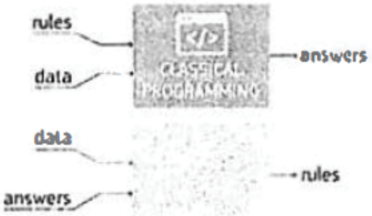
Structure for the White Paper on artificial intelligence – a European approach

1. INTRODUCTION

- The development of artificial intelligence (AI) will have profound impacts on our societies. The purpose of this White Paper is to put forward proposals to develop a European approach to artificial intelligence, which will help prepare our societies for the challenges and opportunities that artificial intelligence is creating. These proposals cover the main building blocks of a European approach, including actions to support the development and uptake of artificial intelligence, actions to facilitate access to data and the key pillars of a future regulatory framework for artificial intelligence. With this White Paper, the Commission launches a broad consultation process and invites all relevant stakeholders to comment on the proposals for this European approach.



- There is currently no consensus at international level on the definition of the term "artificial intelligence". The term is used to describe a variety of technologies with certain common features. The High-Level Expert Group on Artificial Intelligence set up by the Commission described artificial intelligence systems as "*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 (...)*"[1].

- The objective of the European approach to artificial intelligence is to promote the development and uptake of artificial intelligence across Europe, while ensuring that the technology is developed and used in a way that respects European values and principles. Given that other major economies, in particular the US and China, are supporting artificial intelligence, it is essential to ensure that European citizens and companies can both benefit from the technology and shape the way it develops.

- Beyond productivity and efficiency gains, artificial intelligence promises to enable humans to develop analytical capacities not yet reached, opening the way to new discoveries and helping to solve some of the world's biggest challenges: from treating chronic diseases or reducing fatality rates in traffic accidents to fighting climate change or anticipating cybersecurity threats. However, Europe can only seize these opportunities if it can strengthen its leadership in developing artificial intelligence applications and increase the uptake of this technology.

- Artificial intelligence has a transformational potential for the industry. It helps make products, services and processes more efficient. It can also help factories to stay in or return to Europe. It improves products, services, processes and business models in all economic sectors. It can help companies identify, which machines will need maintenance before they break down. It is

[1] For further detail, please see the 8 April 2019 opinion of the High-Level Expert Group on artificial intelligence.

at our fingertips when we translate texts online. It is making life easier for the visually impaired by assisting them in perceiving objects in their daily lives. At home, a smart thermostat can reduce energy bills by up to 25%, by analysing the habits of the people who live in the house and adjusting the temperature accordingly. Artificial intelligence also has the potential to improve the delivery of public services by making them more efficient and accessible and to better allocate scarce resources and budgets

- Artificial intelligence is a strategic technology that can bring tremendous opportunities. At the same time, it has distinct characteristics that raise specific challenges in terms of governance, and in relation to the safety and liability of devices and systems equipped with it. These characteristics include autonomy (e.g. performing tasks in complex environments without constant guidance), opacity ('black-box-effect') and the ability to improve performance by learning from experience. While the promise of artificial intelligence systems is that they will spot patterns in the data and will make decisions faster than humans do, the risk is that they may make inappropriate decisions, and that the reasoning behind those decisions may not be known. This raises concerns related to liability, discrimination and transparency, which should be addressed in a regulatory framework.

- This White Paper is structured as follows:
  - Section 2 describes the existing policy framework for artificial intelligence at the EU level and beyond.
  - Section 3 outlines in more detail the policy actions in support of the development and uptake artificial intelligence across Europe, including on investment, skills, and small and medium-sized enterprises.
  - Section 4 sets out ideas on how best to facilitate access to data, which is a prerequisite for developing the vast majority of todays' artificial intelligence systems.
  - Section 5 constitutes the main part of the White Paper, setting out the key elements of a future comprehensive European legislative framework for artificial intelligence, which respects European values and principles.
  - Section 6 contains the conclusions setting out the Commission's intention of the next steps and the relevant timeline for receiving the contributions of stakeholders.
- The White Paper is accompanied by three other documents:
  - the Report on the broader implications of artificial intelligence, Internet of Things and other digital technologies for the EU safety and liability framework;
  - [the proposal for a new Council Regulation on high performance computing;] and
  - the review of the Coordinated Plan on Artificial Intelligence (COM(2020)xxx).

2. EXISTING POLICY FRAMEWORK

a. EU framework

- This White Paper builds on the existing policy framework, including the Communication on 'Artificial Intelligence for Europe' and the Coordinated Plan on Artificial Intelligence developed together with Member States. The Communication focuses on three key pillars of the European artificial intelligence strategy: support for the EU's technological and industrial capacity and the uptake of artificial intelligence across the economy, preparing for the socio-economic changes brought about by artificial intelligence, and ensuring an appropriate ethical and legal framework. It also launched the work of the European AI Alliance as a forum to bring together a broad range of stakeholders, as well as the High-Level Expert Group on Artificial Intelligence, and the Expert Group on Liability and New Technologies.
- With the subsequent Communication of April 2019 the Commission welcomes the Ethics Guidelines for Trustworthy Artificial Intelligence prepared by the High-Level Expert Group. The guidelines list seven requirements that artificial intelligence systems should meet in order to be trustworthy: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non discrimination and fairness; societal and environmental well-being; and accountability. This approach also includes tools to help put translating the ethical principles into practice. Industry and other stakeholders have recently tested these tools.

b. International aspects

- The EU's work on artificial intelligence has influenced international discussions. When developing its ethical guidelines, the High-Level Expert Group involved a number of non-EU organisations (from US and Canada) and as several governmental observers (from Japan and Singapore). In parallel the EU was closely involved in developing the Organisation for Economic Co-operation and Development's ethical principles for artificial intelligence, which were subsequently endorsed by the G20.
- Given that China and the US remain the most important global players in artificial intelligence, the EU seeks to cooperate with them based on a strategic approach that protects the EU's interests (e.g. mainstreaming European standards, accessing key resources including data, creating a level playing field). The Commission is convinced that international cooperation must be based on a like-minded approach to the EU's fundamental values, such as the respect for human dignity, pluralism, non-discrimination and protection of privacy.[2]

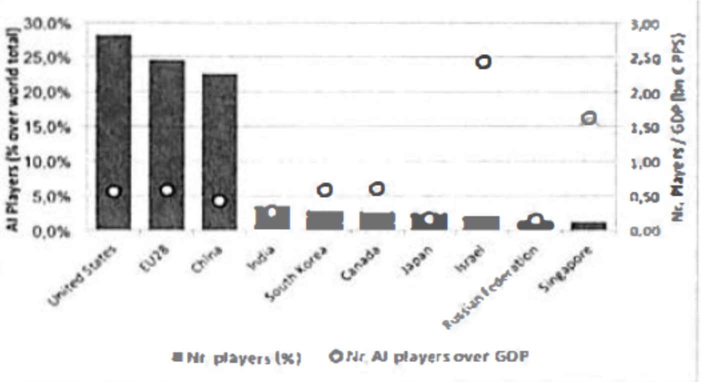
[2] Under the Partnership Instrument, the Commission will finance a €2.5 million project that will facilitate cooperation with like-minded partners, in order to promote the EU artificial intelligence ethical guidelines and to adopt common principles and operational conclusions.

3. SUPPORTING THE DEVELOPMENT AND UPTAKE OF ARTIFICIAL INTELLIGENCE

- Europe is well-placed to benefit from the potential of artificial intelligence, not only as a user but also as a producer of artificial intelligence. It has excellent research centres, which publish more scientific articles related to artificial intelligence than any other region in the world, a world-leading position in robotics, business-to-business markets as well as competitive manufacturing and services sectors, from automotive to healthcare, from energy to financial to agriculture. It holds large amounts of public and industrial data and has well-recognised technology and industrial strengths in low power consumption, and safe and secure digital systems that are essential for the further development of artificial intelligence.

- One reason for Europe's strong position in terms of research is the EU funding programme which has proven instrumental in federating efforts, avoiding duplications, and leveraging public and private investments in the Member States. Over the past two years, the EU funding for activities related to artificial intelligence has gone up by €1.5 billion, i.e. an increase of 70% relative to the previous period.

- However, investment in research and innovation in Europe is still a fraction of the public and private investments in other regions of the world. Some €3.2 billion were invested in artificial intelligence in Europe in 2016, compared to around €12.1 billion in North America and €6.5 billion in Asia. To respond to the challenge, Europe needs to increase significantly investment levels. The Coordinated Plan on Artificial Intelligence developed with Member States, Norway and Switzerland has proven invaluable in building stronger cooperation on artificial intelligence in Europe and in creating synergies for maximising investments into the artificial intelligence value chain.



AI players over GDP; Source: JRC

- Europe should leverage its strengths to expand its position on markets along the value chain, from hardware manufacturing through software all the way to services. This is already happening to an extent: Europe produces more than a quarter of industrial and professional service robots (e.g. for precision farming, security, health, logistics), and plays an important role in the development and exploitation of platforms providing services to companies and organisations (business-to-business), applications to progress towards the "intelligent enterprise" and e-government.

- Europe has a weak position in consumer applications and online platforms which results in a competitive disadvantage in data access. However, there are also opportunities. Whereas around 80% of the current 40 zetabytes of data is stored in data centres, many of which are controlled by non-European operators, the advent of the Internet of Things and edge computing could result in a radical change in the distribution of data. As a result, 80% of the 175 zetabytes of data that is expected to be available in 2025 should be stored locally at the edge of networks in factories, hospitals, etc.

- Similarly, Europe should expand into the area of specialised processors and augment its computing capabilities. Currently, this market is dominated by third countries but this could change with the help of the European Processor Initiative, which addresses the development of low-power computing systems for both edge and next generation high performance computing. Moreover, Europe is leading in neuromorphic solutions that are ideally suited to automation of industrial processes (industry 4.0) and transport modes. They offer improvements of several orders of magnitude in energy efficiency; availability of testing and experimentation facilities will greatly help the application of neuromorphic solutions.

- In parallel, Europe will continue to lead progress in the algorithmic foundations of AI, building on its own scientific excellence. This will require building bridges between existing silos, such as machine learning and symbolic approaches, where Europe is historically very strong. Such efforts will support Europe's technological sovereignty in the long term.

- EU-level funding can ensure cross-fertilisation of European developments in artificial intelligence and federate investments in areas where this will make a difference and the efforts required go beyond what any single Member State can achieve. Key proposals to address the above-mentioned issues include:

    ✓ *Establishing a world-leading artificial intelligence computing and data infrastructure in Europe*: a comprehensive data and computing infrastructure using as a basis High Performance Computing (HPC) centres and edge computing capacities, through the EuroHPC Joint Undertaking. The deployment of a next generation high performance computing infrastructure will be complemented by a European federation of interoperable, flexible and scalable cloud and computing infrastructures and targeted cloud-based artificial intelligence services (to be funded through Horizon Europe and the Digital Europe Programme). Supporting the deployment of common European data spaces to facilitate pooling and sharing of data from across Europe is also crucial and will be addressed in a specific initiative.

    ✓ *Federating knowledge and achieving excellence*: drawing on the Commission's long-term efforts to strengthen the European scientific community and make Europe the "place to be", we will reinforce European excellence centres for artificial intelligence and facilitate their collaboration and networking through strengthened coordination. Currently, there is no single institution which can be recognised as a leader by the entire community in all the four major sub-disciplines where Europe can lead, i.e.: foundational research in artificial intelligence algorithms, perception and interaction, robotics, and next generation of chips for artificial intelligence. As a first step, the Commission helped foster consolidation in the individual sub-branches of artificial intelligence, addressing the fragmentation in these fields. In a second step, the Commission will strengthen coordination of the locally distributed networks. Moreover, it aims at establishing networks of leading universities to attract the best professors and scientists and offer world-leading master programmes in artificial intelligence (to be funded through Horizon Europe and the Digital Europe Programme).

- ✓ *Supporting research and innovation to stay at the forefront and create new markets*: a 'Leaders Group' will be set up with C-level representatives of major stakeholders, to develop an industrial strategy and commit to its implementation. The Leader's Group will also offer strategic guidance to a new public-private partnership on artificial intelligence, data and robotics involving all relevant stakeholders (funded through Horizon Europe), strengthening cooperation between academia and industry. Funding of large-scale testing facilities, including for neuromorphic, under the Digital Europe Programme will help bringing innovation closer to the market.

- ✓ *Fostering the uptake of artificial intelligence*: Improving the uptake of artificial intelligence is a key task of the Digital Innovation Hubs. These Hubs will be strengthened and supported through the Digital Europe Programme which will also support the uptake in high-impact application sectors such as healthcare (e.g. artificial intelligence for health imaging, genomics, testing medicines and medical devices), mobility (cross-border corridors for connected and automated mobility) and environmental modelling and monitoring (e.g. a highly accurate prediction and crisis management capacity). Additionally, the artificial intelligence-on-demand platform should become a reference point for knowledge related to artificial intelligence, algorithms, tools, infrastructure, equipment, and data resources.

- ✓ *Ensuring access to finance for artificial intelligence innovators*: a pilot scheme will be launched under InnovFin to provide equity financing for artificial intelligence and blockchain innovative developments and will be scaled up through InvestEU in 2021.

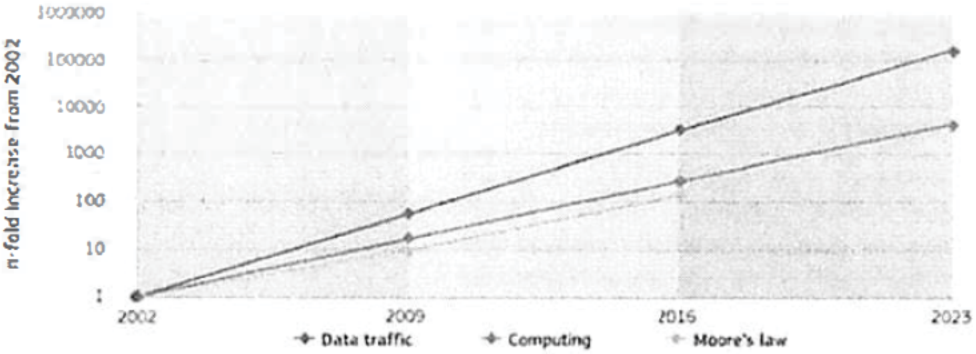
4. FACILITATING ACCESS TO DATA

- Ensuring access to data for EU businesses and the public sector is a prerequisite for developing artificial intelligence. This emerges from national artificial intelligence strategies developed across the EU. Data is an important driver of innovation, and creates new opportunities for growth, including for small and medium-sized enterprises. The optimal use of data can help us live healthier and longer lives that are more environmentally friendly.

- The EU can build on its comprehensive legal framework for data and its use in the economy, including the General Data Protection Regulation, the Regulation on the Free Flow of Data, and the Open Data Directive. The Annex to this White Paper gives an overview of the existing legislation on data access and use in the EU and an assessment of its relevance for the development of artificial intelligence.

- The Commission sees the development of common European data spaces to be a key measure for redressing the problem of data access. These spaces will combine the technical infrastructure for data sharing with governance mechanisms. They will be organised by sector (for example agriculture) or problem area (for example climate change).

- This White Paper presents a series of further measures to ensure data availability in the common European data spaces. On some of these actions, work has already started. Others are to be addressed in the near future. In a separate policy document, the Commission will present its overall strategy on data, including additional measures related to data access and use that require further analysis and discussion.



Projection of data and computing growths (logarithmic scale). Source: JRC based on Kambatla et al., 2014

- Based on the recently revised Open Data Directive, the Commission intends to adopt by early 2021 an implementing act on high-value public sector datasets. These datasets should be made available for free and in machine-readable format, well suited for artificial intelligence development. This concerns geospatial data, environmental and earth observation data, meteorological, mobility and business data, and statistics. Further categories could be added by way of a delegated act.

- A clear link needs to be made between the data policies and the EU-level investments (please see Section 3). In particular, the Commission wants to support the development of common European data spaces under the Digital Europe programme. This includes also support to national agencies for publishing high-value datasets.

<u>Key questions to be further addressed:</u>

- ➢ *What are the main issues concerning data used for training artificial intelligence? Quality of data? Biased data? Interoperability? Access to existing data?*

- ➢ *Are there any existing initiatives in the private sector to improve access to and sharing of data for the purpose of training artificial intelligence? If so, can we help scale them up? If not, why do they not exist?*

- ➢ *What is the existing policy framework to facilitate access to and use of data?*

- ➢ *Which problems could be better solved at national level and which at the EU level?*

5. A REGULATORY FRAMEWORK FOR ARTIFICIAL INTELLIGENCE

- The regulatory framework for artificial intelligence has to be consistent with the overall objectives of the European approach to artificial intelligence, i.e. to promote Europe's innovation capacity in this new and promising field, while simultaneously ensuring that this technology is developed and used in a way that respects European values and principles.

- As any new technology, the use of artificial intelligence brings both new opportunities and new risks. In addition, artificial intelligence poses distinctive challenges from a regulatory point of view, as products and services based on artificial intelligence combine data dependency (data generation, processing and analysis) with almost omnipresent connectivity within new technological ecosystems, such as the Internet of Things and cloud computing.

- Artificial intelligence is already subject to an extensive body of EU legislation, on fundamental rights (e.g. data protection, non discrimination, gender equality, asylum, copyright), consumer law, and product safety and liability. However, given the fast development of the artificial intelligence technology, this legislation might not fully cover all of the specific risks that artificial intelligence brings, possibly revealing certain regulatory gaps or weaknesses that were not apparent before. This also includes a lack of effective regulatory tools to ensure that artificial intelligence complies with existing requirements.

- A balanced and values-based regulatory framework will not only support the widespread adoption of this technology, but will also help European companies to benefit fully from a friction-less single market to scale up their operations across Europe. It must carefully complement and build upon the existing EU and national legal frameworks, to provide policy continuity and ensure legal certainty Such a proportionate approach focused on addressing well-defined risks and gaps will help to avoid unnecessary additional regulatory and administrative burdens, and ensure that European innovation continues to thrive.

A. PROBLEM DEFINITION

- In spite of the opportunities that artificial intelligence can provide, it can also lead to harm. A potential harm brought by artificial intelligence might be both material (loss of life, safety and health of individuals, damage to property, etc.) and immaterial (loss of privacy, limitations to the right of freedom of expression, human dignity, etc.), and can relate to a wide array of risks.

i. Risks for fundamental rights, including discrimination, privacy and data protection

- Bias and discrimination are inherent elements of any societal or economic activity. Human decision-making is also prone to mistakes and biases. However, the same level of bias when present in an artificial intelligence could affect and discriminate many people without the social control mechanisms that govern human behaviour. In addition to discrimination, artificial intelligence may lead to breaches of other fundamental rights[3], including freedom of expression, freedom of assembly, human dignity, private life or right to fair trial and effective remedy.

- These risks might be a result of flawed design of artificial intelligence systems (e.g. the system is programmed to discard female job applications) or the input of biased data (e.g. the system

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

is trained using only data from white males). They can also occur when the system 'learns' during the use phase, for example when an artificial intelligence systems 'learns' that students with the best academic results share the same postal codes which happen to be prevalently white in population. The risks will in such cases not stem from a flaw in the original design of the system but from the practical impacts of the correlations or patterns that the system identifies in a large dataset.

An employer was advertising for a job opening in a male-dominated industry via a social media platform. The platform's ad algorithm pushed jobs to only men to maximize returns on the number and quality of applicants. Souce: *Noam Scheiber, "Facebook Accused of Allowing Bias Against Women in Job Ads." The New York Times, September 18, 2018.*

- Artificial intelligence might also give rise to risks for privacy and protection of personal data.[4] For example, private and public actors can use artificial intelligence to identify people who want to remain anonymous. Employers can use artificial intelligence to observe the working patterns of their employees. Companies can track daily habits of people and listen in to private communication. Artificial intelligence technologies can be used for mass surveillance of the general population by state authorities. By analysing large amounts of non-personal data and identifying links among them, artificial intelligence can also be used to retrace and de-anonymise personal data about certain people.

Artificial intelligence 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: Larry Hardesty, "Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems," MIT News, February 11, 2018.*

ii. Safety and liability risks

- Artificial intelligence technologies may present new safety risks for users when they are embedded in products and services. For example, due to 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, such risks can be a result of flaws in the design of the artificial intelligence technology, problems with the availability and quality of data or problems stemming from machine learning. In case of connected objects, the loss of connectivity may lead to safety risks. While some of these risks are not limited to products and services relying on artificial intelligence, the presence of artificial intelligence may increase or aggravate such safety risks.

- If these risks materialise, the characteristics of artificial intelligence make it more difficult to attribute liability. This in turn makes it difficult for victims of damages to seek remedies under the current EU and national liability legislation.[5]

[4] The General Data Protection Regulation and the ePrivacy Directive (new ePrivacy Regulation under negotiation) broadly address these risks but there might be a need to examine whether artificial intelligence systems pose additional risks. The forthcoming evaluation of the General Data Protection Regulation will be of relevance in this context.

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

In the fatal accident of an Uber autonomous car in Arizona in 2018, the US National Traffic Safety Board observed that the software installed in Uber's vehicles that helps it detect and classify other objects did not include a consideration for jaywalking pedestrians. As a result, the system failed to recognise the woman who was walking her bike across the road as a person. *Source:* *https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf*

- The specific characteristics of artificial intelligence technologies, including complexity, autonomy and opacity ('black box-effect') may hamper the enforcement of existing EU law. Enforcement authorities might lack the means to verify how a given automated decision was taken, or whether existing rules were respected. Individuals and undertakings may face difficulties with access to justice through private enforcement. Developers and users of artificial intelligence do not necessarily keep information that make it possible to trace back problematic decisions that artificial intelligence systems make. Enforcement authorities and victims of possible damage may therefore find it difficult to scrutinise these decisions. Victims of damage may not have effective access to justice and be less protected compared to when damage is caused by traditional technologies. These various risks of harm occurring will increase as the field of applications for artificial intelligence widens and its use becomes more widespread.

In Spain, the complexity of the process used by the public authorities to decide on a discount on energy bills to at-risk individuals and families combined with the malfunctioning software and lack of information about the nature of rejections resulted in only 1,1 million people out of 5,5 potential beneficiaries profiting from the so-called Bono Social. The former government estimated 2,5 million people would receive the subsidy. *Source:* *https://civio.es/novedades/2019/07/12/being-ruled-through-secret-source-code-or-algorithms-should-never-be-allowed-in-a-social-and-democratic-state-under-the-rule-of-law/*

- Member States are already exploring options for national legislation to address the challenges of artificial intelligence. This may risk fragmenting the single market. A number of Member States (e.g. Estonia, Germany, Italy, Latvia and Sweden) have highlighted the need for regulatory action in their national strategies on artificial intelligence. Divergent national rules may create obstacles for companies who want to sell and operate artificial intelligence systems in the single market. Ensuring a common European approach would enable European companies to benefit from smooth access to the single market and support their competitiveness at global markets.

B. EU LEGISLATIVE FRAMEWORK FOR ARTIFICIAL INTELLIGENCE

- The development and use of artificial intelligence is in principle fully covered by a comprehensive body of EU legislation and further complemented by national legislation. As regards the protection of fundamental rights, the EU legislative framework consists of the Charter of Fundamental Rights and sectoral legislation, including the Race Equality Directive, the Employment Equality Directive and the Framework Decision on combating racism and xenophobia. There is also a body of legislation concerning personal data protection and privacy, notably the General Data Protection Regulation.

- The EU also has a legal framework for product safety and liability that consists of the General Product Safety Directive and a number of sector-specific rules covering different categories of products from machines to toys and medical devices. This is complemented by the Product

Liability Directive that provides the rules for compensation for damage suffered by a consumer as a result of defective products.

- While the EU legislation in principle applies to artificial intelligence systems, the question of whether it addresses adequately the risks that artificial intelligence systems pose to fundamental rights.

- In consultation with Member States, businesses and other stakeholders, the Commission identified the following weaknesses of the current legislative framework:

    - *Limitations of scope as regards fundamental rights*: for example, the Charter of Fundamental Rights does not apply to situations involving only private sector parties. Similarly, the EU legislation on fundamental rights covers only certain situations, for example access to employment, social protection, education, public services such as housing. It is does not apply horizontally and does not address all possible grounds of discrimination which the Charter sets out.

    - *Limitations of scope with respect to products*: the EU product safety legislation only applies to the placing of products on the market. Therefore, the safety requirements do not apply to services based on artificial intelligence (e.g. health services, financial services, transport services).

    - *Uncertainty as regards the division of responsibilities between different economic operators in the supply chain*: certain economic actors who develop and integrate artificial intelligence into products are not covered by the EU legislation on product safety. The rules do not apply to the developer of artificial intelligence unless (s)he is at the same time the producer of the product.

    - *Changing nature of products*: the integration of software, including artificial intelligence, into products can modify the functioning of products during their lifecycle. This is particularly true for products that require frequent software updates or which rely on machine learning. These features can give rise to new safety risks that were not present at the time when the product was placed on the market.

    - *Emergence of new risks*: the use of artificial intelligence in products and services can give rise to new safety risks. These may be linked to cyber threats, personal safety risks, risks that result from loss of connectivity, etc. These risks may be present at the time the products are placed products on the market or arise as a result of software updates or machine learning when the product is being used.

    - *Difficulties linked to enforcement*: given the opacity of artificial intelligence ('black-box' characteristics), it may be difficult for authorities to enforce EU legislation, whether on fundamental rights or on safety and liability. The lack of transparency of automated decision-making makes it difficult to prove possible discrimination. The lack of transparency will also make it difficult to attribute liability and prove causality between a damage and a defect in the design of artificial intelligence which in turn will make it difficult to have access to remedies.

- Given the issues identified above the Commission considers it necessary to review and where necessary complement the legislative framework applicable to artificial intelligence to make if fit for the current technological level of development and to take fully into account the human and ethical implications.

C. *LEGAL DEFINITION OF ARTIFICIAL INTELLIGENCE*

- A key issue for the future regulatory framework is the definition of the term "artificial intelligence". From the legal point of view, artificial intelligence is best defined by looking at its functions. A functional definition of artificial intelligence should look at the characteristics that differentiate artificial intelligence from more general terms, such as software. While the term 'software' is not defined in EU law, Directive 2009/24/EC provides a definition of 'computer programme' in recitals. This is defined as including programs in any form, including those which are incorporated into hardware. Therefore, artificial intelligence could be defined as software (integrated in hardware or self-standing) which provides for the following functions:
  - Simulation of human intelligence processes, such as learning, problem-solving, reasoning and self-correction;
  - Performing certain specified complex tasks, such as visual perception, speech recognition, decision-making and translation with a degree of autonomy, including through self-learning processes;
  - Involving the acquisition, processing and rational or reasoned analysis of data, typically in large quantities.
- While other, more technical approaches to the definition are possible[6], the Commission considers that these approaches would be less suitable in view of the fast pace of technological developments. The definition of artificial intelligence must be sufficiently flexible to accommodate technical progress while providing the necessary legal certainty.

D. *ADDRESSEES*

- Many economic actors are involved in the lifecycle of an artificial intelligence system. These include the developer of the algorithm, the producer, distributor or importer of a product based on artificial intelligence, the supplier of services based on artificial intelligence and the operator or user of a product based on artificial intelligence.
- The main principle guiding the attribution of roles and responsibilities in the future regulatory framework should be that the responsibility lies with the actor(s), who is/are best placed to address it. Therefore, while developers of artificial intelligence are best placed to address risks that arise from the development phase, their ability to control risks during the use phase may be more limited. This would also reflect the approach taken in EU safety legislation, which lays down obligations for different economic operators involved in placing products on the market, and to a limited extent for consumers and professional users, taking into account their different roles and knowledge.
- Therefore, the future regulatory framework for artificial intelligence should set out obligations for both developers and users of artificial intelligence. It could also include obligations for other groups, such as suppliers of services (e.g. third-party software update). This approach will require that different requirements are assigned to different types of addressees given the

[6] Such alternative technical approaches to the definition would for instance focus on systems that are trained with the machine learning technique, covering inter alia: deep learning and back-propagation, supervised learning, unsupervised learning, reinforcement learning, generative adversarial networks and symbolic reasoning.

very different roles that these actors have in the lifecycle of products and services based on artificial intelligence. The obligations on developers of artificial intelligence will focus on the risks that can be addressed while artificial intelligence systems are being developed, while the obligations on users of artificial intelligence will target the risks arising when artificial intelligence systems are being used. This approach would ensure that risks are managed comprehensively while not going beyond what is feasible for any given economic actor.

E. *POSSIBLE TYPES OF OBLIGATIONS*

- When designing the future regulatory framework for artificial intelligence, it will be necessary to decide on the types of legal requirements that should be imposed on the developers and users of artificial intelligence. These requirements can have either a preventative *ex ante* character (e.g. process requirements, including transparency and accountability that shape the design of artificial intelligence systems), or an *ex post* character (e.g. requirements on redress, remedies).

- Preventative *ex ante* requirements aim to reduce risks created by artificial intelligence before products or services that rely on artificial intelligence are placed on the market or are provided. *Ex post* requirements address the situations once the harm has materialised and would aim either to facilitate enforcement or to provide possibilities of redress or other types of remedy. While safety risks can largely be addressed through *ex ante* requirements, addressing liability issues requires ex post requirements. Addressing risks to fundamental rights will probably require a combination of *ex ante* and *ex post* requirements.

- *Ex ante* requirements could include:

  - Accountability and transparency requirements for developers (as part of the ex-post mechanism for enforcement) to disclose the design parameters of the artificial intelligence system, metadata of datasets used for training, on conducted audits, etc.;

  - Transparency and information requirements for users towards individuals, including for transparent and clear processes and outcomes for consumers;

  - General design principles for developers to reduce the risks of the artificial intelligence system;

  - Requirements for users regarding the quality and diversity of data used to train artificial intelligence systems;

  - Obligation for developers to carry out an assessment of possible risks and to take steps to minimise them; as well as obligation to keep records of these assessments and the steps to mitigate the risks;

  - Requirements for human oversight or a possible review of the automated decision by artificial intelligence by a human (e.g. in case of denial of social benefits) as regards non-personal data (to complement the obligations for automated decision making under the General Data Protection Regulation);

  - Additional safety requirements for producers of products, notably concerning the risk of cyber threats as well as risks for privacy, data protection and personal security with implications for safety (e.g. obligations for the producer to ensure a certain level of protection against such safety risks);

  - Requirements addressing the changes to the product during its life-cycle that could affect the safety of the product (e.g. machine learning, software updates).

- *Ex post* requirements could include:
  - Requirements on liability for harm/damage caused by a product or a service relying on artificial intelligence, including the necessary procedural guarantees (possibly differentiating between high-risk and low-risk applications); and
  - Requirements on enforcement and redress for individuals and undertakings, including access to existing alternative online dispute resolution systems.

- It is important to note that these requirements focus on process – reducing risks *ex ante* and establishing liability and possible remedies *ex post* – rather than on achieving specific results, i.e. specifying that artificial intelligence shall not discriminate. It would be technically difficult to avoid all risks associated with artificial intelligence. In addition, imposing specific results would likely require establishing new substantive rights for individuals, e.g. non-discrimination by artificial intelligence. This could lead to regulatory differences between artificial intelligence and traditional products and services.

- Based on this, the Commission is of the view that the regulatory framework should be based on requirements for the process rather than requirements for specific results, and that the requirements would need to be both *ex ante* and *ex post*. The stakeholders' input would be particularly welcome on the list of requirements presented above.

F. *POSSIBLE REGULATORY OPTIONS*

- Given the variety of risks covered, the Commission is looking at the following five regulatory options.

Option 1: Voluntary labelling

- This option would consist of a legal instrument setting out a voluntary labelling framework for developers and users of artificial intelligence. They could chose to comply, on a voluntary basis, with requirements for ethical and trustworthy artificial intelligence. If they complied, they would be allowed to use the label of 'ethical/trustworthy artificial intelligence'

- While participation in the labelling scheme would be voluntary, once the developer or user opted to use the label, the requirements would be binding. This scheme would have to include measures to ensure enforcement. It should be recognised that voluntary labelling may not be sufficient to address concerns linked to safety and liability, which are already covered by mandatory requirements in EU legislation. Similarly, a voluntary labelling framework would have limited impact on addressing risks linked to fundamental rights.

- A voluntary framework could nonetheless help to promote 'ethical and trustworthy' artificial intelligence. It would help Europe play an important role in the discussions on 'ethical and trustworthy' artificial intelligence at the international level while limiting the cost implications for both European and foreign developers and users of artificial intelligence.

Option 2: Sectorial requirements for public administration and facial recognition

- This option would focus on a specific area of public concern – the use of artificial intelligence by public authorities. This limited scope could reduce the regulatory and administrative burden and make it easier for developers and users of artificial intelligence systems to ascertain whether or not they fall within the scope of such regulatory instrument. Although this approach would only address the use of artificial intelligence by public authorities, it could have an important signalling effect on the private sector.

- Specific obligations for the use of artificial intelligence by public administrations could follow the model set out by the Canadian directive on automated decision-making[7]. This approach would aim to ensure that public authorities deploy automated decision systems in a way that reduces risks to public institutions, and leads to more efficient, accurate, consistent, and interpretable decisions. It could for instance set out requirements for impact assessments of the algorithms used, quality assurance, redress mechanisms and reporting.

- The requirements for public authorities could be coupled with specific rules on facial recognition systems, irrespective of whether they are used by public or private actors. These rules could regulate in more detail the use of facial recognition technology (also known as biometric remote identification) in public spaces, complementing the provisions of the General Data Protection Regulation.

- The General Data Protection Regulation already stipulates that data subjects shall receive information about the existence of automated decision-making, including profiling, and meaningful information about the logic involved, as well as the significance and the consequences for the data subject. In addition, unless he or she has given explicit consent, the data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects for him or her or significantly affects him or her. This right is subject to some exceptions, notably automated processing is authorised by Union or Member State law (e.g. for border control management). In these cases, the data controller needs to take measures to safeguard the data subject's rights and freedoms and legitimate interests, and to carry out a data protection impact assessment. These provisions mean that citizens must already be informed and consent to the use of artificial technology in situations when this can produce legal effects for them or affect them in a similar way.

- Building on these existing provisions, the future regulatory framework could go further and include a time-limited ban on the use of facial recognition technology in public spaces. This would mean that the use of facial recognition technology by private or public actors in public spaces would be prohibited for a definite period (e.g. 3-5 years) during which a sound methodology for assessing the impacts of this technology and possible risk management measures could be identified and developed. This would safeguard the rights of individuals, in particular against any possible abuse of the technology. It would be necessary to foresee some exceptions, notably for activities in the context of research and development and for security purposes (subject to a decision issued by a relevant court). By its nature, such a ban would be a far-reaching measure that might hamper the development and uptake of this technology. The Commission is therefore of the view that it would be preferable to focus at this stage on full implementation of the provisions in the General Data Protection Regulation. The Commission will consider whether to adopt guidance to facilitate this.

[7] More details are available at: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

Option 3: Mandatory risk-based requirements for high-risk applications

- This option would foresee legally binding requirements for developers and users of artificial intelligence, building on existing EU legislation. Given the need to ensure proportionality, the new requirements could apply only to high-risk applications of artificial intelligence. This risk-based approach would focus on areas where the public is at risk or an important legal interest is at stake. This strictly targeted approach would not add any new additional administrative burden on applications that are deemed 'low-risk'.

- A differentiated risk-based approach would allow for better proportionality of the regulatory intervention, but it also requires clear criteria to differentiate between 'low-risk' and 'high-risk' systems. This is necessary to ensure smooth implementation by all relevant economic actors as well as national competent authorities.

- The criteria to determine the level of risk could include the following:

  a) Defining high-risk sectors (e.g. healthcare, transport), possibly in combination with an indicative or exhaustive list with the possibility to amend such list;

  b) Defining high-risk applications (e.g. predictive policing), possibly in combination with an indicative or exhaustive list with the possibility to amend such list;

  c) (Self-)Identifying the level of risks through a risk assessment carried out by the developer and/or user of artificial intelligence;

  d) Other types of criteria taking into account the context:

     - o whether the individual or legal entities cannot avoid being affected by the output of an artificial intelligence system, or risk suffering serious negative consequences as a result of the decision to 'opt out' (e.g. healthcare applications);

     - o how important the output of the artificial intelligence system is for an individual or legal entity (e.g. social security benefits);

     - o whether the output of the artificial intelligence system with a significant effect for an individual or legal entity is irreversible (e.g. collision avoidance in self-driving vehicles);

     - o whether the individuals or legal entities affected by the output of the artificial intelligence system are in a specific area with a high risk of discrimination (e.g. recruitment proceedings).

- Having considered the different options, the Commission is of the opinion that the definition of 'high-risk applications should rely on a cumulative application of two criteria.

  - o an exhaustive list of sectors (e.g. healthcare, transport, police, judiciary) that would be specified in an annex and subject to amendments by means of delegated acts if necessary, and

  - o a more abstract definition of 'high-risk' applications along the lines of '*high-risk applications means applications of artificial intelligence which can produce legal effects for the individual or the legal entity or pose risk of injury, death or significant material damage for the individual or the legal entity*'.

- Such a combined application of the two criteria would ensure a narrow scope of application while providing the maximum level of legal certainty for relevant economic operators. Only those applications meeting both of these criteria would be subject to the mandatory requirements.

- For 'low-risk' applications, the existing provisions of EU legislation would apply. That includes for example the provisions of the General Data Protection Regulation on the information the data subject must receive about the use of automated processing, including profiling, and the obligation to carry out a data protection impact assessment.

Option 4: Safety and Liability

- The EU acquis includes an extensive body of product safety and liability legislation. While this legal framework has proven its effectiveness, it would be appropriate to consider targeted amendments of the EU safety and liability legislation (including the General Product Safety Directive, the Machinery Directive, the Radio Equipment Directive and the Product Liability Directive) to address the specific risks of artificial intelligence.
- The Report on the broader implications of artificial intelligence, Internet of Things and other digital technologies for the EU safety and liability framework, which accompanies this White Paper, provides an overview of EU legislation and identifies the shortcomings with respect to the specific risks posed by artificial intelligence and other digital technologies. The aim of the targeted adjustments of EU legislation would be to address those shortcomings.
- Specific risks which are currently not addressed or not addressed adequately include the risks of cyber threats, risk to personal security, to privacy and to personal data protection. New requirements should address these issues and the risks that are related to software updates and machine learning when products are being used In addition, adjustments may be needed to clarify the responsibility of developers of artificial intelligence and to distinguish them from the responsibilities of the producer of the products using the artificial intelligence. The scope of the legislation should also be reviewed to determine whether artificial intelligence systems, which are currently not covered by the definition of products, should be covered. Similar changes will be also required to the provisions concerning the liability for damages caused by defective products. Changes to the Product Liability Directive might also aim to facilitate the burden of proof for consumers to ensure easier access to justice.
- To assess the impacts of these targeted changes, the Commission will launch the work on the impact assessment(s). The changes could take the form of specific amendments to individual pieces of EU legislation or a new horizontal piece of legislation that would include the relevant requirements for artificial intelligence.
- This option could be combined with any of the other three options set out above. This combined approach would ensure that all relevant risks posed by artificial intelligence systems are addressed while taking into account the specificities of the existing legal framework.

Option 5: Governance

- To ensure that any future rules on artificial intelligence bring about the anticipated benefits for consumers and businesses, an effective system of enforcement must be an essential component of the future regulatory framework. This will require a strong system of public oversight. This system should, as much as possible, build on the existing network of authorities. It should consist of national authorities that will be entrusted with the implementation and enforcement of the future regulatory framework. In addition, it will be necessary to foresee a mechanism to foster cooperation among national authorities across the EU and facilitate the exchange of information, knowledge, and best practice.

- There are already a number of different authorities involved in implementing and enforcing EU legislation, including in the areas of fundamental rights, data protection and safety. For example, under the General Data Protection Regulation, each Member State had to appoint one or more supervisory authorities to monitor the application of the Regulation. The Regulation also foresees a European Data Protection Board with a number of tasks, including advising the Commission on issues linked to data protection and preparing guidelines, recommendations and best practices. EU safety legislation, including the General Product Safety Directive and the new Regulation on market surveillance and compliance of products, also requires Member States to nominate authorities to monitor the compliance of products with the safety requirements. Both pieces of legislation foresee specific cooperation mechanisms: a Consumer Safety Network and a Union Product Compliance Network.

- Given the specificity and complexity of regulatory challenges posed by artificial intelligence, it would nonetheless be appropriate for Member States to appoint authorities responsible for monitoring the overall application and enforcement of the future regulatory framework for artificial intelligence. Member States will be free to decide that these tasks should be entrusted to existing authorities in order to minimise any additional administrative burden. These authorities could be responsible not only for monitoring the application of the new legislation addressing specifically artificial intelligence but also provide guidance on horizontal questions of relevance for the overall EU regulatory framework for artificial intelligence. The Commission will also set up an appropriate mechanism to promote cooperation between the relevant national authorities.

- [The Commission is of the view that **Option 3** set out above, combined with **Option 4** and **Option 5**, seems to be the most promising to address the risks specific to artificial intelligence. Therefore, the Commission may consider a combination of a horizontal instrument setting out transparency and accountability requirements and covering also the governance framework, complemented by targeted amendments of existing EU safety and liability legislation. The horizontal instrument would be relevant both for enforcing EU fundamental rights legislation as well as existing EU safety and liability legislation, and possibly also national legislation.]

6. CONCLUSION

- [*to be developed, also referring to the broad outline of action after the consultation phase*]

The Commission invites for comments on the proposals set out in the White Paper. They may be sent by XXX 2020, either by e-mail to: YYY or by post to: ZZZ.

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.