they are using (an issue that once again goes back to the issue of re-skilling and training). AI systems should self-present themselves. Users, either medical professionals or patients, should know they are dealing with AI. The AI should be “explainable, transparent, self-indicating and certified”. As for explainability, it is critical that the end user can reach a human expert on demand at any stage of the process.

65. Certification, or validation, is a source of information for the user who is asked to give consent. Certification should not be limited to only the end product but should rather be applied to all stages of AI development and deployment. Certification could be based on a well-defined “rating / scoring” mechanism. Furthermore, demand for certification could be driven by public authorities, where, for example, funding from any State agency could be tied to the certification of AI. The OECD suggests that data intermediaries could act as certification authorities.

66. Alongside certification, a clear and multi-faceted due diligence assessment could be advised. One approach would be, for example, to require that all AI processes and their stages should undergo a “human rights impact assessment (HRIA)”, “privacy impact assessment”, “social impact assessment”, “ethical impact

assessment” or a combined “human rights, ethical and social impact assessment”.²² Indeed, on the occasion of its 40th session, “Guiding Principles on Human Rights Assessment of Economic Reforms” were presented to the Human Rights Council of UN on 28 February 2019. OECD’s 2019 “AI and Society” report also points to the benefits of such impact assessments, stating that “HRIA or similar processes could ensure by-design respect for human rights throughout the lifecycle of the technology”.

8. The need to defend human-centred AI policies and frameworks in health

67. As rapporteur, I am struck by the speed and scope of changes that the AI technologies are bringing into medicine, with challenges for us all - as individuals and as society. I mean in particular the paradigm shift that is shaping up in health care by moving focus from disease and therapy to health / well-being / prevention, away from ‘one-size-fits-all’ treatment protocols to precision medicine responding to specific individual needs. This needs-based approach should guide public policy making for health care and give direction for further technological progress required to ensure that more mature AI mechanisms can be deployed safely from a human rights perspective and that benefits from innovation are spread fairly. We, as lawmakers and ‘digital citizens’, have to better understand risks that may spin out of human control and secure ‘checks-and-balances’ through law to keep up with the pace of change.

68. Further to my liaison with other PACE rapporteurs working on different aspects of AI, I believe that the Assembly should plead for a coordinated and complementary approach by the Council of Europe member States so that human needs, rights and freedoms be put at the centre of the debate and the wealth of opportunities with AI in health care be fully exploited while minimising the risks of harm. It is important for us to discuss and calibrate the essential safeguards (concerning personal health data protection, informed consent, liability of stakeholders) in order to protect the population adequately but also to foster innovation and synergies in the health care sector with increasing emphasis on preventive medicine. It is time for the Council of Europe to start the elaboration of a dedicated legal instrument on AI - such as a convention open to non-member states – with emphasis on human rights implications in general and the right to health in particular.

²² “AI and Big Data: A blueprint for a human rights, social and ethical impact assessment”, Computer Law and Security Review (2018).

Appendix

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.²³ 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:²⁴

- *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

²³ 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*, Samoil, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., and Delipetrev, B., European Commission Joint Research Centre, 2020.

²⁴ 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.

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.

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).