EUROPEAN
COMMISSION

Brussels, 19.2.2020
COM(2020) 65 final

WHITE PAPER

On Artificial Intelligence - A European approach to excellence and trust

White Paper on Artificial Intelligence
A European approach to excellence and trust

Artificial Intelligence is developing fast. It will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans, and in many other ways that we can only begin to imagine. At the same time, Artificial Intelligence (AI) entails a number of potential risks, such as opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes.

Against a background of fierce global competition, a solid European approach is needed, building on the European strategy for AI presented in April 2018[1]. To address the opportunities and challenges of AI, the EU must act as one and define its own way, based on European values, to promote the development and deployment of AI.

The Commission is committed to enabling scientific breakthrough, to preserving the EU's technological leadership and to ensuring that new technologies are at the service of all Europeans – improving their lives while respecting their rights.

Commission President Ursula von der Leyen announced in her political Guidelines[2] a coordinated European approach on the human and ethical implications of AI as well as a reflection on the better use of big data for innovation.

Thus, the Commission supports a regulatory and investment oriented approach with the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology. The purpose of this White Paper is to set out policy options on how to achieve these objectives. It does not address +the development and use of AI for military purposes.The Commission invites Member States, other European institutions, and all stakeholders, including industry, social partners, civil society organisations, researchers, the public in general and any interested party, to react to the options below and to contribute to the Commission's future decision-making in this domain.

1. INTRODUCTION

As digital technology becomes an ever more central part of every aspect of people's lives, people should be able to trust it. Trustworthiness is also a prerequisite for its uptake. This is a chance for Europe, given its strong attachment to values and the rule of law as well as its proven capacity to build safe, reliable and sophisticated products and services from aeronautics to energy, automotive and medical equipment.

Europe's current and future sustainable economic growth and societal wellbeing increasingly draws on value created by data. AI is one of the most important applications of the data economy. Today most data are related to consumers and are stored and processed on central cloud-based infrastructure. By contrast a large share of tomorrow's far more abundant data will come from industry, business and the public sector, and will be stored on a variety of systems, notably on computing devices working at the edge of the network. This opens up new opportunities for Europe, which has a strong position in

[1] AI for Europe, COM/2018/237 final
[2] https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.

digitised industry and business-to-business applications, but a relatively weak position in consumer platforms.

Simply put, AI is a collection of technologies that combine data, algorithms and computing power. Advances in computing and the increasing availability of data are therefore key drivers of the current upsurge of AI. Europe can combine its technological and industrial strengths with a high-quality digital infrastructure and a regulatory framework based on its fundamental values to **become a global leader in innovation in the data economy and its applications** as set out in the European data strategy[3]. On that basis, it can develop an AI ecosystem that brings the benefits of the technology to the whole of European society and economy:

- for **citizens** to reap new benefits for example improved health care, fewer breakdowns of household machinery, safer and cleaner transport systems, better public services;
- for **business** development, for example a new generation of products and services in areas where Europe is particularly strong (machinery, transport, cybersecurity, farming, the green and circular economy, healthcare and high-value added sectors like fashion and tourism); and
- for services of **public interest**, for example by reducing the costs of providing services (transport, education, energy and waste management), by improving the sustainability of products[4] and by equipping law enforcement authorities with appropriate tools to ensure the security of citizens[5], with proper safeguards to respect their rights and freedoms.

Given the major impact that AI can have on our society and the need to build trust, it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection.

Furthermore, the impact of AI systems should be considered not only from an individual perspective, but also from the perspective of society as a whole. The use of AI systems can have a significant role in achieving the Sustainable Development Goals, and in supporting the democratic process and social rights. With its recent proposals on the European Green Deal[6], Europe is leading the way in tackling climate and environmental-related challenges. Digital technologies such as AI are a critical enabler for attaining the goals of the Green Deal. Given the increasing importance of AI, the environmental impact of AI systems needs to be duly considered throughout their lifecycle and across the entire supply chain, e.g. as regards resource usage for the training of algorithms and the storage of data.

A common European approach to AI is necessary to reach sufficient scale and avoid the fragmentation of the single market. The introduction of national initiatives risks to endanger legal certainty, to weaken citizens' trust and to prevent the emergence of a dynamic European industry.

This White Paper presents policy options to enable a trustworthy and secure development of AI in Europe, in full respect of the values and rights of EU citizens. The main building blocks of this White Paper are:

---

[3] COM(2020) 66 final.

[4] AI and digitalisation in general are critical enablers of Europe's Green deal ambitions. However, the current environmental footprint of the ICT sector is estimated at more than 2% of all global emissions. The European digital strategy accompanying this White Paper proposes green transformation measures for digital.

[5] AI tools can provide an opportunity for better protecting EU citizens from crime and acts of terrorism. Such tools could, for example, help identify online terrorist propaganda, discover suspicious transactions in the sales of dangerous products, identify dangerous hidden objects or illicit substances or products, offer assistance to citizens in emergencies and help guide first responders.

[6] COM(2019) 640 final.

- The policy framework setting out measures to align efforts at European, national and regional level. In partnership between the private and the public sector, the aim of the framework is to mobilise resources to achieve an **'ecosystem of excellence'** along the entire value chain, starting in research and innovation, and to create the right incentives to accelerate the adoption of solutions based on AI, including by small and medium-sized enterprises (SMEs).
- The key elements of a future regulatory framework for AI in Europe that will create a unique **'ecosystem of trust'.** To do so, it must ensure compliance with EU rules, including the rules protecting fundamental rights and consumers' rights, in particular for AI systems operated in the EU that pose a high risk[7]. Building an ecosystem of trust is a policy objective in itself, and should give citizens the confidence to take up AI applications and give companies and public organisations the legal certainty to innovate using AI. The Commission strongly supports a human-centric approach based on the Communication on Building Trust in Human-Centric AI[8] and will also take into account the input obtained during the piloting phase of the Ethics Guidelines prepared by the High-Level Expert Group on AI.

The European strategy for data, which accompanies this White Paper, aims to enable Europe to become the most attractive, secure and dynamic data-agile economy in the world – empowering Europe with data to improve decisions and better the lives of all its citizens. The strategy sets out a number of policy measures, including mobilising private and public investments, needed to achieve this goal. Finally, the implications of AI, Internet of Things and other digital technologies for safety and liability legislation are analysed in the Commission Report accompanying this White Paper.

2. CAPITALISING ON STRENGTHS IN INDUSTRIAL AND PROFESSIONAL MARKETS

Europe is well placed to benefit from the potential of AI, not only as a user but also as a creator and a producer of this technology. It has excellent research centres, innovative start-ups, a world-leading position in robotics and competitive manufacturing and services sectors, from automotive to healthcare, energy, financial services and agriculture. Europe has developed a strong computing infrastructure (e.g. high-performance computers), essential to the functioning of AI. Europe also holds large volumes of public and industrial data, the potential of which is currently under-used. It has well-recognised industrial strengths in safe and secure digital systems with low-power consumption that are essential for the further development of AI.

Harnessing the capacity of the EU to invest in next generation technologies and infrastructures, as well as in digital competences like data literacy, will increase Europe's technological sovereignty in key enabling technologies and infrastructures for the data economy. The infrastructures should support the creation of European data pools enabling trustworthy AI, e.g. AI based on European values and rules.

Europe should leverage its strengths to expand its position in the ecosystems and along the value chain, from certain hardware manufacturing sectors to software all the way to services. This is already happening to an extent. Europe produces more than a quarter of all industrial and professional service robots (e.g. for precision farming, security, health, logistics.), and plays an important role in developing and using software applications for companies and organisations (business-to-business applications such as Enterprise Resource Planning, design and engineering software) as well as applications to support e-government and the "intelligent enterprise".

[7] Although further arrangements may need to be put in place to prevent and counter misuse of AI for criminal purposes, this is outside the scope of this white paper.

[8] COM(2019) 168.

Europe leads the way in deploying AI in manufacturing. Over half of the top manufacturers implement at least one instance of AI in manufacturing operations[9].

One reason for Europe's strong position in terms of research is the EU funding programme that has proven instrumental in pooling action, avoiding duplications, and leveraging public and private investments in the Member States. Over the past three years, EU funding for research and innovation for AI has risen to €1.5 billion, i.e. a 70% increase compared to the previous period.

However, investment in research and innovation in Europe is still a fraction of the public and private investment in other regions of the world. Some €3.2 billion were invested in AI in Europe in 2016, compared to around €12.1 billion in North America and €6.5 billion in Asia[10]. In response, Europe needs to increase its investment levels significantly. The Coordinated plan on AI[11] developed with Member States is proving to be a good starting point in building closer cooperation on AI in Europe and in creating synergies to maximise investment in the AI value chain.

3. SEIZING THE OPPORTUNITIES AHEAD: THE NEXT DATA WAVE

Although Europe currently is in a weaker position in consumer applications and on online platforms, which results in a competitive disadvantage in data access, major shifts in the value and re-use of data across sectors are underway. The volume of data produced in the world is growing rapidly, from 33 zettabytes in 2018 to an expected 175 zettabytes in 2025[12]. Each new wave of data brings opportunities for Europe to position itself in the data-agile economy and to become a world leader in this area. Furthermore, the way in which data are stored and processed will change dramatically over the coming five years. Today 80% of data processing and analysis that takes place in the cloud occurs in data centres and centralised computing facilities, and 20% in smart connected objects, such as cars, home appliances or manufacturing robots, and in computing facilities close to the user ("edge computing"). By 2025 these proportions are set to change markedly[13].

Europe is a global leader in low-power electronics which is key for the next generation of specialised processors for AI. This market is currently dominated by non-EU players. This could change with the help of initiatives such as the European Processor Initiative, which focuses on developing low-power computing systems for both edge and next generation high-performance computing, and the work of the Key Digital Technology Joint Undertaking, proposed to start in 2021. Europe also leads in neuromorphic solutions[14] that are ideally suited to automating industrial processes (industry 4.0) and transport modes. They can improve energy efficiency by several orders of magnitude.

Recent advances in quantum computing will generate exponential increases in processing capacity[15]. Europe can be at the forefront of this technology thanks to its academic strengths in quantum computing, as well as European industry's strong position in quantum simulators and programming environments for quantum computing. European initiatives that aim to increase the availability of quantum testing and experimentation facilities will help apply these new quantum solutions to a number of industrial and academic sectors.

---

[9] Followed by Japan (30%) and the US (28%). Source: CapGemini (2019).
[10] 10 imperatives for Europe in the age of AI and automation, McKinsey (2017).
[11] COM(2018) 795.
[12] IDC (2019).
[13] Gartner (2017).
[14] Neuromorphic solutions means any very large-scale system of integrated circuits that mimic neuro-biological architectures present in the nervous system.
[15] Quantum computers will have the capacity to process in less than seconds many fold larger data sets than today's highest performance computers allowing for the development of new AI applications across sectors.

In parallel, Europe will continue to lead progress in the algorithmic foundations of AI, building on its own scientific excellence. There is a need to build bridges between disciplines that currently work separately, such as machine learning and deep learning (characterised by limited interpretability, the need for a large volume of data to train the models and learn through correlations) and symbolic approaches (where rules are created through human intervention). Combining symbolic reasoning with deep neural networks may help us improve explainability of AI outcomes.

4. AN ECOSYSTEM OF EXCELLENCE

To build an ecosystem of excellence that can support the development and uptake of AI across the EU economy and public administration, there is a need to step up action at multiple levels.

A. WORKING WITH MEMBER STATES

Delivering on its strategy on AI adopted in April 2018,[16] in December 2018 the Commission presented a Coordinated Plan - prepared together with the Member States - to foster the development and use of AI in Europe[17].

This plan proposes some 70 joint actions for closer and more efficient cooperation between Member States, and the Commission in key areas, such as research, investment, market uptake, skills and talent, data and international cooperation. The plan is scheduled to run until 2027, with regular monitoring and review.

The aim is to maximise the impact of investment in research, innovation and deployment, assess national AI strategies and build on and extend the Coordinated Plan on AI with Member States:

- *Action 1: The Commission, taking into account the results of the public consultation on the White Paper, will propose to the Member States a revision of the Coordinated Plan to be adopted by end 2020*

EU-level funding in AI should attract and pool investment in areas where the action required goes beyond what any single Member State can achieve. The objective is to attract over €20 billion[18] of total investment in the EU per year in AI over the next decade. To stimulate private and public investment, the EU will make available resources from the Digital Europe Programme, Horizon Europe as well as from the European Structural and Investment Funds to address the needs of less-developed regions as well as rural areas.

The Coordinated Plan could also address societal and environmental well-being as a key principle for AI. AI systems promise to help tackling the most pressing concerns, including climate change and environmental degradation. It is also important that this happens in an environmentally friendly manner. AI can and should itself critically examine resource usage and energy consumption and be trained to make choices that are positive for the environment. The Commission will consider options to encourage and promote AI solutions that do this together with the Member States.

B. FOCUSING THE EFFORTS OF THE RESEARCH AND INNOVATION COMMUNITY

[16] Artificial Intelligence for Europe, COM(2018) 237.
[17] Coordinated Plan on Artificial Intelligence, COM(2018) 795.
[18] COM(2018) 237.

Europe cannot afford to maintain the current fragmented landscape of centres of competence with none reaching the scale necessary to compete with the leading institutes globally. It is imperative to create more synergies and networks between the multiple European research centres on AI and to align their efforts to improve excellence, retain and attract the best researchers and develop the best technology. Europe needs a lighthouse centre of research, innovation and expertise that would coordinate these efforts and be a world reference of excellence in AI and that can attract investments and the best talents in the field.

The centres and the networks should concentrate in sectors where Europe has the potential to become a global champion such as industry, health, transport, finance, agrifood value chains, energy/environment, forestry, earth observation and space. In all these domains, the race for global leadership is ongoing, and Europe offers significant potential, knowledge and expertise[19]. Equally important is to create testing and experimentation sites to support the development and subsequent deployment of novel AI applications.

- *Action 2: the Commission will facilitate the creation of excellence and testing centres that can combine European, national and private investments, possibly including a new legal instrument. The Commission has proposed an ambitious and dedicated amount to support world reference testing centres in Europe under the Digital Europe Programme and complemented where appropriate by research and innovation actions of Horizon Europe as part of the Multiannual Financial Framework for 2021 to 2027.*

C. SKILLS

The European approach to AI will need to be underpinned by a strong focus on skills to fill competence shortages.[20] The Commission will soon present a reinforcement of the Skills Agenda, which aims to ensure that everyone in Europe can benefit from the green and digital transformations of the EU economy. Initiatives could also include the support of sectoral regulators to enhance their AI skills in order to effectively and efficiently implement relevant rules. The updated Digital Education Action Plan will help make better use of data and AI-based technologies such as learning and predictive analytics with the aim to improve education and training systems and make them fit for the digital age. The Plan will also increase awareness of AI at all levels of education in order to prepare citizens for informed decisions that will be increasingly affected by AI.

Developing the skills necessary to work in AI and upskilling the workforce to become fit for the AI-led transformation will be a priority of the revised Coordinated Plan on AI to be developed with Member States. This could include transforming the assessment list of the ethical guidelines into an indicative "curriculum" for developers of AI that will be made available as a resource for training institutions. Particular efforts should be undertaken to increase the number of women trained and employed in this area.

In addition, a lighthouse centre of research and innovation for AI in Europe would attract talent from all over the world due to the possibilities it could offer. It would also develop and spread excellence in skills that take root and grow across Europe.

[19] The future European Defence Fund and Permanent Structured Cooperation (PESCO) will also provide opportunities for research and development in AI. These projects should be synchronized with the wider EU civilian programmes devoted to AI.

[20] https://ec.europa.eu/jrc/en/publication/academic-offer-and-demand-advanced-profiles-eu

- *Action 3: Establish and support through the advanced skills pillar of the Digital Europe Programme networks of leading universities and higher education institutes to attract the best professors and scientists and offer world-leading masters programmes in AI.*

Beyond upskilling, workers and employers are directly affected by the design and use of AI systems in the workplace. The involvement of social partners will be a crucial factor in ensuring a human-centred approach to AI at work.

D. FOCUS ON SMES

It will also be important to ensure that SMEs can access and use AI. To this end, the Digital Innovation Hubs[21] and the AI-on-demand platform[22] should be strengthened further and foster collaboration between SMEs. The Digital Europe Programme will be instrumental in achieving this. While all Digital Innovation Hubs should provide support to SMEs to understand and adopt AI, it will be important that at least one innovation hub per Member State has a high degree of specialisation in AI.

SMEs and start-ups will need access to finance in order to adapt their processes or to innovate using AI. Building on the forthcoming pilot investment fund of €100 million in AI and blockchain, the Commission plans to further scale up access to finance in AI under InvestEU[23]. AI is explicitly mentioned among the eligible areas for the use of the InvestEU guarantee.

- *Action 4: the Commission will work with Member States to ensure that at least one digital innovation hub per Member State has a high degree of specialisation on AI. Digital Innovation Hubs can be supported under the Digital Europe Programme.*
- *The Commission and the European Investment Fund will launch a pilot scheme of €100 million in Q1 2020 to provide equity financing for innovative developments in AI. Subject to final agreement on the MFF, the Commission's intention is to scale it up significantly from 2021 through InvestEU.*

E. PARTNERSHIP WITH THE PRIVATE SECTOR

It is also essential to make sure that the private sector is fully involved in setting the research and innovation agenda and provides the necessary level of co-investment. This requires setting up a broad-based public private partnership, and securing the commitment of the top management of companies.

- *Action 5: In the context of Horizon Europe, the Commission will set up a new public private partnership in AI, data and robotics to combine efforts, ensure coordination of research and innovation in AI, collaborate with other public-private partnerships in Horizon Europe and work together with the testing facilities and the Digital Innovation Hubs mentioned above.*

[21] ec.europe.eu/digital-single-market/en/news/digital-innovation-hubs-helping-companies-across-economy-make-most-digital-opportunities.
[22] www.Ai4eu.eu.
[23] Europe.eu/investeu.

F. PROMOTING THE ADOPTION OF AI BY THE PUBLIC SECTOR

It is essential that public administrations, hospitals, utility and transport services, financial supervisors, and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities. A specific focus will be in the areas of healthcare and transport where technology is mature for large-scale deployment.

- *Action 6: The Commission will initiate open and transparent sector dialogues giving priority to healthcare, rural administrations and public service operators in order to present an action plan to facilitate development, experimentation and adoption. The sector dialogues will be used to prepare a specific 'Adopt AI programme' that will support public procurement of AI systems, and help to transform public procurement processes themselves.*

G. SECURING ACCESS TO DATA AND COMPUTING INFRASTRUCTURES

The areas for action set out in this White Paper are complementary to the plan presented in parallel under the European data strategy. Improving access to and the management of data is fundamental. Without data, the development of AI and other digital applications is not possible. The enormous volume of new data yet to be generated constitutes an opportunity for Europe to position itself at the forefront of the data and AI transformation. Promoting responsible data management practices and compliance of data with the FAIR principles will contribute to build trust and ensure re-usability of data[24]. Equally important is investment in key computing technologies and infrastructures.

The Commission has proposed more than €4 billion under the Digital Europe Programme to support high-performance and quantum computing, including edge computing and AI, data and cloud infrastructure. The European data strategy develops these priorities further.

H. INTERNATIONAL ASPECTS

Europe is well positioned to exercise global leadership in building alliances around shared values and promoting the ethical use of AI. The EU's work on AI has already influenced international discussions. When developing its ethical guidelines, the High-Level Expert Group involved a number of non-EU organisations and several governmental observers. In parallel, the EU was closely involved in developing the OECD's ethical principles for AI[25]. The G20 subsequently endorsed these principles in its June 2019 Ministerial Statement on Trade and Digital Economy.

In parallel, the EU recognises that important work on AI is ingoing in other multilateral fora, including the Council of Europe, the United Nations Educational Scientific and Cultural Organization (UNESCO), the Organisation for Economic Co-operation and Development's (OECD), the World Trade Organisation and the International Telecommunications Union (ITU). At the UN, the EU is involved in the follow-up of the report of the High-Level Panel on Digital Cooperation, including its recommendation on AI.

The EU will continue to cooperate with like-minded countries, but also with global players, on AI, based on an approach based on EU rules and values (e.g. supporting upward regulatory convergence, accessing key resources including data, creating a level playing field). The Commission will closely monitor the policies of third countries that limit data flows and will address undue restrictions in

[24] Findable, Accessible, Interoperable and Reusable as stated in the Final Report and Action Plan from the Commission Expert Group on FAIR data, 2018, https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

[25] https://www.oecd.org/going-digital/ai/principles/

bilateral trade negotiations and through action in the context of the World Trade Organization. The Commission is convinced that international cooperation on AI matters must be based on an approach that promotes the respect of fundamental rights, including human dignity, pluralism, inclusion, non-discrimination and protection of privacy and personal data[26] and it will strive to export its values across the world[27]. It is also clear that the responsible development and use of AI can be a driving force to achieve the Sustainable Development Goals and advance the 2030 Agenda.

5. AN ECOSYSTEM OF TRUST: REGULATORY FRAMEWORK FOR AI

As with any new technology, the use of AI brings both opportunities and risks. Citizens fear being left powerless in defending their rights and safety when facing the information asymmetries of algorithmic decision-making, and companies are concerned by legal uncertainty. While AI can help protect citizens' security and enable them to enjoy their fundamental rights, citizens also worry that AI can have unintended effects or even be used for malicious purposes. These concerns need to be addressed. Moreover, in addition to a lack of investment and skills, lack of trust is a main factor holding back a broader uptake of AI.

That is why the Commission set out an AI strategy[28] on 25 April 2018 addressing the socioeconomic aspects in parallel with an increase in investment in research, innovation and AI-capacity across the EU. It agreed a Coordinated Plan[29] with the Member States to align strategies. The Commission also established a High-Level Expert Group that published Guidelines on trustworthy AI in April 2019[30].

The Commission published a Communication[31] welcoming the seven key requirements identified in the Guidelines of the High-Level Expert Group:

- Human agency and oversight,
- Technical robustness and safety,
- Privacy and data governance,
- Transparency,
- Diversity, non-discrimination and fairness,
- Societal and environmental wellbeing, and
- Accountability.

In addition, the Guidelines contain an assessment list for practical use by companies. During the second half of 2019, over 350 organisations have tested this assessment list and sent feedback. The High-Level Group is in the process of revising its guidelines in light of this feedback and will finalise this work by June 2020. A key result of the feedback process is that while a number of the requirements are already reflected in existing legal or regulatory regimes, those regarding transparency, traceability and human oversight are not specifically covered under current legislation in many economic sectors.

On top of this set of non-binding Guidelines of the High-Level Expert Group, and in line with the President's political guidelines, a clear European regulatory framework would build trust among

[26] Under the Partnership Instrument, the Commission will finance a €2.5 million project that will facilitate cooperation with like-minded partners, in order to promote the EU AI ethical guidelines and to adopt common principles and operational conclusions.
[27] President Von der Leyen, A Union that strives for more – My agenda for Europe, page 17.
[28] COM(2018) 237.
[29] COM(2018) 795.
[30] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
[31] COM(2019) 168.

consumers and businesses in AI, and therefore speed up the uptake of the technology. Such a regulatory framework should be consistent with other actions to promote Europe's innovation capacity and competitiveness in this field. In addition, it must ensure socially, environmentally and economically optimal outcomes and compliance with EU legislation, principles and values. This is particularly relevant in areas where citizens' rights may be most directly affected, for example in the case of AI applications for law enforcement and the judiciary.

Developers and deployers of AI are already subject to European legislation on fundamental rights (e.g. data protection, privacy, non-discrimination), consumer protection, and product safety and liability rules. Consumers expect the same level of safety and respect of their rights whether or not a product or a system relies on AI. However, some specific features of AI (e.g. opacity) can make the application and enforcement of this legislation more difficult. For this reason, there is a need to examine whether current legislation is able to address the risks of AI and can be effectively enforced, whether adaptations of the legislation are needed, or whether new legislation is needed.

Given how fast AI is evolving, the regulatory framework must leave room to cater for further developments. Any changes should be limited to clearly identified problems for which feasible solutions exist.

Member States are pointing at the current absence of a common European framework. The German Data Ethics Commission has called for a five-level risk-based system of regulation that would go from no regulation for the most innocuous AI systems to a complete ban for the most dangerous ones. Denmark has just launched the prototype of a Data Ethics Seal. Malta has introduced a voluntary certification system for AI. If the EU fails to provide an EU-wide approach, there is a real risk of fragmentation in the internal market, which would undermine the objectives of trust, legal certainty and market uptake.

A solid European regulatory framework for trustworthy AI will protect all European citizens and help create a frictionless internal market for the further development and uptake of AI as well as strengthening Europe's industrial basis in AI.

A. PROBLEM DEFINITION

While AI can do much good, including by making products and processes safer, it can also do harm. This harm might be both material (safety and health of individuals, including loss of life, damage to property) and immaterial (loss of privacy, limitations to the right of freedom of expression, human dignity, discrimination for instance in access to employment), and can relate to a wide variety of risks. A regulatory framework should concentrate on how to minimise the various risks of potential harm, in particular the most significant ones.

The main risks related to the use of AI concern the application of rules designed to protect fundamental rights (including personal data and privacy protection and non-discrimination), as well as safety[32] and liability-related issues.

Risks for fundamental rights, including personal data and privacy protection and non-discrimination

[32] This includes issues of cybersecurity, issues associated with AI applications in critical infrastructures, or malicious use of AI.

The use of AI can affect the values on which the EU is founded and lead to breaches of fundamental rights[33], including the rights to freedom of expression, freedom of assembly, human dignity, non-discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, as applicable in certain domains, protection of personal data and private life, [34] or the right to an effective judicial remedy and a fair trial, as well as consumer protection. These risks might result from flaws in the overall design of AI systems (including as regards human oversight) or from the use of data without correcting possible bias (e.g. the system is trained using only or mainly data from men leading to suboptimal results in relation to women).

AI can perform many functions that previously could only be done by humans. As a result, citizens and legal entities will increasingly be subject to actions and decisions taken by or with the assistance of AI systems, which may sometimes be difficult to understand and to effectively challenge where necessary. Moreover, AI increases the possibilities to track and analyse the daily habits of people. For example, there is a potential risk that AI may be used, in breach of EU data protection and other rules, by state authorities or other entities for mass surveillance and by employers to observe how their employees behave. By analysing large amounts of data and identifying links among them, AI may also be used to retrace and de-anonymise data about persons, creating new personal data protection risks even in respect to datasets that per se do not include personal data. AI is also used by online intermediaries to prioritise information for their users and to perform content moderation. The processed data, the way applications are designed and the scope for human intervention can affect the rights to free expression, personal data protection, privacy, and political freedoms.

Certain AI algorithms, when exploited for predicting criminal recidivism, can display gender and racial bias, demonstrating different recidivism prediction probability for women vs men or for nationals vs foreigners. *Source: Tolan S., Miron M., Gomez E. and Castillo C. "Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia", Best Paper Award, International Conference on AI and Law, 2019*

Certain AI programmes for facial analysis display gender and racial bias, demonstrating low errors for determining the gender of lighter-skinned men but high errors in determining gender for darker-skinned women. Source: *Joy Buolamwini, Timnit Gebru; Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91, 2018.*

Bias and discrimination are inherent risks of any societal or economic activity. Human decision-making is not immune to mistakes and biases. However, the same bias when present in AI could have a much larger effect, affecting and discriminating many people without the social control mechanisms that govern human behaviour[35]. This can also happen when the AI system 'learns' while in operation.

[33] Council of Europe research shows that a large number of fundamental rights could be impacted from the use of AI, https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5.

[34] The General Data Protection Regulation and the ePrivacy Directive (new ePrivacy Regulation under negotiation) address these risks but there might be a need to examine whether AI systems pose additional risks. The Commission will be monitoring and assessing the application of the GDPR on a continuous basis.

[35] The Commission's Advisory Committee on Equal Opportunities for Women and Men is currently preparing an "Opinion on Artificial Intelligence" analysing inter alia the impacts of Artificial Intelligence on gender equality which is expected to be adopted by the Committee in early 2020. The EU Gender Equality Strategy 2020-2024 also addresses the link between AI on gender equality; The European Network of Equality Bodies (Equinet) will publish a report (by Robin

In such cases, where the outcome could not have been prevented or anticipated at the design phase, the risks will not stem from a flaw in the original design of the system but rather from the practical impacts of the correlations or patterns that the system identifies in a large dataset.

The specific characteristics of many AI technologies, including opacity ('black box-effect'), complexity, unpredictability and partially autonomous behaviour, may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights. Enforcement authorities and affected persons might lack the means to verify how a given decision made with the involvement of AI was taken and, therefore, whether the relevant rules were respected. Individuals and legal entities may face difficulties with effective access to justice in situations where such decisions may negatively affect them.

Risks for safety and the effective functioning of the liability regime

AI technologies may present new safety risks for users when they are embedded in products and services. For example, as result of a flaw in the object recognition technology, an autonomous car can wrongly identify an object on the road and cause an accident involving injuries and material damage. As with the risks to fundamental rights, these risks can be caused by flaws in the design of the AI technology, be related to problems with the availability and quality of data or to other problems stemming from machine learning. While some of these risks are not limited to products and services that rely on AI , the use of AI may increase or aggravate the risks.

A lack of clear safety provisions tackling these risks may, in addition to risks for the individuals concerned, create legal uncertainty for businesses that are marketing their products involving AI in the EU. Market surveillance and enforcement authorities may find themselves in a situation where they are unclear as to whether they can intervene, because they may not be empowered to act and/or don't have the appropriate technical capabilities for inspecting systems[36]. Legal uncertainty may therefore reduce overall levels of safety and undermine the competitiveness of European companies.

If the safety risks materialise, the lack of clear requirements and the characteristics of AI technologies mentioned above make it difficult to trace back potentially problematic decisions made with the involvement of AI systems. This in turn may make it difficult for persons having suffered harm to obtain compensation under the current EU and national liability legislation.[37]

Allen and Dee Masters) on "Regulating AI: the new role for Equality Bodies – Meeting the new challenges to equality and non-discrimination from increased digitalisation and the use of AI", expected early 2020.

[36] An example may be the smart watch for children. This product may cause no direct harm to the child wearing it, but lacking a minimum level of security, it can be easily used as a tool to have access to the child. Market surveillance authorities may find it difficult to intervene in cases where the risk is not linked to the product as such.

[37] The implications of AI, Internet of Things and other digital technologies for safety and liability legislation are analysed in the Commission Report accompanying this White Paper.

Under the Product Liability Directive, a manufacturer is liable for damage caused by a defective product. However, in the case of an AI based system such as autonomous cars, it may be difficult to prove that there is a defect in the product, the damage that has occurred and the causal link between the two. In addition, there is some uncertainty about how and to what extent the Product Liability Directive applies in the case of certain types of defects, for example if these result from weaknesses in the cybersecurity of the product.

Thus, the difficulty of tracing back potentially problematic decisions taken by AI systems and referred to above in relation to fundamental rights applies equally to safety and liability-related issues. Persons having suffered harm may not have effective access to the evidence that is necessary to build a case in court, for instance, and may have less effective redress possibilities compared to situations where the damage is caused by traditional technologies. These risks will increase as the use of AI becomes more widespread.

B. POSSIBLE ADJUSTMENTS TO EXISTING EU LEGISLATIVE FRAMEWORK RELATING TO AI

An extensive body of existing EU product safety and liability legislation[38], including sector-specific rules, further complemented by national legislation, is relevant and potentially applicable to a number of emerging AI applications.

As regards the protection of fundamental rights and consumer rights, the EU legislative framework includes legislation such as the Race Equality Directive[39], the Directive on equal treatment in employment and occupation[40], the Directives on equal treatment between men and women in relation to employment and access to goods and services[41], a number of consumer protection rules[42], as well as rules on personal data protection and privacy, notably the General Data Protection Regulation and other sectorial legislation covering personal data protection, such as the Data Protection Law Enforcement Directive[43]. In addition, as from 2025, the rules on accessibility requirements for goods and services, set out in the European Accessibility Act will apply[44]. In addition, fundamental rights need to be respected when implementing other EU legislation, including in the field of financial services, migration or responsibility of online intermediaries.

While the EU legislation remains in principle fully applicable irrespective of the involvement of AI, it is important to assess whether it can be enforced adequately to address the risks that AI systems create, or whether adjustments are needed to specific legal instruments.

---

38 The EU legal framework for product safety consists of the General Product Safety Directive (Directive 2001/95/EC), as a safety net, and a number of sector-specific rules covering different categories of products ranging from machines, planes and cars to toys and medical devices aiming to provide a high level of health and safety. Product liability law is complemented by different systems of civil liability for damages caused by products or services.

39 Directive 2000/43/EC.

40 Directive 2000/78/EC.

41 Directive 2004/113/EC; Directive 2006/54/EC.

42 Such as the Unfair Commercial Practices Directive (Directive 2005/29/EC) and the Consumer Rights Directive (Directive 2011/83/EC).

43 Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.

44 Directive (EU) 2019/882 on the accessibility requirements for products and services.

For example, economic actors remain fully responsible for the compliance of AI to existing rules that protects consumers, any algorithmic exploitation of consumer behaviour in violation of existing rules shall be not permitted and violations shall be accordingly punished.

The Commission is of the opinion that the legislative framework could be improved to address the following risks and situations:

- *Effective application and enforcement of existing EU and national legislation*: the key characteristics of AI create challenges for ensuring the proper application and enforcement of EU and national legislation. The lack of transparency (opaqueness of AI) makes it difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation. Therefore, in order to ensure an effective application and enforcement, it may be necessary to adjust or clarify existing legislation in certain areas, for example on liability as further detailed in the Report, which accompanies this White Paper.

- *Limitations of scope of existing EU legislation*: an essential focus of EU product safety legislation is on the placing of products on the market. While in EU product safety legislation software, when is part of the final product, must comply with the relevant product safety rules, it is an open question whether stand-alone software is covered by EU product safety legislation, outside some sectors with explicit rules[45]. General EU safety legislation currently in force applies to products and not to services, and therefore in principle not to services based on AI technology either (e.g. health services, financial services, transport services).

- *Changing functionality of AI systems*: the integration of software, including AI, into products can modify the functioning of such products and systems during their lifecycle. This is particularly true for systems that require frequent software updates or which rely on machine learning. These features can give rise to new risks that were not present when the system was placed on the market. These risks are not adequately addressed in the existing legislation which predominantly focuses on safety risks present at the time of placing on the market.

- *Uncertainty as regards the allocation of responsibilities between different economic operators in the supply chain*: in general, EU legislation on product safety allocates the responsibility to the producer of the product placed on the market, including all components e.g. AI systems. But the rules can for example become unclear if AI is added after the product is placed on the market by a party that is not the producer. In addition, EU product liability legislation provides for liability of producers and leaves national liability rules to govern liability of others in the supply chain.

- *Changes to the concept of safety*: the use of AI in products and services can give rise to risks that EU legislation currently does not explicitly address. These risks may be linked to cyber threats, personal security risks (linked for example to new applications of AI such as to home appliances), risks that result from loss of connectivity, etc. These risks may be present at the time of placing products on the market or arise as a result of software updates or self-learning when the product is being used. The EU should make full use of the tools at its disposal to

[45] For instance software intended by the manufacturer to be used for medical purposes is considered a medical device under the Medical Device Regulation (Regulation (EU) 2017/745).

enhance its evidence base on potential risks linked to AI applications, including using the experience of the EU Cybersecurity Agency (ENISA) for assessing the AI threat landscape.

As indicated earlier, several Member States are already exploring options for national legislation to address the challenges created by AI. This raises the risk that the single market may be fragmented. Divergent national rules are likely to create obstacles for companies that want to sell and operate AI systems in the single market. Ensuring a common approach at EU level would enable European companies to benefit from smooth access to the single market and support their competitiveness on global markets.

Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics

The Report, which accompanies this White Paper, analyses the relevant legal framework. It identifies uncertainties as to the application of this framework with respect to the specific risks posed by AI systems and other digital technologies.

It concludes that the current product safety legislation already supports an extended concept of safety protecting against all kind of risks arising from the product according to its use. However, provisions explicitly covering new risks presented by the emerging digital technologies could be introduced to provide more legal certainty.

- The autonomous behaviour of certain AI systems during its life cycle may entail important product changes having an impact on safety, which may require a new risk assessment. In addition, human oversight from the product design and throughout the lifecycle of the AI products and systems may be needed as a safeguard.
- Explicit obligations for producers could be considered also in respect of mental safety risks of users when appropriate (ex. collaboration with humanoid robots).
- Union product safety legislation could provide for specific requirements addressing the risks to safety of faulty data at the design stage as well as mechanisms to ensure that quality of data is maintained throughout the use of the AI products and systems.
- The opacity of systems based on algorithms could be addressed through transparency requirements.
- Existing rules may need to be adapted and clarified in the case of a stand-alone software placed as it is on the market or downloaded into a product after its placing on the market, when having an impact on safety.
- Given the increasing complexity of supply chains as regards new technologies, provisions specifically requesting cooperation between the economic operators in the supply chain and the users could provide legal certainty.

The characteristics of emerging digital technologies like AI, the IoT and robotics may challenge aspects of the liability frameworks and could reduce their effectiveness. Some of these characteristics could make it hard to trace the damage back to a person, which would be necessary for a fault-based claim in accordance with most national rules. This could significantly increase the costs for victims and means that liability claims against others than producers may be difficult to make or prove.

- Persons having suffered harm caused with the involvement of AI systems need to enjoy the same level of protection as persons having suffered harm caused by other technologies, whilst technological innovation should be allowed to continue to develop.
- All options to ensure this objective should be carefully assessed, including possible amendments to the Product Liability Directive and possible further targeted harmonisation of national liability rules. For example, the Commission is seeking views whether and to what extent it may be needed to mitigate the consequences of complexity by adapting the burden of proof required by national liability rules for damage caused by the operation of AI applications.

From the discussion above, the Commission concludes that – in addition to the possible adjustments to existing legislation – a new legislation specifically on AI may be needed in order to make the EU legal framework fit for the current and anticipated technological and commercial developments.

C. SCOPE OF A FUTURE EU REGULATORY FRAMEWORK

A key issue for the future specific regulatory framework on AI intelligence is to determine the scope of its application. The working assumption is that the regulatory framework would apply to products and services relying on AI. AI should therefore be clearly defined for the purposes of this White Paper, as well as any possible future policy-making initiative.

In its Communication on AI for Europe the Commission provided a first definition of AI[46]. This definition was further refined by the High Level Expert Group[47].

In any new legal instrument, the definition of AI will need to be sufficiently flexible to accommodate technical progress while being precise enough to provide the necessary legal certainty.

For the purposes of this White Paper, as well as of any possible future discussions on policy initiatives, it seems important to clarify the main elements that compose AI, which are "data" and "algorithms". AI can be integrated in hardware. In case of machine learning techniques, which constitute a subset of AI, algorithms are trained to infer certain patterns based on a set of data in order to determine the actions needed to achieve a given goal. Algorithms may continue to learn when in use. While AI-based products can act autonomously by perceiving their environment and without following a pre-determined set of instructions, their behaviour is largely defined and constrained by its developers. Humans determine and programme the goals, which an AI system should optimise for.

> In autonomous driving for example, the algorithm uses, in real time, the data from the car (speed, engine consumption, shock-absorbers, etc..) and from the sensors scanning the whole environment of the car (road, signs, other vehicles, pedestrians etc..) to derive which direction, acceleration and speed the car should take to reach a certain destination. Based on the data observed, the algorithm adapts to the situation of the road and to the outside conditions, including other drivers' behaviour, to derive the most comfortable and safest drive.

The EU has a strict legal framework in place to ensure inter alia consumer protection, to address unfair commercial practices and to protect personal data and privacy. In addition, the acquis contains specific rules for certain sectors (e.g. healthcare, transport). These existing provisions of EU law will continue to apply in relation to AI, although certain updates to that framework may be necessary to reflect the digital transformation and the use of AI (see section B). As a consequence, those aspects that are

[46] COM(2018) 237 final, p. 1: "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.
AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)."

[47] High Level Expert Group, A definition of AI, p. 8: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions."

already addressed by existing horizontal or sectoral legislation (e.g. on medical devices[48], in transport systems) will continue to be governed by this legislation.

As a matter of principle, the new regulatory framework for AI should be effective to achieve its objectives while not being excessively prescriptive so that it could create a disproportionate burden, especially for SMEs. To strike this balance, the Commission is of the view that it should follow a risk-based approach.

A risk-based approach is important to help ensure that the regulatory intervention is proportionate. However, it requires clear criteria to differentiate between the different AI applications, in particular in relation to the question whether or not they are 'high-risk'[49]. The determination of what is a high-risk AI application should be clear and easily understandable and applicable for all parties concerned. Nevertheless even if an AI application is not qualified as high-risk, it remains entirely subject to already existing EU-rules.

The Commission is of the opinion that a given AI application should generally be considered high-risk in light of what is at stake, considering whether both the sector and the intended use involve significant risks, in particular from the viewpoint of protection of safety, consumer rights and fundamental rights. More specifically, an AI application should be considered high-risk where it meets the following two cumulative criteria:

- First, the AI application is employed in a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur. This first criterion ensures that the regulatory intervention is targeted on the areas where, generally speaking, risks are deemed most likely to occur. The sectors covered should be specifically and exhaustively listed in the new regulatory framework. For instance, healthcare; transport; energy and parts of the public sector.[50] The list should be periodically reviewed and amended where necessary in function of relevant developments in practice;

- Second, the AI application in the sector in question is, in addition, used in such a manner that significant risks are likely to arise. This second criterion reflects the acknowledgment that not every use of AI in the selected sectors necessarily involves significant risks. For example, whilst healthcare generally may well be a relevant sector, a flaw in the appointment scheduling system in a hospital will normally not pose risks of such significance as to justify legislative intervention. The assessment of the level of risk of a given use could be based on the impact on the affected parties. For instance, uses of AI applications that produce legal or similarly significant effects for the rights of an individual or a company; that pose risk of injury, death or significant material or immaterial damage; that produce effects that cannot reasonably be avoided by individuals or legal entities.

The application of the two cumulative criteria would ensure that the scope of the regulatory framework is targeted and provides legal certainty. The mandatory requirements contained in the new regulatory framework on AI (see section D below) would in principle apply only to those applications identified as high-risk in accordance with these two cumulative criteria.

---

[48] For example, there are different safety considerations and legal implications concerning AI systems that provide specialized medical information to physicians, AI systems providing medical information directly to the patient and AI systems performing medical tasks themselves directly on a patient. The Commission is examining these safety and liability challenges that are distinct to healthcare.

[49] EU legislation may categorise "risks" differently to what is described here, depending on the area, such as for example, product safety

[50] The public sector could include areas like asylum, migration, border controls and judiciary, social security and employment services.

Notwithstanding the foregoing, there may also be exceptional instances where, due to the risks at stake, the use of AI applications for certain purposes is to be considered as high-risk as such – that is, irrespective of the sector concerned and where the below requirements would still apply.[51] As an illustration, one could think in particular of the following:

- In light of its significance for individuals and of the EU acquis addressing employment equality, the use of AI applications for recruitment processes as well as in situations impacting workers' rights would always be considered "high-risk" and therefore the below requirements would at all times apply. Further specific applications affecting consumer rights could be considered.

- the use of AI applications for the purposes of remote biometric identification[52] and other intrusive surveillance technologies, would always be considered "high-risk" and therefore the below requirements would at all times apply.

D. TYPES OF REQUIREMENTS

When designing the future regulatory framework for AI, it will be necessary to decide on the types of mandatory legal requirements to be imposed on the relevant actors. These requirements may be further specified through standards. As noted in section C above and in addition to already existing legislation, those requirements would apply to high-risk AI applications only, thus ensuring that any regulatory intervention is focused and proportionate.

Taking into account the guidelines of the High Level Expert Group and what has been set out in the foregoing, the requirements for high-risk AI applications could consist of the following key features, which are discussed in further detail in the subsections below:

- training data;
- data and record-keeping;
- information to be provided;
- robustness and accuracy;
- human oversight;
- specific requirements for certain particular AI applications, such as those used for purposes of remote biometric identification.

To ensure legal certainty, these requirements will be further specified to provide a clear benchmark for all the actors who need to comply with them.

a) Training data

It is more important than ever to promote, strengthen and defend the EU's values and rules, and in particular the rights that citizens derive from EU law. These efforts undoubtedly also extend to the high-risk AI applications marketed and used in the EU under consideration here.

[51] It is important to highlight that other pieces of EU legislation may also apply. For example, when incorporated into a consumer product, the General Product Safety Directive may apply to the safety of AI applications.

[52] Remote biometric identification should be distinguished from biometric authentication (the latter is a security process that relies on the unique biological characteristics of an individual to verify that he/she is who he/she says he/she is). Remote biometric identification is when the identities of multiple persons are established with the help of biometric identifiers (fingerprints, facial image, iris, vascular patterns, etc.) at a distance, in a public space and in a continuous or ongoing manner by checking them against data stored in a database.

As discussed earlier, without data, there is no AI. The functioning of many AI systems, and the actions and decisions to which they may lead, very much depend on the data set on which the systems have been trained. The necessary measures should therefore be taken to ensure that, where it comes to the data used to train AI systems, the EU's values and rules are respected, specifically in relation to safety and existing legislative rules for the protection of fundamental rights. The following requirements relating to the data set used to train AI systems could be envisaged:

- Requirements aimed at providing reasonable assurances that the subsequent use of the products or services that the AI system enables is safe, in that it meets the standards set in the applicable EU safety rules (existing as well as possible complementary ones). For instance, requirements ensuring that AI systems are trained on data sets that are sufficiently broad and cover all relevant scenarios needed to avoid dangerous situations.

- Requirements to take reasonable measures aimed at ensuring that such subsequent use of AI systems does not lead to outcomes entailing prohibited discrimination. These requirements could entail in particular obligations to use data sets that are sufficiently representative, especially to ensure that all relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination are appropriately reflected in those data sets;

- Requirements aimed at ensuring that privacy and personal data are adequately protected during the use of AI-enabled products and services. For issues falling within their respective scope, the General Data Protection Regulation and the Law Enforcement Directive regulate these matters.

b) Keeping of records and data

Taking into account elements such as the complexity and opacity of many AI systems and the related difficulties that may exist to effectively verify compliance with and enforce the applicable rules, requirements are called for regarding the keeping of records in relation to the programming of the algorithm, the data used to train high-risk AI systems, and, in certain cases, the keeping of the data themselves. These requirements essentially allow potentially problematic actions or decisions by AI systems to be traced back and verified. This should not only facilitate supervision and enforcement; it may also increase the incentives for the economic operators concerned to take account at an early stage of the need to respect those rules.

To this aim, the regulatory framework could prescribe that the following should be kept:

- accurate records regarding the data set used to train and test the AI systems, including a description of the main characteristics and how the data set was selected;

- in certain justified cases, the data sets themselves;

- documentation on the programming[53] and training methodologies, processes and techniques used to build, test and validate the AI systems, including where relevant in respect of safety and avoiding bias that could lead to prohibited discrimination.

The records, documentation and, where relevant, data sets would need to be retained during a limited, reasonable time period to ensure effective enforcement of the relevant legislation. Measures should be

[53] For instance, documentation on the algorithm including what the model shall optimise for, which weights are designed to certain parameters at the outset etc.

taken to ensure that they are made available upon request, in particular for testing or inspection by competent authorities. Where necessary, arrangements should be made to ensure that confidential information, such as trade secrets, is protected.

c) Information provision

Transparency is required also beyond the record-keeping requirements discussed in point c) above. In order to achieve the objectives pursued – in particular promoting the responsible use of AI, building trust and facilitating redress where needed – it is important that adequate information is provided in a proactive manner about the use of high-risk AI systems.

Accordingly, the following requirements could be considered:

- Ensuring clear information to be provided as to the AI system's capabilities and limitations, in particular the purpose for which the systems are intended, the conditions under which they can be expected to function as intended and the expected level of accuracy in achieving the specified purpose. This information is important especially for deployers of the systems, but it may also be relevant to competent authorities and affected parties.

- Separately, citizens should be clearly informed when they are interacting with an AI system and not a human being. Whilst EU data protection legislation already contain certain rules of this kind[54], additional requirements may be called for to achieve the abovementioned objectives. If so, unnecessary burdens should be avoided. Therefore, no such information needs to be provided, for instance, in situations where it is immediately obvious to citizens that they are interacting with AI systems. It is furthermore important that the information provided is objective, concise and easily understandable. The manner in which the information is to be provided should be tailored to the particular context.

d) Robustness and accuracy

AI systems – and certainly high-risk AI applications – must be technically robust and accurate in order to be trustworthy. That means that such systems need to be developed in a responsible manner and with an ex-ante due and proper consideration of the risks that they may generate. Their development and functioning must be such to ensure that AI systems behave reliably as intended. All reasonable measures should be taken to minimise the risk of harm being caused.

Accordingly, the following elements could be considered:

- Requirements ensuring that the AI systems are robust and accurate, or at least correctly reflect their level of accuracy, during all life cycle phases;

- Requirements ensuring that outcomes are reproducible;

- Requirements ensuring that AI systems can adequately deal with errors or inconsistencies during all life cycle phases.

[54] In particular, pursuant to Art. 13(2)(f) GDPR, controllers must, at the time when the personal data are obtained, provide the data subjects with further information necessary to ensure fair and transparent processing about the existence of automated decision-making and certain additional information.

- Requirements ensuring that AI systems are resilient against both overt attacks and more subtle attempts to manipulate data or algorithms themselves, and that mitigating measures are taken in such cases.

e) Human oversight

Human oversight helps ensuring that an AI system does not undermine human autonomy or cause other adverse effects. The objective of trustworthy, ethical and human-centric AI can only be achieved by ensuring an appropriate involvement by human beings in relation to high-risk AI applications.

Even though the AI applications considered in this White paper for a specific legal regime are all considered high-risk, the appropriate type and degree of human oversight may vary from one case to another. It shall depend in particular on the intended use of the systems and the effects that the use could have for affected citizens and legal entities. It shall also be without prejudice to the legal rights established by the GDPR when the AI system processes personal data. For instance, human oversight could have the following, non-exhaustive, manifestations:

- the output of the AI system does not become effective unless it has been previously reviewed and validated by a human (e.g. the rejection of an application for social security benefits may be taken by a human only);

- the output of the AI system becomes immediately effective, but human intervention is ensured afterwards (e.g. the rejection of an application for a credit card may be processed by an AI system, but human review must be possible afterwards);

- monitoring of the AI system while in operation and the ability to intervene in real time and deactivate (e.g. a stop button or procedure is available in a driverless car when a human determines that car operation is not safe);

- in the design phase, by imposing operational constraints on the AI system (e.g. a driverless car shall stop operating in certain conditions of low visibility when sensors may become less reliable or shall maintain a certain distance in any given condition from the preceding vehicle).

f) Specific requirements for remote biometric identification

The gathering and use of biometric data[55] for remote identification[56] purposes, for instance through deployment of facial recognition in public places, carries specific risks for fundamental rights[57]. The

[55] Biometric data is defined as "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique authentification or identification of that natural person, such as facial images or dactyloscopic [fingerprint] data." (Law Enforcement Directive, Art. 3 (13); GDPR, Art. 4 (14); Regulation (EU) 2018/1725, Art. 3 (18).

[56] In connection to facial recognition, identification means that the template of a person's facial image is compared to many other templates stored in a database to find out if his or her image is stored there. Authentication (or verification) on the other hand is often referred to as one-to-one matching. It enables the comparison of two biometric templates, usually assumed to belong to the same individual. Two biometric templates are compared to determine if the person shown on the two images is the same person. Such a procedure is, for example, used at Automated Border Control (ABC) gates used for border checks at airports.

[57] For example on people's dignity. Relatedly, the rights to respect for private life and protection of personal data are at the core of fundamental rights concerns when using facial recognition technology. There is also a potential impact on non-discrimination and rights of special groups, such as children, older persons and persons with disabilities. Moreover, freedom of expression, association and assembly must not be undermined by the use of the technology. See: Facial recognition technology: fundamental rights considerations in the context of law enforcement, https://fra.europa.eu/en/publication/2019/facial-recognition.

fundamental rights implications of using remote biometric identification AI systems can vary considerably depending on the purpose, context and scope of the use.

EU data protection rules prohibit in principle the processing of biometric data for the purpose of uniquely identifying a natural person, except under specific conditions[58]. Specifically, under the GDPR, such processing can only take place on a limited number of grounds, the main one being for reasons of substantial public interest. In that case, the processing must take place on the basis of EU or national law, subject to the requirements of proportionality, respect for the essence of the right to data protection and appropriate safeguards. Under the Law Enforcement Directive, there must be a strict necessity for such processing, in principle an authorisation by EU or national law as well as appropriate safeguards. As any processing of biometric data for the purpose of uniquely identifying a natural person would relate to an exception to a prohibition laid down in EU law, it would be subject to the Charter of Fundamental Rights of the EU.

It follows that, in accordance with the current EU data protection rules and the Charter of Fundamental Rights, AI can only be used for remote biometric identification purposes where such use is duly justified, proportionate and subject to adequate safeguards.

In order to address possible societal concerns relating to the use of AI for such purposes in public places, and to avoid fragmentation in the internal market, the Commission will launch a broad European debate on the specific circumstances, if any, which might justify such use, and on common safeguards.

E. ADDRESSEES

In relation to the addressees of the legal requirements that would apply in relation to the high-risk AI applications referred to above, there are two main issues to be considered.

First, there is the question how obligations are to be distributed among the economic operators involved. Many actors are involved in the lifecycle of an AI system. These include the developer, the deployer (the person who uses an AI-equipped product or service) and potentially others (producer, distributor or importer, service provider, professional or private user).

It is the Commission's view that, in a future regulatory framework, each obligation should be addressed to the actor(s) who is (are) best placed to address any potential risks. For example, while the developers of AI may be best placed to address risks arising from the development phase, their ability to control risks during the use phase may be more limited. In that case, the deployer should be subject to the relevant obligation. This is without prejudice to the question whether, for the purpose of liability to end-users or other parties suffering harm and ensuring effective access to justice, which party should be liable for any damage caused. Under EU product liability law, liability for defective products is attributed to the producer, without prejudice to national laws which may also allow recovery from other parties.

Second, there is the question about the geographic scope of the legislative intervention. In the view of the Commission, it is paramount that the requirements are applicable to all relevant economic operators providing AI-enabled products or services in the EU, regardless of whether they are established in the EU or not. Otherwise, the objectives of the legislative intervention, mentioned earlier, could not fully be achieved.

[58] Article 9 GDPR, Article 10 Law Enforcement Directive. See also Article 10 Regulation (EU) 2018/1725 (applicable to the EU institutions and bodies).

F. COMPLIANCE AND ENFORCEMENT

In order to ensure that AI is trustworthy, secure and in respect of European values and rules, the applicable legal requirements need to be complied with in practice and be effectively enforced both by competent national and European authorities and by affected parties. Competent authorities should be in a position to investigate individual cases, but also to assess the impact on society.

In view of the high risk that certain AI applications pose for citizens and our society (see section A above), the Commission considers at this stage that an objective, prior conformity assessment would be necessary to verify and ensure that certain of the above mentioned mandatory requirements applicable to high-risk applications (see section D above) are complied with. The prior conformity assessment could include procedures for testing, inspection or certification[59]. It could include checks of the algorithms and of the data sets used in the development phase.

The conformity assessments for high-risk AI applications should be part of the conformity assessment mechanisms that already exist for a large number of products being placed on the EU's internal market. Where no such existing mechanisms can be relied on, similar mechanisms may need to be established, drawing on best practice and possible input of stakeholders and European standards organisations. Any such new mechanism should be proportionate and non-discriminatory and use transparent and objective criteria in compliance with international obligations.

When designing and implementing a system relying on prior conformity assessments, particular account should be taken of the following:

- Not all requirements outlined above may be suitable to be verified through a prior conformity assessment. For instance, the requirement about information to be provided generally does not lend itself well for verification through such an assessment.

- Particular account should be taken of the possibility that certain AI systems evolve and learn from experience, which may require repeated assessments over the life-time of the AI systems in question.

- The need to verify the data used for training and the relevant programming and training methodologies, processes and techniques used to build, test and validate AI systems.

- In case the conformity assessment shows that an AI system does not meet the requirements for example relating to the data used to train it, the identified shortcomings will need to be remedied, for instance by re-training the system in the EU in such a way as to ensure that all applicable requirements are met.

The conformity assessments would be mandatory for all economic operators addressed by the requirements, regardless of their place of establishment[60]. In order to limit the burden on SMEs, some support structure might be envisaged including through the Digital Innovation Hubs. In addition, standards as well as dedicated online tools could facilitate compliance.

[59] The system would be based on conformity assessment procedures in the EU, see Decision 768/2008/EC or on Regulation (EU) 2019/881 (Cybersecurity Act), taking into account the specificities of AI. See the Blue Guide on the Implementation of EU product rules, 2014.

[60] As regards the relevant governance structure, including the bodies designated to carry out the conformity assessments, see section H below.

Any prior conformity assessment should be without prejudice to monitoring compliance and ex post enforcement by competent national authorities. That holds true in respect of high-risk AI applications, but also in respect of other AI applications subject to legal requirements, although the high-risk nature of the applications at issue may be reason for the competent national authorities to give particular attention to the former. Ex-post controls should be enabled by adequate documentation of the relevant AI application (see section E above) and, where appropriate, a possibility for third parties such as competent authorities to test such applications. This may be especially important where risks to fundamental rights arise, which are context dependent. Such monitoring of compliance should be part of a continuous market surveillance scheme. Governance-related aspects are further discussed in section H below.

Moreover, both for high- risk AI applications and for other AI applications, effective judicial redress for parties negatively affected by AI systems should be ensured. Issues related to liability are further discussed in the Report on the safety and liability framework accompanying this White Paper.

G. VOLUNTARY LABELLING FOR NO-HIGH RISK AI APPLICATIONS

For AI applications that do not qualify as 'high-risk' (see section C above) and that are therefore not subject to the mandatory requirements discussed above (see sections D, E and F above), an option would be, in addition to applicable legislation, to establish a voluntary labelling scheme.

Under the scheme, interested economic operators that are not covered by the mandatory requirements could decide to make themselves subject, on a voluntary basis, either to those requirements or to a specific set of similar requirements especially established for the purposes of the voluntary scheme. The economic operators concerned would then be awarded a quality label for their AI applications.

The voluntary label would allow the economic operators concerned to signal that their AI-enabled products and services are trustworthy. It would allow users to easily recognise that the products and services in question are in compliance with certain objective and standardised EU-wide benchmarks, going beyond the normally applicable legal obligations. This would help enhance the trust of users in AI systems and promote the overall uptake of the technology.

This option would entail the creation of a new legal instrument that sets out the voluntary labelling framework for developers and/or deployers of AI systems that are not be considered as high-risk. While participation in the labelling scheme would be voluntary, once the developer or the deployer opted to use the label, the requirements would be binding. The combination of *ex ante* and *ex post* enforcement would need to ensure that all requirements are complied with.

H. GOVERNANCE

A European governance structure on AI in the form of a framework for cooperation of national competent authorities is necessary to avoid fragmentation of responsibilities, increase capacity in Member States, and make sure that Europe equips itself progressively with the capacity needed for testing and certification of AI-enabled products and services. In this context, it would be beneficial to support competent national authorities to enable them to fulfil their mandate where AI is used.

A European governance structure could have a variety of tasks, as a forum for a regular exchange of information and best practice, identifying emerging trends, advising on standardisation activity as well as on certification. It should also play a key role in facilitating the implementation of the legal framework, such as through issuing guidance, opinions and expertise. To that effect, it should rely on a network of national authorities, as well as sectorial networks and regulatory authorities, at national and EU level. Moreover, a committee of experts could provide assistance to the Commission.

The governance structure should guarantee maximum stakeholders participation. Stakeholders – consumer organisation and social partners, businesses, researchers, and civil society organisations – should be consulted on the implementation and the further development of the framework.

Given already existing structures such as in finance, pharmaceuticals, aviation, medical devices, consumer protection, data protection, the proposed governance structure should not duplicate existing functions. It should instead establish close links with other EU and national competent authorities in the various sectors to complement existing expertise and help existing authorities in monitoring and the oversight of the activities of economic operators involving AI systems and AI-enabled products and services.

Finally, if this option is pursued, the carrying out of conformity assessments could be entrusted to notified bodies designated by Member States. Testing centres should enable the independent audit and assessment of AI-systems in accordance with the requirements outlined above. Independent assessment will increase trust and ensures objectivity. It could also facilitate the work of relevant competent authorities.

The EU enjoys excellent testing and assessment centres and should develop its capacity also in the area of AI. Economic operators established in third countries wanting to enter the internal market could either make use of designated bodies established in the EU or, subject to mutual recognition agreements with third countries, have recourse to third-country bodies designated to carry out such assessment.

The governance structure relating to AI and the possible conformity assessments at issue here would leave the powers and responsibilities under existing EU law of the relevant competent authorities in specific sectors or on specific issues (finance, pharmaceuticals, aviation, medical devices, consumer protection, data protection, etc.) unaffected.

6. CONCLUSION

AI is a strategic technology that offers many benefits for citizens, companies and society as a whole, provided it is human-centric, ethical, sustainable and respects fundamental rights and values. AI offers important efficiency and productivity gains that can strengthen the competitiveness of European industry and improve the wellbeing of citizens. It can also contribute to finding solutions to some of the most pressing societal challenges, including the fight against climate change and environmental degradation, the challenges linked to sustainability and demographic changes, and the protection of our democracies and, where necessary and proportionate, the fight against crime.

For Europe to seize fully the opportunities that AI offers, it must develop and reinforce the necessary industrial and technological capacities. As set out in the accompanying European strategy for data, this also requires measures that will enable the EU to become a global hub for data.

The European approach for AI aims to promote Europe's innovation capacity in the area of AI while supporting the development and uptake of ethical and trustworthy AI across the EU economy. AI should work for people and be a force for good in society.

With this White Paper and the accompanying Report on the safety and liability framework, the Commission launches a broad consultation of Member States civil society, industry and academics, of concrete proposals for a European approach to AI. These include both policy means to boost investments in research and innovation, enhance the development of skills and support the uptake of AI by SMEs, and proposals for key elements of a future regulatory framework. This consultation will

allow a comprehensive dialogue with all concerned parties that will inform the next steps of the Commission.

The Commission invites for comments on the proposals set out in the White Paper through an open public consultation available at https://ec.europa.eu/info/consultations_en. The consultation is open for comments until 19 May 2020.

It is standard practice for the Commission to publish submissions received in response to a public consultation. However, it is possible to request that submissions, or parts thereof, remain confidential. Should this be the case, please indicate clearly on the front page of your submission that it should not be made public and also send a non-confidential version of your submission to the Commission for publication.