Council of the
European Union

Brussels, 23 April 2021
(OR. en)

8115/21
ADD 2

Interinstitutional File:
2021/0106(COD)

TELECOM 156
JAI 429
COPEN 191
CYBER 108
DATAPROTECT 103
EJUSTICE 41
COSI 69
IXIM 74
ENFOPOL 148
FREMP 103
RELEX 347
MI 271
COMPET 275
IA 60
CODEC 573

COVER NOTE

From:	Secretary-General of the European Commission, signed by Ms Martine DEPREZ, Director
date of receipt:	22 April 2021
To:	Mr Jeppe TRANHOLM-MIKKELSEN, Secretary-General of the Council of the European Union
No. Cion doc.:	SWD(2021) 84 final PART 1/2
Subject:	COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT Accompanying the Proposal for a Regulation of the European Parliament and of the Council LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

Delegations will find attached document SWD(2021) 84 final PART 1/2.

Encl.: SWD(2021) 84 final PART 1/2

EUROPEAN
COMMISSION

Brussels, 21.4.2021
SWD(2021) 84 final

PART 1/2

COMMISSION STAFF WORKING DOCUMENT

IMPACT ASSESSMENT

Accompanying the

Proposal for a Regulation of the European Parliament and of the Council

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{COM(2021) 206 final} - {SEC(2021) 167 final} - {SWD(2021) 85 final}

Table of contents

1. INTRODUCTION: TECHNOLOGICAL, SOCIO-ECONOMIC, LEGAL AND POLITICAL CONTEXT 1

1.1. Technological context 2
1.2. Socio-economic context 3
1.3. Legal context 5
1.3.1. Relevant fundamental rights legislation 5
1.3.2. Relevant product safety legislation 6
1.3.3. Relevant liability legislation 7
1.4. Political context 9
1.5. Scope of the impact assessment 12

2. PROBLEM DEFINITION 13

2.1. What are the problems? 13
2.2. What are the main problem drivers? 28
2.3. How will the problem evolve? 30

3. WHY SHOULD THE EU ACT? 30

3.1. Legal basis 30
3.2. Subsidiarity: Necessity of EU action 31
3.3. Subsidiarity: Added value of EU action 32

4. OBJECTIVES: WHAT IS TO BE ACHIEVED? 32

4.1. General objectives 32
4.2. Specific objectives 32
4.3. Objectives tree/intervention logic. 34

5. WHAT ARE THE AVAILABLE POLICY OPTIONS? 36

5.1. What is the baseline from which options are assessed? 37
5.2. Option 1: EU legislative instrument setting up a voluntary labelling scheme 39
5.3. Option 2: A sectoral, 'ad-hoc' approach 43
5.4. Option 3: Horizontal EU legislative instrument establishing mandatory requirements for high-risk AI applications 48
5.5. Option 3+: Horizontal EU legislative instrument establishing mandatory requirements for high-risk AI applications + co-regulation through codes of conduct for non-high risk applications 61
5.6. Option 4: Horizontal EU legislative instrument establishing mandatory requirements for all AI applications, irrespective of the risk they pose 62
5.7. Options discarded at an early stage 62

6. WHAT ARE THE IMPACTS OF THE POLICY OPTIONS? 64

6.1. Economic impacts 64
6.1.1. Functioning of the internal market 64
6.1.2. Impact on uptake of AI 64
6.1.3. Costs and administrative burdens 65
6.1.4. SME test 70
6.1.5. Competitiveness and innovation 72
6.2. Costs for public authorities 74
6.3. Social impact 75
6.4. Impacts on safety 76

6.5. Impacts on fundamental rights 76
6.6. Environmental impacts 78
7. HOW DO THE OPTIONS COMPARE? 79
7.1. Criteria for comparison 79
7.2. Achievement of specific objectives 80
7.2.1. First specific objective: Ensure that AI systems placed on the market and used are safe and respect fundamental rights and Union values 80
7.2.2. Second specific objective: Ensure legal certainty to facilitate investment and innovation 81
7.2.3. Third specific objective: Enhance governance and effective enforcement of fundamental rights and safety requirements applicable to AI systems 82
7.2.4. Fourth specific objective: Facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation 82
7.3. Efficiency 83
7.4. Coherence 84
7.5.Proportionality 85
8. PREFERRED OPTION 85
9. HOW WILL ACTUAL IMPACTS BE MONITORED AND EVALUATED? 89

1. INTRODUCTION: TECHNOLOGICAL, SOCIO-ECONOMIC, LEGAL AND POLITICAL CONTEXT

As part of the Commission's overarching agenda of making Europe ready for the digital age, the Commission is undertaking considerable work on Artificial Intelligence (AI). The overall EU strategy proposed in the White Paper on AI proposes an ecosystem of excellence and trust for AI.[1] The concept of an **ecosystem of excellence** in Europe refers to measures which support research, foster collaboration between Member States and increase investment into AI development and deployment. The **ecosystem of trust** is based on EU values and fundamental rights, and foresees robust requirements that would give citizens the confidence to embrace AI-based solutions, while encouraging businesses to develop them. 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.[2]

The development of an ecosystem of trust is intended as a comprehensive package of measures to address problems posed by the introduction and use of AI. In accordance with the White Paper and the Commission Work Programme, the EU plans to adopt a set of three inter-related initiatives related to AI:

(1) **European legal framework for AI** to address fundamental rights and safety risks specific to the AI systems (Q2 2021);
(2) EU rules to address **liability issues** related to new technologies, including AI systems (Q4 2021-Q1 2022);
(3) **Revision of sectoral safety legislation** (e.g. Machinery directive, Q1 2021, General Product Safety Directive, Q2 2021).

These three initiatives would be **complementary and their adoption will proceed in stages.** Firstly, as entrusted by the European Council, requested by the European Parliament and supported by the results of the public consultation on the White Paper on AI, the European Commission will adopt European legal framework for AI. **This legal framework should set the ground for other forthcoming initiatives** by providing: (1) a definition of an AI system; (2) a definition of 'high risk' AI system, and (3) common rules to ensure that AI systems placed or put into service in the Union market are trustworthy. The introduction of the European legal framework for AI will be **supplemented with revisions of the sectoral safety legislation** and **changes to the liability rules.** This staged, step-by-step and complementary approach to regulate AI aims to ensure regulatory coherence throughout the Union, therefore contributing to legal certainty for developers and users of AI systems and citizens. More details on the scope of the existing safety and liability legislation are discussed in section 1.3. (legal context) and the interaction between the three initiatives are presented in section 8 (preferred option).

This impact assessment focuses on the first AI initiative, the **European legal framework for AI**. The purpose of this document is to assess the case for action, the objectives, and the impact of different policy options for a European framework for AI, as envisaged by the 2020 Commission work programme.

The Proposal for a European legal framework for AI and this impact assessment build on two years of analysis of evidence and involvement of stakeholders, including academics, businesses, non-governmental organisations, Member States and citizens. The preparatory work started in 2018 with the setting up of a High-Level Expert Group on AI (HLEG) which had an inclusive and broad

[1] European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, 2020.

[2] See above, European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, 2020. p. 25.

composition of 52 well-known experts tasked to advise the Commission on the implementation of the Commission's Strategy on Artificial Intelligence. In April 2019, the Commissioned welcomed[3] the key requirements set out in the HLEG ethics guidelines for Trustworthy AI,[4] which had been revised to take into account more than 500 submissions from stakeholders. The Assessment List for Trustworthy Artificial Intelligence (ALTAI)[5] made these requirements operational in a piloting process with over 350 organisations. The White Paper on Artificial Intelligence further developed this approach, inciting comments from more than 1250 stakeholders. As a result, the Commission published an Inception Impact Assessment that in turn attracted more than 130 comments.[6] Additional stakeholder workshops and events were also organised, the results of which support the analysis and the proposals made in this impact assessment.[7]

1.1. Technological context

Today, AI is one of the most vibrant domains in scientific research and innovation investment around the world. Approaches and techniques differ according to fields, but overall **AI is best defined as an emerging general-purpose technology**: a very powerful family of computer programming techniques that can be deployed for desirable uses, as well as more harmful ones.[8] The precise definition of AI is highly contested.[9] In 2019, the Organisation for Economic Co-operation and Development (OECD) adopted the following definition of an **AI system**: 'An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.'[10]

The *OECD Report on Artificial Intelligence in Society* provides a further explanation on what an AI system is.[11] An AI system, also referred to as 'intelligent agent', "consists of **three main elements**: sensors, operational logic and actuators. Sensors collect raw data from the environment, while actuators act to change the state of the environment. Sensors and actuators are either machines or humans.[12] The key power of an AI system resides in its operational logic. For a given set of objectives and based on input data from sensors, the operational logic provides output for the actuators. These take the form of recommendations, predictions or decisions that can influence the state of the environment."[13]

---

3 European Commission, *Building Trust in Human-Centric Artificial Intelligence*, COM(2019) 168.

4 HLEG, *Ethics Guidelines for Trustworthy AI*, 2019.

5 HLEG, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, 2020.

6 European Commission, *Inception Impact Assessment For a Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence.*

7 For details of all the consultations that have been carried out see Annex 2.

8 For the discussion on why AI can be considered as an emerging general purpose technology see for instance Agrawal, A., J. Gans and A. Goldfarb, *Economic policy for artificial intelligence*, NBER Working Paper No. 24690, 2018; Brynjolfsson, E., D. Rock and C. Syverson, *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics*, NBER Working Paper No. 24001, 2017.

9 For the analysis of the available definitions and they scope see e.g. JRC, *Defining Artificial Intelligence, Towards an operational definition and taxonomy of artificial intelligence*, 2020. As well as the forthcoming update to this JRC Technical Report. The forthcoming update provides a qualitative analysis of 37 AI policy and institutional reports, 23 relevant research publications and 3 market reports, from the beginning of AI in 1955 until today.

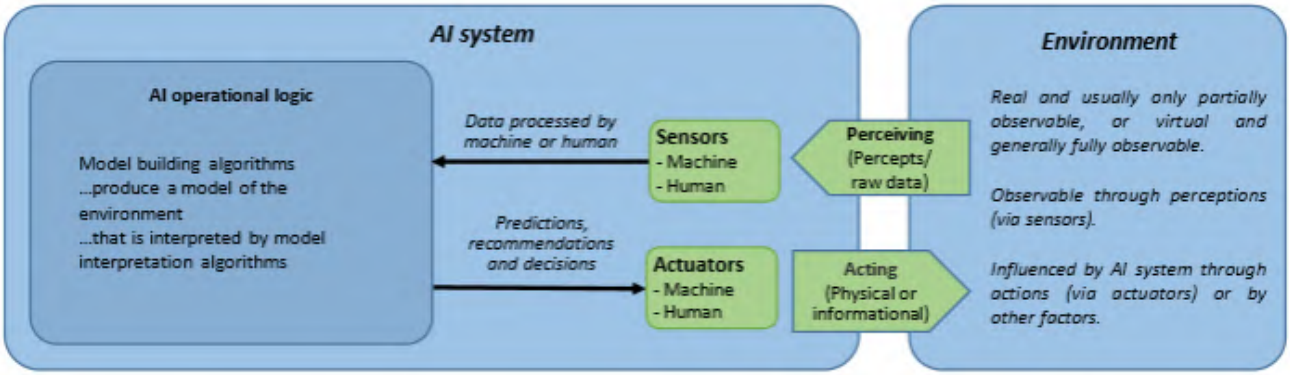
10 OECD, *Recommendation of the Council on Artificial Intelligence, 2019*.

11 OECD, *Artificial Intelligence in Society*, 2019, p. 23.

12 See above, OECD, *Artificial Intelligence in Society*, 2019.

13 See above, OECD, *Artificial Intelligence in Society*, 2019.

Figure 1: A high-level conceptual view of an AI system



Source: OECD, Report – Artificial Intelligence in Society, p.23.

AI systems are typically **software-based, but** often **also embedded in hardware-software systems**. Traditionally AI systems have focused on **'rule-based algorithms'** able to perform complex tasks by automatically executing rules encoded by their programmers.[14] However, recent developments of AI technologies have increasingly been on so called '**learning algorithms'**. In order to successfully 'learn', many machine learning systems require substantial computational power and availability of large datasets ('big data'). This is why, among other reasons,[15] despite the development of 'machine learning' (ML),[16] AI scientists continue to combine traditional rule-based algorithms and 'new' learning based AI techniques.[17] As a result, the **AI systems currently in use often include both rule-based and learning-based algorithms.**

1.2. Socio-economic context

The use of AI systems leads to important breakthroughs in a number of domains. By improving prediction, optimising operations and resource allocation, and personalizing service delivery, the use of **AI can support socially and environmentally beneficial outcomes and provide key competitive advantages to companies**. The use of AI systems in healthcare, farming, education, infrastructure management, energy, transport and logistics, public services, security, and climate change mitigation, can help solve complex problems for the public good. Combined with robotics and the Internet of Things (IoT), AI systems are increasingly acquiring the potential to carry out complex tasks that go far beyond human capacity.[18]

A recent Nature article found that AI systems could enable the accomplishment of 134 targets across all the Sustainable Development Goals, including finding **solutions to global climate problems, reducing poverty, improving health and the quality and access to education, and making our cities safer and greener**.[19] In the ongoing Covid-19 pandemic, AI systems are being

---

[14] Rule-based algorithms are well-suited to execute applications and tasks that require high reliability and robustness. They can be used for complex simulations and can be adapted by adding new information to the system that is then processed with the help of the established rules.

[15] Rule-based algorithms are well-suited to execute applications and tasks that require high reliability and robustness.

[16] One of the best known subfields of AI technology where algorithms 'learn' from data is 'machine learning' (ML) that predicts certain features, also called 'outputs', based on a so called 'input'. 'Learning' takes place when the ML algorithm progressively improves its performance on the given task.

[17] For now, the majority of AI systems are rule-based.

[18] AI is a technology, thus it cannot be directly compared or equated with human intelligence. However, to explain how AI systems achieve 'artificial intelligence' the parallel to humans is telling. The AI 'brain' is increasingly acquiring the potential to carry our complex tasks which require a 'body' (sensors, actuators) and a nervous system (embedded AI). This combination of 'brain' and 'body' connected through 'a nervous system' allows AI systems to perform tasks such as exploring space, or the bottom of the oceans. For a graphical overview, see Figure 1.

[19] Vinuesa, R. et al., 'The role of artificial intelligence in achieving the Sustainable Development Goals', *Nature communications* 11(1), 2020, pp. 1-10.

used, for example, in the quest for vaccines, in disease detection via pattern recognition using medical imagery, in calculating probabilities of infection, or in emergency response with robots replacing humans for high-exposure tasks in hospitals.[20] This example alone already indicates the breadth of possible benefits of AI systems. Other practical applications further show how citizens can reap a lot of benefits whern accessing improved services such as personalised telemedicine care, personalised tutoring tailored to each student, or enhanced security through applications that ensure more efficient protection against cybersecurity risks.

The successful uptake of AI technologies also has the potential to accelerate **Europe's economic growth and global competitiveness**.[21] McKinsey Global Institute estimated that by 2030 AI technologies could contribute to about 16% higher cumulative global gross domestic product (GDP) compared with 2018, or about 1.2% additional GDP growth per year.[22] AI systems and the new business models they enable are progressively developing to at-scale deployment. Accordingly, those AI systems will increasingly impact all sectors of the economy. The International Data Corporation AI market development forecast suggests that global revenues for the AI market are expected to double and surpass USD 300 billion by as early as 2024.[23] Many businesses in various sectors of the EU economy are already seizing these opportunities.[24] In addition to the ICT sector, the sectors using AI most intensively are education, health, social work and manufacturing.[25] However, Europe is home to only 3 of the top 25 AI clusters worldwide and has only a third as many AI companies per million employees as the US.[26]

Table 1: AI technologies adopted in European businesses

AI TECHNOLOGIES	CURRENTLY USE IT	PLAN TO USE IT
Process or equipment optimisation	13%	11%
Anomaly detection	13%	7%
Process automation	12%	11%
Forecasting, price-optimisation and decision-making	10%	10%
Natural language processing	10%	8%
Autonomous machines	9%	7%
Computer vision	9%	7%
Recommendation/personalisation engines	n/a	7%
Creative and experimentation activities	7%	4%
Sentiment analysis	3%	3%

Source: Ipsos Survey, 2020[27]

The same elements and techniques that power socio-economic benefits of AI systems can also bring about **risks or negative consequences for individuals or for society as a whole**.[28] For example,

[20] OECD, *Using artificial intelligence to help combat COVID-19*, 2020.

[21] According to McKinsey, the cumulative additional GDP contribution of new digital technologies could amount to €2.2 trillion in the EU by 2030, a 14.1% increase from 2017, McKinsey, *Shaping the Digital Transformation in Europe*, 2020). PwC comes to an almost identical forecast increase at global level, amounting to USD 15.7 trillion, PwC, *Sizing the prize: What's the real value of AI for your business and how can you capitalise?*, 2017.

[22] For a comparison, the introduction of steam engines in the 1800s boosted labour productivity by 0.3% a year and spread of IT during the 2000s by 0.6% a year (ITU/McKinsey, *Assessing the Economic Impact of Artificial Intelligence*, 2018).

[23] IDC, *IDC Forecasts Strong 12.3% Growth for AI Market in 2020 Amidst Challenging Circumstances*, 2020.

[24] OECD, *Artificial Intelligence in Society*, 2019.

[25] European Commission, Ipsos Report, *European enterprise survey on the use of technologies based on artificial intelligence*, 2020.

[26] McKinsey, *How nine digital frontrunner can lead on AI in Europe*, 2020.

[27] European Commission, Ipsos Survey, *European enterprise survey on the use of technologies based on artificial intelligence*, 2020. (Company survey across 30 European countries, N= 9640).

deployment of AI systems may be intentionally used by a developer or an operator to deceive or manipulate human choices, or altogether disable human agency, control and intermediation. This possible use of AI could have strong negative consequences for the protection of fundamental rights and for human safety.[29] In the world of work, AI systems could also undermine the effective enforcement of labour and social rights.

In light of the speed of technological change and possible challenges, the EU is committed to strive for a balanced approach. European Commission President von der Leyen stated: 'In order to release that potential **we have to find our European way**, balancing the flow and wide use of data while preserving high privacy, security, safety and ethical standards.'[30]

1.3. Legal context

European Union law does not have a specific legal framework for AI. Thus, as it currently stands, EU law **does not provide for a definition of an AI system, nor for horizontal rules related to the classification of risks related to AI technologies**. The development and uptake of AI systems more broadly, as outlined in this section, takes place in the context of the existing body of EU law that provides non-AI specific principles and rules on protection of fundamental rights, product safety, services or liability issues.

1.3.1. Relevant fundamental rights legislation

The Union is founded on the values of human dignity and respect of human rights that are further specified in **the EU Charter of Fundamental Rights** (the Charter). The provisions of the Charter are addressed to the institutions and bodies of the Union and to the Member States only when they are implementing Union law. Some fundamental rights obligations are further provided for in EU secondary legislation, including in the field of data protection, non-discrimination and consumer protection. This body of EU secondary legislation is applicable to both public and private actors whenever they are using AI technology.[31]

In this context, the EU acquis on data protection is particularly relevant. The **General Data Protection Regulation**[32] and the **Law Enforcement Directive**[33] aim to protect the fundamental rights and freedoms of natural persons, and in particular their right to the protection of personal data, whenever their personal data are processed. This covers the processing of personal data through 'partially or solely automated means',[34] including any AI system.[35] Users that determine

---

[28] For detailed review of various human rights related risks see e.g. Horizon 2020 funded SIENNA project, Rodrigues, R, Siemaszko, K and Warso. Z, D4.2: *Analysis of the legal and human rights requirements for AI and robotics in and outside the EU* (Version V2.0). Zenodo, 2019. The researchers in this project identified the following main concerns related to fundamental rights and AI systems: lack of algorithmic transparency / transparency in automated decision-making; unfairness, bias, discrimination and lack of contestability; intellectual property issues; issues related to AI vulnerabilities in cybersecurity; issues related to impacts on the workplace and workers; privacy and data protection issues; and liability issues related to damage caused by AI systems and applications. See also, JRC Report, *Artificial Intelligence: A European Perspective*, 2018.

[29] For the discussion see Problem 2 below.

[30] Ursula von der Leyen, *Political Guidelines for the Next European Commission 2019-2024*, 2019, p. 13.

[31] For a comprehensive overview of applicable EU primary and secondary legislation see SIENNA project, ibid.

[32] Regulation (EU) 2016/679 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 and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[33] 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, and repealing Council Framework Decision 2008/977/JHA.

[34] This is a rather broad term, encompassing in principle any AI and automated decision-making systems.

[35] For an overview on how GDPR applies to AI, see e.g. Spanish Data Protection Agency, *RGPD compliance of processings that embed Artificial Intelligence: An introduction*, 2020. See also European Data Protection Supervisor

the purpose and means of the AI processing ('data controllers') have to comply with a number of data processing principles such as lawfulness, transparency, fairness, accuracy, data minimization, purpose and storage limitation, confidentiality and accountability. On the other hand, natural persons, whose personal data are processed, have a number of rights, for instance, the right to access, correction, not to be subject to solely automated decision-making with legal or similarly significant effects unless specific conditions apply. Stricter conditions also apply for the processing of sensitive data, including biometric data for identification purposes, while processing that poses high risk to natural persons' rights and freedoms requires a data protection impact assessment.

Users of AI systems are also bound by existing **equality directives.** The EU equality acquis prohibits discrimination based on a number of protected grounds (such as racial and ethnic origin, religion, sex, age, disability and sexual orientation) and in specific context and sectors (for example, employment, education, social protection, access to goods and services).[36] This existing acquis has been complemented with the new EU Accessibility Act setting requirements for the accessibility of goods and services, to become applicable as of 2025.[37]

Consumer protection law and obligations to abstain from any unfair commercial practices listed in the Unfair Commercial Practice Directive[38] are also highly relevant for businesses using AI systems.

Furthermore, EU secondary law in the area areas of **asylum, migration, judicial cooperation in criminal matters, financial services and online platforms** is also relevant from a fundamental rights perspective when AI is developed and used in these specific contexts.

1.3.2. Relevant product safety legislation

In addition, there is a solid body of EU **secondary law on product safety.**[39] The EU safety legislation aims to ensure that only safe products are placed on the Union market. The overall EU architecture on safety is based on the combination of horizontal and sectoral rules. This includes the General Product Safety Directive (GPSD)[40] applicable to consumer products, insofar there are not more specific provisions in harmonised sector-specific safety legislation, as for example, the

---

Opinion on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence and trust, 2020.

36 E.g. Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin; Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation; Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast); Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services.

37 Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services, OJ L 151, 7.6.2019, pp. 70–115.

38 Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive').

39 In the context of EU sector specific safety legislation, so-called old and new approaches are traditionally distinguished. The 'Old Approach' refers to the very initial phase of EU regulation on products, whose main feature was the inclusion of detailed technical requirements in the body of the legislation. Certain sectors such as food or transport are still being regulated on the basis of 'old approach' legislations with detailed product requirements for reasons of public policy or because of their reliance on international traditions and/or agreements which cannot be changed unilaterally. The so-called 'New Approach' was developed in 1985, whose main objective was to restrict the content of legislation to 'essential (high-level) requirements' leaving the technical details to European harmonised standards. On the basis of the New Approach, the New Legislative Framework (NLF) was then developed in 2008, introducing harmonised elements for conformity assessment, accreditation of conformity assessment bodies and market surveillance. Today more than 20 sectors are regulated at EU level based on the NLF approach, e.g. medical devices, toys, radio-equipment or electrical appliances.

40 Directive 2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety.

Machinery Directive (MD),[41] the Medical Device Regulation (MDR) and the EU framework on the approval and market surveillance of motor vehicles[42] (in particular Vehicle Safety Regulation).[43]

Reviews of both the MD and the GPSD are currently under way.[44] Those reviews aim to respond, among other things, to the challenges of new technologies, such as IoT, robotics and AI. In addition, delegated acts are expected to be soon adopted by the Commission under the Radio Equipment Directive[45] to enact certain new requirements on data protection and privacy, cybersecurity and harm to the network. Moreover, in the automotive sector new rules on automated vehicles, cybersecurity and software updates of vehicles will become applicable as part of the vehicle type approval and market surveillance legislation from 7 July 2022.

While the European Commission *Report on safety and liability implications of AI, the Internet of Things and Robotics* identifies the review of the General Product Safety Directive, the Machinery Directive and the Radio Equipment Directive as priorities, other pieces of product legislation may well be updated in the future in order to address existing gaps linked to new technologies.

The **product safety legislation is technology-neutral and focuses on the safety of the final product as a whole.** The revisions of the product safety legislation do not have the objective to regulate AI as such, but aim primarily at ensuring that the integration of AI systems into the overall product will not render a product unsafe and the compliance with the sectoral rules will not be affected.[46]

1.3.3. Relevant liability legislation

Safety legislation sets rules to ensure that products are safe and safety risks are addressed, nevertheless, damages can still occur. For that purpose, the liability rules at national and EU level complement the safety legislation and determine **which party is liable for harm**, and under which conditions **a victim can be compensated**. A longstanding approach within the EU with regard to product legislation is based on a combination of both safety and liability rules. In practice, while being driven by different regulatory rationales and objectives, safety and liability initiatives are essential and complementary in nature.

At EU level, the **Product Liability Directive**[47] (PLD) is currently **the only EU legal framework that harmonizes part of national liability law**, introducing a system of 'strict' liability without

---

[41] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast).

[42] New rules on automated vehicles, cybersecurity and software updates of vehicles will become applicable as part of the vehicle type approval and market surveillance legislation as from 7 July 2022, providing notably for obligations for the manufacturer to perform an exhaustive risk assessment (including risks linked to the use of AI) and to put in place appropriate risk mitigations, as well as to implement a comprehensive risk management system during the lifecycle of the product.

[43] Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users, amending Regulation (EU) 2018/858 of the European Parliament and of the Council (OJ L 325, 16.12.2019), p. 1.

[44] The Commission intends to adopt proposals in the first quarter of 2021 for the revision of the Machinery Directive and second quarter of 2021 for the revision of the General Product Safety Directive.

[45] Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC.

[46] Safety legislation assesses a broad spectrum of risks and ensures that the overall interplay between different types and elements of risks does not render a product or service as a whole unsafe. These measures will also facilitate the uptake and increase certainty, by ensuring that the integration of new technologies in the product does not endanger the overall safety of a product or service. More detailed explanation about the interaction between the AI initiative examined in this impact assessment and the sectoral product legislation can be found in Annex 5.3.

[47] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products. See in particular Article 6(1), listing out

fault of the producer for physical or material damage caused by a defect in products placed on the Union market.[48] While the PLD provides legal certainty and uniform consumer protection as a safety net applicable to all products, its rules also face increasing challenges posed by emerging technologies, including AI.[49] As part of the upcoming revision, the Commission will be exploring several options to adapt the EU product liability rules to the digital world, for instance by adapting the definition of product and producer, or by extending the application of the strict liability regime to certain other types of damages (e.g. to privacy or personal data). The Commission will also explore options on how to address the current imbalance between consumers and producers by reversal or alleviation of the burden of proof (with access to information and presumption of defectiveness under certain circumstances), and explore the abolishment of existing timelines and threshold (€500). At national level, non-harmonised civil liability frameworks complement these Union rules by ensuring compensation for damages from products and services and by addressing different liable persons. National liability systems usually include fault-based and strict liability regimes.[50]

In order to ensure that victims who suffer damage to their life, health or property as a result of AI technologies have access to the same compensation as victims of other technologies, the Commission has announced possible revision of rules on liability.[51] The main objective of the revision is to ensure that damages caused by AI systems are covered. In order to achieve this objective, together with the update of the PLD, the Commission is also considering possible new AI-specific rules harmonising certain aspects of national civil liability frameworks with regard to the liability for certain AI systems. In particular, options which are currently under evaluation include the possible setting of strict liability for AI operators, possibly combined with mandatory insurance for AI applications with a specific risk profile as well as adaptation of burden of proof concerning causation and fault for all other AI applications.[52]

In addition to the general review of liability rules, the Commission is examining liability challenges which are specific to certain sectors, such as health-care, and which may deserve specific considerations.

The relationship between the proposed European legal framework for AI analysed in this impact assessment and the forthcoming new rules on liability is further discussed under the preferred option in section 8. In terms of timing for the adoption of the new rules on liability, the Commission decided for a **staged approach**. First, the Commission will propose in Q2 2021 the **AI horizontal**

---

the relevant circumstances under which a product is considered defective, i.e. when not providing the 'safety which a person is entitled to expect'; see also whereas 6 and 8 of this Directive.

[48] To obtain compensation, the injured party shall prove in court three elements: defect, damage, causal link between the two. The PLD is technology-neutral by nature.

[49] It is unclear whether the PLD still provides the intended legal certainty and consumer protection when it comes to AI systems and the review of the directive will aim to address that problem. Software, artificial intelligence and other digital components play an increasingly important role in the safety and functioning of many products, but are not expressly covered by the PLD. The PLD also lacks clear rules in relation to changes, updates or refurbishments of products, plus it is not always clear who the producer is where products are adapted or combined with services. Finally, establishing proof of defect, harm and causation is in many cases excessively difficult for consumers, who are at a disadvantage in terms of technical information about the product, especially in relation to complex products such as AI. See Evaluation SWD(2018)157 final of Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, accompanying Report COM(2018) 246 from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the application of the Directive; See also Report COM(202064 on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics.

[50] For the overview of national liability regimes applicable to the AI technologies, see e.g. European Parliamentary Research Service, Civil liability regime for artificial intelligence: European Added Value Assessment, 2020.

[51] See introduction above for references.

[52] For additional details, see the European Commission *Report on safety and liability implications of AI, the Internet of Things and Robotics*, 2020 and the Report on *Liability for Artificial Intelligence and other emerging technologies*.

framework (the current initiative) and then**,** the EU rules to address **liability issues** related to new technologies, including AI systems (expected Q4 2021 – Q1 2022).[53] The future changes to the liability rules will take into account the elements of the horizontal framework with a view to designing the most effective and proportionate solutions with regard to liability. Moreover, compliance with the requirements of the AI horizontal framework will be taken into account for assessing liability of actors under future liability rules.[54]

With regard to intermediary liability, for example when sellers place faulty products through online marketplaces, the E-Commerce Directive regulates the liability exemptions for online intermediaries. This framework is currently updated in the Commission's proposal for a Digital Services Act.

1.3.4. Relevant legislation on services

The EU also has a comprehensive **legal framework on services** that is applicable whenever AI software is provided as a stand-alone service or integrated into other services, including in particular digital services, audio-visual media services, financial services, transport services, professional services and others. In the context of information society services, the e-Commerce Directive provides for an applicable regulatory framework laying down horizontal rules for provisions of such services in the Union. The proposed Digital Services Act includes rules on liability exemption for providers of intermediary services (i.e. mere conduit; caching; hosting), which are to date contained in the e-Commerce Directive that remains largely unchanged and fully applicable. At the same time, the Digital Services Act introduces due diligence obligations for providers of intermediary services so as to keep users safe from illegal goods, content or services and to protect their fundamental rights online.[55] These due diligence obligations are adapted to the type and nature of the intermediary service concerned and transparency and accountability rules will apply for algorithmic systems, including those based on AI, used by online platforms.

In the field of financial services, the risk governance requirements under the existing legislation provide a strong regulatory and supervisory framework for assessment and management of risks. Specific rules additionally apply in relation to trading algorithms.[56] With respect to creditworthiness assessment, European Banking Authority guidelines[57] have been recently adopted to improve regulated financial institutions' practices and associated governance arrangements, processes and mechanisms in relation to credit granting, management and monitoring, including when using automated models in the creditworthiness assessment and credit decision-making processes.[58] However, no sector-specific EU guidance or rules currently apply to non-regulated entities when assessing the creditworthiness of consumers.

1.4. Political context

To address the opportunities and challenges of AI, in April 2018 the **European Commission** put forward a European approach to AI in its Communication "Artificial Intelligence for Europe."[59] In

---

[53] See section 8 for a more detailed analysis.

[54] The relevant recital provision to this extent would be included in the proposed horizontal framework initiative.

[55] European Commission, *Digital Services Act – deepening the internal market and clarifying responsibilities for digital services*, 2020.

[56] Commission Delegated Regulation (EU) 2017/589 of 19 July 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council with regard to regulatory technical standards specifying the organisational requirements of investment firms engaged in algorithmic trading.

[57] European Banking Authority, *Guidelines on loan origination and monitoring*, 2020.

[58] See, also mapping of national approaches in relation to creditworthiness assessment under Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC. As set out in the recently adopted digital finance strategy, the Commission will invite the European Supervisory Authorities and the European Central Bank to develop guidance on the use of AI applications in 2021.

[59] European Commission, *Artificial Intelligence for Europe*, COM(2018) 327 final, 2018.

June 2018, the Commission appointed the High-Level Expert Group on Artificial Intelligence,[60] which produced two deliverables: the Ethics guidelines for trustworthy AI[61] and the Policy and investment recommendations for trustworthy AI.[62] In December 2018, the Commission presented a Coordinated Plan on AI[63] with Member States to foster the development and use of AI.[64]

In June 2019, in its Communication "Building Trust in Human-Centric Artificial Intelligence",[65] the Commission endorsed the seven key requirements for Trustworthy AI identified by the HLEG. After extensive consultation, on 17 July 2020, the HLEG published an Assessment List for Trustworthy Artificial Intelligence for self-assessment (ALTAI)[66] which was tested by over 350 organisations.

In February 2020, the European Commission published a White Paper on AI[67] setting out policy options for a regulatory and investment oriented approach. It was accompanied by a Commission Report on the safety and liability implications of AI.[68] The White Paper opened a wide public consultation where more than 1 215 contributions were received from a wide variety of stakeholders, including representatives from industry, academia, public authorities, international organisations, standardisation bodies, civil society organisations and citizens. This clearly showed the great interest from stakeholders around the globe in shaping the future EU regulatory approach to AI, as assessed in this impact assessment.

The European Council and the European Parliament (EP) also repeatedly called for the Commission to take legislative action to ensure a well-functioning internal market for AI systems where both benefits and risks are adequately addressed at EU level.

In 2017, the **European Council** called for a 'sense of urgency to address emerging trends' including 'issues such as artificial intelligence …, while at the same time ensuring a high level of data protection, digital rights and ethical standards'.[69] In its 2019 Conclusions on the Coordinated Plan on the development and use of artificial intelligence Made in Europe,[70] the Council further highlighted the importance of ensuring that European citizen's rights are fully respected and called for a review of the existing relevant legislation to make it fit for purpose for the new opportunities and challenges raised by AI. The European Council has also called for a clear determination of what should be considered as high-risk AI applications.[71]

The most recent Conclusions from 21 October 2020 further called for addressing the opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behaviour of certain

---

[60] European Commission, *High-Level Expert Group on Artificial Intelligence*, 2020.

[61] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019.

[62] High-Level Expert Group on Artificial Intelligence, *Policy and investment recommendations for trustworthy AI*, 2019.

[63] European Commission, *Coordinated Plan on Artificial Intelligence*, 2018.

[64] The Plan builds on a Declaration Cooperation on Artificial Intelligence, signed by EU Member States and Norway in April 2018. An AI Member States group has been regularly meeting since 2018, discussing among other things the ethical and regulatory aspects of AI. A review of the Coordinated plan is foreseen in 2021.

[65] European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Building Trust in Human Centric Artificial Intelligence,* COM(2019)168 final, 2019.

[66] High-Level Expert Group on Artificial Intelligence, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, 2020.

[67] European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, 2020.

[68] European Commission, *Staff Working Document on Liability for emerging digital technologies*, SWD 2018/137 final.

[69] European Council, *European Council meeting (19 October 2017) – Conclusion* EUCO 14/17, 2017, p. 8.

[70] Council of the European Union, *Artificial intelligence b) Conclusions on the coordinated plan on artificial intelligence-Adoption* 6177/19, 2019.

[71] European Council, *Special meeting of the European Council (1and 2 October 2020) – Conclusions* EUCO 13/20, 2020.

AI systems, to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules.[72]

In 2017, the **European Parliament** (EP) adopted a Resolution on Civil Law Rules on Robotics urging the Commission to analyse the impact of use of AI technologies in the main areas of EU legislative concern, including ethics, liability, standardisation and institutional coordination and oversight, and to adopt legislation where necessary.[73] In 2019, the EP adopted a Resolution on a Comprehensive European Industrial Policy on Artificial Intelligence and Robotics.[74] In June 2020, the EP also set up a Special Committee on Artificial Intelligence in a Digital Age (AIDA) tasked to analyse the future impact of AI systems in the digital age on the EU economy and orient future EU priorities.[75]

In October 2020, the EP adopted a number of resolutions related to AI, including on ethics,[76] liability[77] and copyright.[78] EP resolutions on AI in criminal matters[79] and AI in education, culture and the audio-visual sector[80] are forthcoming. The EP Resolution on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies, specifically recommends to the Commission to propose a legislative action to harness the opportunities and benefits of AI, but also to ensure protection of ethical principles.[81] The EP resolution on internal market aspects of the Digital Services Act presents the challenges identified by the EP as regards AI-driven services. [82]

At international level, the ramifications in the use of AI systems and related challenges have also received significant attention.[83] The **Council of Europe** started work on an international legal framework for the development, design and application of AI, based on the Council of Europe's standards on human rights, democracy and rule of law. It has also recently issued guidelines and proposed safeguards and certain prohibitions of the use of facial recognition technology considered particularly intrusive and interfering with human rights.[84] The **OECD** adopted a Council Recommendation on Artificial Intelligence.[85] The **G20** adopted human-centred AI Principles that draw on the OECD AI Principles.[86] **UNESCO** is also starting to develop a global standard setting instrument on AI.[87] Furthermore, the EU, together with many advanced economies, set up the

---

[72] Council of the European Union, *Presidency conclusions - The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change*, 11481/20, 2020.

[73] European Parliament resolution of 16 February 2017 on Civil Law Rules on Robotics, 2015/2103(INL).

[74] European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics, 2018/2088(INI).

[75] European Parliament decision of 18 June 2020 on setting up a special committee on artificial intelligence in a digital age, and defining its responsibilities, numerical strength and term of office, 2020/2684(RSO).

[76] European Parliament resolution of 20 October 2020 on a framework of ethical aspects of artificial intelligence, robotics and related technologies, 2020/2012(INL).

[77] European Parliament resolution of 20 October 2020 on a civil liability regime for artificial intelligence, 2020/2014(INL).

[78] European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI).

[79] European Parliament Draft Report, Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, 2020/2016(INI).

[80] European Parliament Draft Report, Artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI).

[81] More details of the EP proposals are presented in section 5 when various policy options are discussed.

[82] European Parliament resolution of 20 October 2020 on the Digital Services Act: Improving the functioning of the Single Market (2020/2018(INL)).

[83] See for an overview: Fundamental Rights Agency, *AI Policy Initiatives 2016-2020*, 2020; or Council of Europe, *Artificial Intelligence*, 2020.

[84] Consultative Committee of The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data Convention 108 Guidelines on Facial Recognition, 28 January 2021, T-PD(2020)03rev4.

[85] OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 2019.

[86] G20, *G20 Ministerial Statement on Trade and Digital Economy*, 2019.

[87] UNESCO, *Artificial intelligence with human values for sustainable development*, 2020.

Global Partnership on Artificial Intelligence (GPAI).[88] As part of the EU-Japan Partnership for Sustainable Connectivity and Quality Infrastructure, concluded in September 2019, both sides reconfirmed their intention to continue promoting policies that boost innovation including in Artificial Intelligence, cloud, quantum computing and blockchain.

In addition to EU and international initiatives, many countries around the world started to consider adopting their own ethical and accountability frameworks on AI and/or automated decision-making systems. In **Canada**, a Directive on Automated Decision-Making came into effect on April 1, 2020 and applies to the use of automated decision systems in the public sector that "provide external services and recommendations about a particular client, or whether an application should be approved or denied." The Directive includes an Algorithmic Impact Assessment and transparency obligations vis-à-vis persons affected by the automated decision. In 2020, the Government of **New Zealand**, together with the World Economic Forum, was spearheading a multi-stakeholder, policy project, structured around three focus areas: 1) obtaining of a social licence for the use of AI through an inclusive national conversation; 2) the development of in-house understanding of AI to produce well-informed policies; and 3) the effective mitigation of risks associated with AI systems to maximize their benefits. In early 2020, the **United States'** government adopted overall regulatory principles. On this basis, the White House released the first-ever guidance for Federal agencies on the regulation of artificial intelligence applications in the public sector that should comply with key principles for Trustworthy AI.[89] Other countries with regulatory initiatives on AI include, for example, **Singapore**, **Japan**, **Australia**, the **United Kingdom** and **China**.[90] Annex 5.1 summarises the main ongoing initiatives in these third countries undertaken to address the challenges posed by AI and to harness its potential for good.

1.5. Scope of the impact assessment

This report assesses the case for an EU regulatory framework for the development and use of AI systems and examines the impact of different policy options. The use of AI for exclusive military purposes remains outside the scope of the present initiative due to its implications for the Common Foreign and Security Policy (CFSP).[91] Insofar as 'dual use' products and technologies[92] have AI features and can be used for both military and civil purposes, these goods will fall into the scope of the current initiative on AI.

The forthcoming initiative on liability and the ongoing revisions of sectoral safety legislation are subject to separate initiatives and remain equally outside the scope of this impact assessment, as discussed in section 1.3 above.

---

[88] The EU is one of the founding members, alongside Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, the United States of America. The aim of this initiative is to bring together leading experts from industry, civil society, governments, and academia to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities.

[89] See the most recent Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government from 3 December 2020, stipulates that when designing, developing, acquiring, and using AI in the Federal Government, agencies shall adhere to the following Principles: (a) Lawful and respectful of our Nation's values; (b) Purposeful and performance-driven; (c) Accurate, reliable, and effective; (d) Safe, secure, and resilient; (e) Understandable; (f) Responsible and traceable; (g) Regularly monitored; (h) Transparent; (i) Accountable.

[90] Fjeld, J., N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center Research Publication No. 2020-1, 2020.

[91] The Common Foreign and Security Policy is regulated under Title V of the Treaty on European Union, which would be applicable for the use of AI for such exclusive military purposes.

[92] Modernized rules for the export control of such dual use products and technologies were agreed by the EU co-legislators in November based on the Commission's Proposal for a Regulation setting up a Union regime for the control of exports, transfer, brokering, technical assistance and transit of dual-use items (recast), COM(2016) 616 final. 2016/0295 (COD).

2. PROBLEM DEFINITION

2.1. What are the problems?

The analysis of the available evidence[93] suggests that there are six main related problems triggered by the development and use of AI systems that the current initiative aims to address.

Table 2: Main problems

MAIN PROBLEMS	STAKEHOLDERS CONCERNED
1. Use of AI poses increased risks to safety and security of citizens	Citizens, consumers and other victims Affected businesses
2. Use of AI poses increased risk of violations of citizens' fundamental rights and Union values	Citizens, consumers and other victims Whole groups of the society, Users of AI systems liable for fundamental rights violations
3. Authorities do not have powers, procedural frameworks and resources to ensure and monitor compliance of AI development and use with applicable rules	National authorities responsible for compliance with safety and fundamental rights rules
4. Legal uncertainty and complexity on how existing rules apply to AI systems dissuade businesses from developing and using AI systems	Businesses and other providers developing AI systems Businesses and other users using AI systems
5. Mistrust in AI would slow down AI development in Europe and reduce the global competitiveness of the EU economy	Businesses and other users using AI systems Citizens using AI systems or being affected by them
6. Fragmented measures create obstacles for cross-border AI single market and threaten Union's digital sovereignty	Businesses developing AI, mainly SMEs affected Users of AI system, including consumers, businesses and public authorities

Problem 1: The use of AI poses increased risks to safety and security of citizens

The overall EU architecture of safety frameworks is based on a combination of horizontal and sectoral rules.[94] This includes the horizontal GPSD and sector-specific legislation, as for example, the Machinery Directive. The EU safety legislation has significantly contributed to the high-level of safety of products put into circulation in the EU Single Market. However, it is increasingly confronted with the challenges posed by new technologies, some of which specifically relate to AI technologies.[95]

Two main reasons explain the limitations of the existing EU safety and security framework in relation to the application to AI technologies.

[93] The analysed evidence includes results of the public consultation on White Paper on AI, responses to the inception impact assessment, stakeholder consultations carried out within the framework of this impact assessment, European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies, European Council Presidency Conclusion of 21 October 2020, ongoing work of international organizations, as well as secondary literature.

[94] For more details about the existing product safety legislation see section legal context 1.3.2.

[95] In this respect the Commission's White Paper on AI was accompanied by *Commission Report on safety and liability implications of AI, the Internet of Things and Robotics*, 2020. Both the White Paper and the Report point to the combination of revision of existing EU safety legislation and the horizontal framework on AI to address those risks as explained in the section on the legal context 1.3.2. and 1.3.3.

Firstly, the **nature of safety risks caused by AI**: The main causes of safety risks generated by complex software or algorithms are qualitatively different from risks caused by physical products.[96] The specific characteristics of AI systems,[97] could lead to safety risks including:[98]

- **Biases in data or models:** training data can have hidden biases that can be 'learnt' by the AI model and reproduced in its outputs. AI algorithms can also introduce biases in their reasoning mechanisms by preferring certain characteristics of the data. For example, a medical imaging recognition device trained on data biased towards a specific segment of population may lead in certain cases to safety risks due to misdiagnosis in patients not well represented by the data.
- **Edge cases:** unexpected or confusing input data can result in failures due to the limited ability that AI models might exhibit to generalise well from training data. For example, image recognition systems in an autonomous vehicle could malfunction in the case of unexpected road situations, endangering passengers and pedestrians.
- **Negative side effects:** the realization of a task by an autonomous AI system could lead to harmful effects if the scope of action of the system is not correctly defined and does not consider the context of use and the state of the environment. For instance, an industrial robot with an AI system designed to maximize the work rate in a workshop could damage property or accidentally hurt people in situations not foreseen in its design.

AI models which are run on top of ICT infrastructure are composed of a diverse range of digital assets. Therefore, cybersecurity issues affecting this broader digital ecosystem also extend to AI systems[99] and can result in important AI safety risks. This is particularly relevant considering the new systems enabled by AI, such as cyber-physical systems, where cybersecurity risks have direct safety implications.[100]

Moreover, AI systems can also be subject to malicious attempts to exploit AI specific vulnerabilities, including:[101]

- **Evasion –** an attack when an attacker modifies input data, sometimes in an imperceptible manner, so that the AI model cannot correctly identify the input and this leads to wrong outputs. Examples include attacks to evade anti-spam filters, spoofing attacks against biometric verification systems or stickers added to stop signs to make autonomous vehicles to perceive them as speed signs.
- **Data poisoning –** an action aiming to modify the behaviour of the AI model by altering the training datasets, especially when the data used is scraped from the web, sourced from data exchanges or from open datasets. The 'learning' systems where model parameters are

[96] Those risks very much pertain to the quality and reliability of information which results from the output of a computing operation. Qualitatively different means that the nature (the cause/ driver) of safety risks generated by complex software or algorithms are different from risks caused by physical products.

[97] For explanation of AI Characteristics please see section 2.2. 'Drivers' below and Annex 5.2.: Five specific characteristics of AI.

[98] This refers primarily to the ill-designed systems, see Russell, Stuart J., Peter Norvig. Artificial Intelligence: A Modern Approach. Pearson, 4th ed., 2020.

[99] The "AI Cybersecurity Challenges: Threat landscape for Artificial Intelligence" report published by ENISA with the support to the Ad-Hoc Working Group of Artificial Intelligence Cybersecurity presents a mapping of the AI cybersecurity ecosystem and its Threat Landscape, highlighting the importance of cybersecurity for secure and trustworthy AI.

[100] ENISA, JRC, Cybersecurity challenges in the uptake of Artificial Intelligence in Autonomous Driving, 2021.

[101] These vulnerabilities and risks in addition to their implications on safety, could also lead in certain scenarios to the situations that could have significant negative impacts on fundamental rights and Union values (see Problem 2), in particular when AI systems are used to make decisions based on personal data.

constantly updated using new data, are particularly sensitive to data poisoning.[102] For example, poisoning the data used to train a chatbot could make it disclose sensitive information or adopt inappropriate behaviour.

- **Model extraction –** attacks that aim to build a surrogate system that imitates the targeted AI model. The goal is to get access to a representation of the AI model that will allow the attacker to understand and mimic the logic of the system and possibility build more sophisticated attacks, like evasion or data poisoning or steal sensitive information from training data.
- **Backdoor –** refers to a typical risk in programming which is not limited to AI applications, but is more difficult to detect and avoid in the case of AI due to its opacity.[103] The presence of a so-called 'backdoor'[104] makes unauthorised access to a system possible.

AI specific risks are not or are only partly covered by the current Union safety and security legislation. While Union safety legislation covers more generally the risks stemming from software being a safety component (and usually embedded) in a hardware product, stand-alone software (except in the medical device framework) – including when used in services – or software uploaded to a hardware device after this device is placed on the market are not currently covered.[105] Thus, services based on AI technology, such as transport services or infrastructure management, are not covered.

Moreover, even when software is covered by EU safety legislation, **no specific safety or performance requirements are set for AI systems**. For example, there are no AI-technology specific requirements ensuring reliability and safety over its lifecycle. The EU legal framework on cybersecurity applies to ICT products, services and processes, and therefore could also potentially cover AI technologies.[106] However, no scheme for AI currently exists and there are no established AI cybersecurity standards of best practice for developers in the design phase, notably when it comes to 'security by design' for AI. Moreover, the certification schemes for cybersecurity are of a voluntary nature.[107]

This lack of clear safety provisions covering specific AI risks, both for AI systems being safety components of products and AI systems used in services, **can be highly problematic for users and consumers of AI applications**.

Secondly, the lifecycle of an AI product: Under the current legal framework, ex-ante conformity assessment procedures are mainly conceptualized for products that are 'stable' in time after deployment. The current safety legislation does not contain specific provisions for products that are

---

[102] The use of 'Evasion', 'data poisoning' attacks or task misspecification may also have an objective to misdirect reinforcement learning behaviour.

[103] For the definition of a term, please see Annex 5.2.: Five specific characteristics of AI.

[104] A backdoor usually refers to any method by which authorized and unauthorized users are able to get around normal security measures and gain high level user access on a computer system, network, or software application.

[105] The ongoing review of certain sectorial legislations is considering these aspects (e.g. Machinery Directive and General Product Safety Directive).

[106] Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). The cybersecurity framework allows the development of dedicated certification schemes. Each scheme establishes and lists the relevant standards, however, any certification scheme established under this Regulation is of a voluntary nature.

[107] However, product specific legislation already exists in some sector: e.g. in the automotive sector new rules on automated vehicles, cybersecurity and software updates of vehicles will become applicable as part of the vehicle type approval and market surveillance legislation as from 7 July 2022, providing notably for obligations for the manufacturer to perform an exhaustive risk assessment (including risks linked to the use of AI) and to put in place appropriate risk mitigations, as well as to implement a comprehensive risk management system during the lifecycle of the product.

possibly subject to evolution during their lifecycle. Yet, certain AI systems are subject to considerable change after their first deployment on the market.

Such potential and unanticipated modifications to the performance of those AI systems after their placement on the market could still occur and cause injuries and physical damage.[108]A product that is already on the market is normally required to undergo a new conformity assessment if there are substantial modifications of the product. However, the exact interpretation of the notion of "substantial modification" in an AI context needs to be clearly defined, also in light of the complexity of the AI value chain,[109] lest it should lead to legal uncertainty.

As a general conclusion, the specificities of AI applications might require the establishment of some specific safety requirements to ensure a high level of protection of users and consumers as well as to provide legal certainty for businesses, notably for use cases where the lack of proper performance or reliability of AI could have a severe impact on life and health of individuals. This is regardless of whether AI systems are a safety component of products or are used in services.

Stakeholders views: In the Public consultation on White Paper on AI, 83% of all respondents consider that the fact that AI may endanger safety' is 'important' (28%) or 'very important' (55%). Among SMEs, 72% found safety to be an important or very important concern, whereas only 12% said it was not important or not important at all. This position was even more pronounced among large businesses, with 83% saying that safety was (very) important and only 4% finding the issue unimportant. 80% of academic and other research institutions and 88% of civil society organisations agreed that safety was a (very) important concern. Among EU citizens, 73% found safety to be an important or very important issue. Of those stakeholders who said safety was not a (very) important concern, 43% were EU citizens (which make up 35% of all respondents) and 20% were SMEs (7%).

Problem 2: Use of AI poses an increased risk of citizens' fundamental rights and Union values violations

The use of AI can have a significant impact on virtually all fundamental rights as enshrined in the EU Charter of Fundamental Rights. AI use can be positive, promoting certain rights (e.g. the right to education, health care) and contributing to important public interests (e.g. public security, public health, protection of financial interests). It can also help reduce some adverse impacts on fundamental rights by improving the accuracy or efficiency of decision-making processes and addressing biases, delays or errors in individual human decisions.[110] On the other hand, AI used to replace or support human decision-making or for other activities such as surveillance may also infringe upon individual's rights.[111] This is not a flaw of the technology *per se*, but the responsibility of the humans who are designing and using it and who must ensure these violations do not happen in the first place.

If breaches of fundamental rights do happen, these can also be very difficult to detect and prove, especially when the system is not transparent. This challenges the effective enforcement of the existing EU legislation aimed at safeguarding fundamental rights, as listed in section 1.3.

[108] The current legal framework, does not provide conditions when self-learning AI should undergo a new conformity assessment.

[109] For example, developers, installation/operation/maintenance service providers at the point of use, actors responsible for the operation and maintenance of networks/platforms.

[110] To this end, the fundamental rights of all persons concerned must be looked at and all remedies and safeguards applicable to an AI systems considered. The potential positive or adverse impact on the society as a whole and on general public interests such as public security, fight against crime, good administration, public health, protection of public finances should also be taken into account.

[111] EU Fundamental Rights Agency, *Getting the future right – Artificial intelligence and fundamental rights*, 2020. Raso, F. et al., *Artificial Intelligence & Human Rights: Opportunities & Risks*, Berkman Klein Center for Internet & Society Research Publication, 2018.

While it remains difficult to quantify the real magnitude of the risks to fundamental rights, a growing amount of evidence[112] suggests that Union citizens might be affected in an increasingly wide range. Moreover, a growing body of case law and legal challenges to the use of AI breaching fundamental rights is also emerging across different Member States.[113] The following sections will focus on some of the most prominent fundamental rights risks.[114]

2.1.1. Use of AI may violate human dignity and personal autonomy

The right to human dignity is an inviolable right that requires every individual to be treated with respect as a human being and not as a mere 'object' and their personal autonomy respected. Depending on the circumstances of the case and the nature of the interaction, this might be challenging if people are **misled in believing that they are interacting with another person when they are actually interacting with an AI system**.

Moreover, AI is often used to **sort and classify traits and characteristics that emanate from datasets that are not based on the individual concerned**. Organisations that use such data to determine individual's status as a victim (e.g. women at risk of domestic violence)[115] or individual's likelihood to reoffend in case of predictive risk assessments could violate individuals' right to human dignity since the assessment is no longer based on the personal individual situation and merits.

AI can also be used for **manipulation** that can be particularly harmful for certain users. While psychological science shows that these problems are not new, the growing manipulative capabilities of algorithms that collect and can predict very sensitive and privacy intrusive personal information can make people extremely vulnerable, easily deceived or hyper-nudged towards specific decisions that do not align with their goals or run counter to their interests.[116] Evidence suggests that AI supported products or services (toys, personal assistants etc.) can be intentionally designed or used in ways that appeal to the subliminal perception of individuals, thus causing them to take decisions that are beyond their cognitive capacities.[117] Even if the techniques used are not subliminal, for certain categories of vulnerable subjects, in particular children, these might have the same adverse manipulative effects if their mental infirmity, age or credulity are exploited in harmful ways.[118] As the AI application areas develop, these (mis)uses and risks will likely increase.

---

[112] Reports and case studies published among others by research and civil society organisations such as AlgorithmWatch and Bertelsmann Stiftung, *Automating Society – Taking Stock of Automated Decision-Making in the EU*, 2019. AlgorithmWatch and Bertelsmann Stiftung, *Report Automating Society*, 2020. EDRi, *Use cases: Impermissible AI and fundamental rights breaches*, 2020.

[113] See e.g. *Decision 216/2017* of the National Non-Discrimination and Equality Tribunal of Finland of 21 March 2017. The Joint Council for the Welfare of Immigrants, *We won! Home Office to stop using racist visa algorithm*, 2020. The decision of the Hague District Court regarding the use of the SyRi scheme by the Dutch authorities etc.

[114] Broader considerations and analysis of the impact on all fundamental rights can be found in the study supporting this impact assessment as well as studies on AI and human rights, for example, commissioned by the EU Fundamental Rights Agency and the Council of Europe.

[115] For example the VioGén protocol in Spain which includes an algorithm that evaluates the risk that victims of domestic violence are going to be attacked again by their partners or ex-partners. See AlgorithmWatch and Bertelsmann Stiftung, *Report Automating Society*, 2020.

[116] See Council of Europe, Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, February 2019;

[117] U.S. White House Office of Science and Technology Policy, Request for Information on the Future of Artificial Intelligence, September 1, 2016; Maurice E. Stucke Ariel Ezrachi, 'The Subtle Ways Your Digital Assistant Might Manipulate You', Wired, 2016; Judith Shulevitz, 'Alexa, Should We Trust You?, The Atlantic, November 2018.

[118] Anna-Lisa Vollmer, Children conform, adults resist: A robot group induced peer pressure on normative social conformity, Science Robotics, Vol. 3, Issue 21, 15 Aug 2018; Hasse, A., Cortesi, S. Lombana Bermudez, A. and Gasser, U. (2019). 'Youth and Artificial Intelligence: Where We Stand', Berkman Klein Center for Internet & Society at Harvard University; UNICEF, 'Safeguarding Girls and Boys: When Chatbots Answer Their Private Questions', 6 August 2020.

2.1.2. Use of AI may reduce privacy protection and violate the right to data protection

The EU has a strong and modern legal framework on data protection with the Law Enforcement Directive and the General Data Protection Regulation recently evaluated as fit for purpose.[119] Still, the use of AI systems might challenge the effective protection of individuals since the right to private life and other fundamental rights violations can occur even if non-personal, including anonymized data is processed.

The arbitrary use of algorithmic tools gives unprecedented opportunities for **indiscriminate or mass surveillance, profiling and scoring of citizens** and significant intrusion into people's privacy and other fundamental rights. Beyond affecting the individuals concerned, such use of technology has also an impact on society as a whole and on broader Union values such as democracy, freedom, rule of law, etc.[120]

A particularly sensitive case is the increasing use of **remote biometric identification systems** in publicly accessible spaces.[121] Currently, the most advanced variety of this family of applications is facial recognition, but other varieties exist, such as gait recognition or voice recognition. Wherever such a system is in operation, the whereabouts of persons included in the reference-database can be followed, thus impacting their personal data, privacy, autonomy and dignity. Moreover, freedom of expression, association and assembly might be undermined by the use of the technology resulting in a chilling effect on democracy. On the other hand, the use of such systems has been considered by some justified in limited cases when strictly necessary and proportionate for safeguarding important public interests of public security.[122] Public and private operators are already using such systems in Europe,[123] but because of privacy and other fundamental rights violations, their operation has been blocked by data protection authorities in schools or in other publicly accessible spaces.[124] Despite these serious concerns and potential legal challenges, many countries consider using biometric identification systems at a much larger scale to cope with increasing security risks.[125]

Apart from identification, facial, gait, iris or voice recognition technology is also used to attempt to predict individual's characteristics (e.g. sex, race or even sexual orientation), emotions and to detect

---

[119] See Communication from the Commission, *Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the General Data Protection Regulation* COM(2020) 264 final, 2020.

[120] *EDPS Opinion on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence and trust*, 2020.

[121] AlgorithmWatch and Bertelsmann Stiftung, *Report Automating Society*, 2019 and 2020.

[122] In their submissions to the public consultation on the White paper, some countries (e.g. France, Finland, the Check republic, Denmark) submit that the use of remote biometric identification systems in public spaces might be justified for important public security reasons under strict legal conditions and safeguards.

[123] For example, the Italian ministry of interior plans to employ the SARI facial recognition in Italy Cameras with facial recognition technology have also been used in a train station in Madrid or in the bus terminal. See also EU Fundamental Rights Agency, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, 2019.

[124] See e.g. EDPB, *Facial recognition in school renders Sweden's first GDPR fine*, 2019. Politico, *French privacy watchdog says facial recognition trial in high schools is illegal*, 2019. In the UK, the Court of Appeal found that the facial recognition programme used by the South Wales police was unlawful and that '[i]t is not clear who can be placed on the watch list, nor is it clear that there are any criteria for determining where [the facial recognition technology] can be deployed' (UK, *Court of Appeal, R (Bridges) v. CC South Wales*, EWCA Civ 1058, 11 August 2020).

[125] Germany put on hold its plans to use facial recognition at 134 railway stations and 14 airports, while France plans to establish a legal framework permitting video surveillance systems to be embedded with facial recognition (Stolton, S., '*After Clearview AI scandal, Commission 'in close contact' with EU data authorities'*, *Euroactiv,* 2020). In 2019, the Hellenic Police signed a €4 million contract with Intracom Telecom for a smart policing project (Homo Digitalis, *The Greek DPA investigates the Greek Police*, 2020). Italy also considers using facial recognition in all football stadiums (Chiusi, F., *In Italy, an appetite for face recognition in football stadiums*, 2020).

whether people are lying or telling the truth.[126] **Biometrics for categorisation and emotion recognition** might lead to serious infringements of peoples' privacy and their right to the protection of personal data as well as to their manipulation. In addition, there are serious doubts as to the scientific nature and reliability of such systems.[127]

While EU data protection rules in principle prohibit the processing of biometric data for the purpose of uniquely identifying a natural person except under specific conditions permitted by law,[128] the White Paper on AI opened a discussion on the specific circumstances, if any, which might justify such use, and on common safeguards.

Stakeholders views: In a recent survey, between 45% and 60% of consumers believed that AI will lead to more abuse of personal data (BEUC, Consumers see potential of artificial intelligence but raise serious concerns, 2020). 76.7 % of respondents to the White Paper on AI consider that the systems for remote biometric identification in public spaces have to be regulated in one way or another, 28.1% consider that they should never be authorized at publicly accessible spaces. Recently, 12 NGOs also started an EU-wide campaign called 'Reclaim Your Face' to urge EU to ban facial recognition in public spaces. The Commission has also registered a European Citizens' Initiative entitled 'Civil society initiative for a ban on biometric mass surveillance practices'.

2.1.3. Use of AI may lead to discriminatory outcomes

Algorithmic discrimination can occur for several reasons at many stages and it is often difficult to detect and mitigate.[129] Problems may arise due to flawed design and developers who unconsciously embed their own biases and stereotypes when making the classification choices. Users might also misinterpret the AI output in concrete situations or use it in a way that is not fit for the intended purpose. Moreover, bias causes specific concerns for **AI techniques dependent on data,** which might be unrepresentative, incomplete or contain historical biases that can cement existing injustices with the 'stamp' of what appears to be scientific and evidence-based legitimacy.[130] Developers or users could also intentionally or unintentionally use proxies that correlate with protected characteristics under EU non-discrimination legislation such as race, sex, disability etc. Although being based on seemingly neutral criteria, this may disproportionately affect certain protected groups giving rise to indirect discrimination (e.g., using proxies such as postal codes to account for ethnicity and race).[131] As explained in the driver section 2.2., the algorithms can also introduce themselves biases in their reasoning mechanisms by favouring certain characteristics of the data on which they have been trained. Varying levels of accuracy in the performance of AI systems may also disproportionately affect certain groups, for example facial recognition systems that detect gender well for white men, but not for black women[132] or that do not detect as person those using wheelchairs.

The use of discriminatory AI systems notably in sectors such as employment, public administration, judiciary or law enforcement, might also violate many other fundamental rights (e.g. right to education, social security and social assistance, good administration etc.) and lead to broader

[126] This was researched at selected EU external borders (Greece, Hungary and Latvia) in the framework of the Integrated Portable Control System (iBorderCtrl) project, which integrates facial recognition and other technologies to detect if a person is saying the truth.

[127] Vincent, J., *AI 'emotion recognition' can't be trusted*, The Verge, 2019.

[128] See Article 9(2) of the GDPR and Article 10 of the Law Enforcement Directive.

[129] Fundamental Rights Agency, *#BigData: Discrimination in data-supported decision-making*, 2018, p. 3. Despite the new risks posed by AI to the right to non-discrimination, the FRA report also highlights that human-decision-making is similarly prone to bias and if AI systems are properly designed and used, they offer opportunities to limit discriminatory treatment based on biased human decisions.

[130] Bakke, E., 'Predictive policing: The argument for public transparency', *New York University Annual Survey of American Law*, 2018, pp. 139-140.

[131] Postal codes were used, for example, in the Amsterdam risk assessment tool ProKid (now discontinued) to assess the risk of recidivism – future criminality – of children and young people, even if postal codes are often proxies for ethic origin as ruled by the CJEU, Case C-83/14.

[132] The Gender Shades project evaluates the accuracy of AI powered gender classification products.

societal consequences, **reinforcing existing or creating new forms of structural discrimination and exclusion.**

For example, evidence suggests that in the **employment sector** AI is playing an increasingly key role in making hiring decisions mainly facilitated by intermediary tech service providers.[133] This can negatively affect potential candidates in terms of discriminatory filtering at different moments of recruitment procedures or afterwards.[134] Another problematic area is the **administration of social welfare assistance,** with some recent cases of suspected discriminatory profiling of unemployed people in Denmark, Poland or Austria.[135] Financial institutions and other organisations might also use AI for **assessing individual's creditworthiness** to support decisions determining the access to credit and other services such as housing. While this can increase the opportunities for some people to get access to credit on the basis of more diverse data points, there is also a risk that systems for assessing scores might unintentionally induce biases, if not properly designed and validated.[136] In **law enforcement and criminal justice**, AI models trained with past data can be used to forecast trends in the development of criminality in certain geographic areas, to identify potential victims of criminal offences such as domestic violence or to assess the threats posed by individuals to commit offences based upon their criminal records and overall behaviour. In the EU, cases of these predictive policing systems exist in a number of Member States.[137] At the borders, specific groups such as migrants and asylum seekers can also have their rights significantly affected if discriminatory AI systems are used by public authorities.[138]

Under the existing EU and national anti-discrimination law, it could be very difficult to launch a complaint as the affected person most likely do not know that an AI systems is used and even if they do, they are not aware how it functions and how its outputs are applied in practice. This makes it very difficult, if not impossible, for the persons concerned to establish the facts needed to establish *prima facie* discrimination, or prove it. It might also be very challenging for supervisory authorities and courts to detect and assess discrimination, in particular in cases when there is no readily available and relevant statistical evidence.[139]

---

[133] Research suggests that hiring platforms such as PeopleStrong or TribePad, HireVue, Pymetrics and Applied use such kind of algorithmic tools for supporting recruitment decisions, see Sánchez-Monedero et al., *What does It mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems*, 2020. See also VZBV, *Artificial Intelligence: Trust Is Good, Control Is Better*, 2018.

[134] Algorithms used to serve ads were found to generally prefer men over women for high-paying jobs. See Upturn, *An Examination of Hiring Algorithms, Equity, Bias*, 2018.

[135] For example, the Dutch SyRI system used to identify the risk of abusing the social welfare state, was recently found by the court to be intransparent and unduly interfering with the rights to private life of all affected persons. AI systems for social welfare also exist in Finland, Germany, Estonia and other countries, see AlgorithmWatch and Bertelsmann Stiftung, 2020.

[136] For example, in Germany the leading company for scoring individuals, SCHUFA run an AI system that was found by researchers and civil society to suffer from various anomalies in the data. In 2018, the Finnish National Non-Discrimination and Equality Tribunal prohibited a financial company, specialising in credits, from using certain statistical methods in credit scoring decisions. For more cases see also AlgorithmWatch and Bertelsmann Stiftung, 2019 and 2020.

[137] E.g. the Dutch ProKid (now discontinued) and the e-Criminality Awareness System; Precobs, Krim and SKALA in Germany; KeyCrime and eSecurity in Italy, Pred-Crime in Spain. For more cases see also AlgorithmWatch and Bertelsmann Stiftung, 2019 and 2020.

[138] For example, the UK Home Office stopped using an algorithm streaming visa applicants, because of allegations for unlawful discrimination against people of certain nationalities. Also in the UK, a speech recognition system used to detect fraud among those sitting English language exams in order to fulfil student visa requirements, reportedly resulted in the wrongful deportation of up to 7,000 people. The Algorithm Watch and Bertelsmann Stiftung, 2020 mentions several other AI tools and pilot projects in EU Member States where AI is used in the context of migration, border control and asylum procedures, e.g. p, 26, 85, 115 and 199.

[139] Wachter, S., B. Mittelstadt, and C. Russell, *Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI*, Oxford Internet Institute, University of Oxford, 2020.

Stakeholders views: 2020 BEUC survey - Consumers see potential of artificial intelligence but raise serious concerns: In a recent consumer organization survey in nine Member States, between 37% and 51% of respondents agree or strongly agree that AI will lead to unfair discrimination based on individual characteristics or social categories. **In the public consultation on the AI White Paper**, 89% of all respondents found that AI leading to discriminatory outcomes is an important or very important concern. This was a (very) important concern for 76% of SMEs and only 5% found it not important (at all). Large businesses were even more concerned: 89% said discrimination was a (very) important concern. Similarly, 91% of academic and research institutions and 90% of civil society organisations thought this was (very) important concern. Meanwhile, EU citizens were less concerned, although 78% still found this to be (very) important. Of those stakeholders stating that discriminatory outcomes were not important or very important concern, EU citizens (35%), academic and research institutions (19%) and SMEs (15%) were represented the most. For academic and research institutions and SMEs, this share was significantly larger than their representation in the overall sample.

2.1.4. Use of AI might violate the right to an effective remedy, fair trial and good administration

One prominent threat to the right to an effective remedy is **the lack of transparency in the use and operation of AI systems**.[140] Without access given to relevant information, individuals may not be able to defend themselves and challenge any decision taken or supported by AI systems that might adversely affect them. This jeopardizes their **right to be heard as well as the right to an effective remedy and fair trial.**[141]

Furthermore, the use of **automated decision-making in judicial proceedings** might particularly affect the right of affected persons to access to court and to fair trial, if these systems are not subject to appropriate safeguards for transparency, accuracy, non-discrimination and human oversight.[142]

The opacity of AI could also hamper the ability of persons charged with a crime to defend themselves and challenge the evidence used against them.[143] In the context of AI-enabled individual risk assessments increasingly used in law enforcement, singling out people without reasonable suspicion or on the basis of biased or flawed data[144] might also threaten **the presumption of innocence**.[145] Public authorities may also not be able to properly reason their individual administrative decisions which is required as part of the principle and the right to good administration.[146]

---

[140] Fundamental Rights Agency, *Artificial intelligence and fundamental rights*, 2020, Ferguson A. G., *Policing Predictive Policing*, Washington University Law Review, 2017, pp. 1165-1167.

[141] Ibidem, see also Council of Europe, *Algorithms and human rights*, 2017, pp.11 and 24. See also a recent judgment from Italy (T.A.R., Rome, sect. III-bis, 22 mars 2017, n 3769) that ruled that the simple description of the algorithm, in terms of decision-making process steps, without the disclosure of the specific sequence of instructions contained in the algorithm, would not constitute an effective protection of the subjective right concerned.

[142] EU Fundamental Rights Agency, *Artificial intelligence and fundamental rights*, 2020. See also the European Commission for the Efficiency of Justice, *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, 2018.

[143] See Erik van de Sandt et al. *Towards Data Scientific Investigations: A Comprehensive Data Science Framework and Case Study for Investigating Organized Crime & Serving the Public Interest*, November 2020.

[144] Meijer, A. and M. Wessels, 'Predictive Policing: Review of Benefits and Drawbacks', *International Journal of Public Administration* 42:12, 2019, p. 1032.

[145] The CJEU has ruled that the inclusion of a natural person in databases of potential suspects interferes with the presumption of innocence and can be proportionate 'only if there are sufficient grounds to suspect the person concerned' (CJEU, *Peter Puskar,* Case C-73/16, para 114).

[146] EU Fundamental Rights Agency, *Artificial intelligence and fundamental rights*, 2020.

Stakeholders views: In the public consultation on the White Paper on AI**,** only 2% of all respondents said that AI breaching fundamental rights is not (at all) an important concern. 85% of SMEs considered this (very) important, while none found the issue to be unimportant. Similarly, 87% large businesses found this to be (very) important– only one respondent (1%) found it not important. 93% and 94% of academic/research institutions and civil society organisations, respectively, were (very) concerned about fundamental rights breaches. EU citizens were also concerned: 83% found potential breaches of fundamental rights (very) important. Among those stakeholders who found this not to be a (very) important concern, academic and research institutions were the largest group with 33% (much higher than their 14% share of the entire sample).

Problem 3: Competent authorities do not have powers, resources and/or procedural frameworks to ensure and monitor compliance of AI use with fundamental rights and safety rules

The specific characteristics of many AI technologies, set out in section 2.2., often **make it hard to verify how outputs and decisions have been reached where AI is used**. As a consequence, it may become impossible to verify compliance with existing EU law meant to guarantee safety and protect fundamental rights. For example, to determine whether a recruitment decision is justified or involved discrimination, enforcement authorities need to determine how this decision was reached. Yet, since there is no requirement for producers and users of AI systems to keep proper documentation and ensure traceability of these decision-making processes, **public authorities may not be able to properly investigate, prove and sanction a breach**.

The governance and enforcement mechanisms under existing sectoral legislation also suffer from shortcomings. Firstly, the use of AI systems may lead to situations where market surveillance and supervisory authorities **may not be empowered to act and/or do not have the appropriate technical capabilities and expertise to inspect these systems**.

Secondly*,* existing secondary legislation on data protection, consumer protection and non-discrimination legislation relies primarily on ex-post mechanisms for enforcement and focuses on individual remedies for 'data subjects' or 'consumers'. To evaluate compliance with fundamental rights, the purpose and use of an application needs to be assessed in context and it is the responsibility of every actor to comply with their legal obligations. Unless legal compliance in view of the intended purpose and context is taken into account already at the design stage, **harmful AI systems might be placed on the market and violate individual fundamental rights at scale** before any enforcement action is taken by competent authorities.

Thirdly, as set out above, the current safety legislation does not provide yet for clear and specific requirements for AI systems that are embedded in products.[147] Outside the scope of product safety legislation, **there is also no binding obligation for prior testing and validation** of the systems before they are placed on the market. Moreover, after systems are placed on the market and deployed, there is no strict ex post obligation for continuous monitoring which is, however, essential given the continuous learning capabilities of certain AI system or their changing performance due to regular software updates.

Fourthly*,* the secondary legislation on fundamental rights **primarily places the burden for compliance on the user and often leaves the provider of the AI system outside its scope**.[148] However, while users remain responsible for a possible breach of fundamental rights obligations, providers might be best placed to prevent and mitigate some of the risks already at an early development stage. Users are also often unable to fully understand the workings of AI applications if not provided with all the necessary information. Because of these gaps in the existing legislation, procedures by supervisory authorities may not result in useful findings.

[147] However, this is going to be covered in the new Machinery legal act being revised for AI systems/ components having safety functions.

[148] See Recital 78 of the GDPR which states that producers of systems are not directly bound by the data protection legislation.

Finally, given the complexity and rapid speed of AI development, **competent authorities often lack the necessary resources, expertise and technological tools** to effectively supervise risks posed by the use of AI systems to safety and fundamental rights. They also do not have sufficient tools for cooperation with authorities in other Member States for carrying out joint investigations,[149] or even at national level where, for example, various sectoral legislation might intersect and lead to violations of multiple fundamental rights or to risks to both safety and fundamental rights.

Problem 4: Legal uncertainty and complexity on how to ensure compliance with rules applicable to AI systems dissuade businesses from developing and using the technology

Due to the specific characteristics of AI set out in section 2.2., businesses using AI technology are also facing increasing **legal uncertainty and complexity on how to comply with existing legislation**. Considering the various sources of risks at all different levels, organisations involved in the complex AI value chain might be unclear who exactly should ensure compliance with the different requirements under the existing legislation.[150]

For example, providers of AI systems might be unclear on what measures they should integrate to minimize the risks to safety and fundamental rights' violation, while users might not be able to remedy features of the design that are inadequate for the context of application. In this context, the evolving nature of risks also pose particular problems to correctly attribute responsibilities. Providers of AI systems may have limited information with regard to the harm that AI can produce post deployment (especially if the application context has not been taken into account in the design of the system), while users may be unable to exercise due care when operating the AI system if not properly informed about its nature and provided with guidance about the required oversight and ex post control. The lack of clear distribution of obligations across the AI value chain taking into account all these specific features of the AI technology leads to significant legal uncertainty for companies, while failing to effectively minimise the increasing risks to safety and fundamental rights, as identified above.

Another problem is that there are **no harmonized standards under EU law** as to how general principles or requirements on security, non-discrimination, transparency, accuracy, human oversight should be implemented specifically as regards AI systems in the design and development stage.[151] This results in legal uncertainty on the business side, which affects both the developer and the user of the AI system.

Also, there are **no clear red lines** when companies should not engage in the use of AI for certain particularly harmful practices beyond the practices explicitly listed in the Unfair Commercial Practice Directive.[152] **Certification** for trustworthy AI product and services that are currently available on the Union market is also missing and creates uncertainties across the AI value chain.

Without a clear legal framework, start-ups and developers working in this field will **not be able to attract the required investments**. Similarly, without certainty on applicable rules and clear common standards on what is required for a trustworthy, safe and lawful AI, developers and providers of AI systems and other innovators are less likely to pursue developments in this field.

---

[149] In sectors other than under the GDPR and the Law Enforcement Directive where data protection authorities from different Member States can cooperate.

[150] See, for example, Mahieu, R. and J. van Hoboken, *Fashion-ID: Introducing a Phase-Oriented Approach to Data Protection?*, European Law Blog, 2019.

[151] For example, assurance and quality control, metrics and thresholds used, testing and validation procedures, good practice risk prevention and mitigation measures, data governance and quality management procedures, or disclosure obligations.

[152] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation.

As a result**, businesses and public authorities using AI technology fear becoming responsible for fundamental rights infringements,** which may dissuade them from using AI. At European level, 69% of companies cite a 'need for new laws or regulation' as an obstacle to adoption of AI.[153] SMEs were more likely than large companies to say that this was not a challenge or barrier, whereas small enterprises were more likely to identify this as a major barrier (31% for companies with 5 to 9 employees and 28% for those with 10 to 49 employees) compared to medium and large companies.

Figure 2: Obstacles to the use of AI (by companies)

Obstacles by relevance	Overall	Adopters	Non-adopters	Plan to use
Strict standards for data exchange	77%	83%	72%	81%
The need for new laws or regulation	69%	76%	64%	73%
Lack of public or external funding	65%	71%	60%	69%
Lack of trust amongst citizens	65%	71%	61%	69%
Lack of access to or availability of public data	62%	68%	58%	65%
Reputational risks linked to using artificial intelligence	61%	69%	56%	64%
Liability for damage caused by artificial intelligence	59%	66%	54%	64%
Lack of access to high quality private data	58%	64%	54%	61%

Base question Q3: I will name potential EXTERNAL obstacles to the use of artificial intelligence. Please indicate all that your company has experienced as a challenge or a barrier. Base size: EU27, N=8661. (Base size represents only EU27 Member States, excluding the UK, Iceland and Norway).

Source: Ipsos Survey, 2020

At national level, a survey of 500 companies by the German Technical Inspection Association TÜV found that 42% of businesses expect legal challenges due to the use of AI, while 84% said that there was uncertainty around AI within their company. As a result, 84% wanted AI products to be clearly marked for users, and 87% were in favour of a risk-based approach to regulation.[154]

Yet, a McKinsey global survey (2017) showed that companies that are committed to AI have significantly higher profit margins across sectors. These companies also expect a margin increase of up to five percentage points more than industry average in the following three years.[155] Hence, direct costs from legal uncertainty in the market of AI development and use are also accompanied by a missed potential for business innovation. In the end, even consumers will suffer, as they will miss out beneficial products and services not developed by businesses due to the fear of legal consequences.

Problem 5: Mistrust in AI would slow down AI development in Europe and reduce the global competitiveness of the EU economy

If citizens observe that AI repeatedly endangers the safety of individuals or infringes their fundamental rights, **they are unlikely to be willing to accept the use of AI technologies for themselves or by other users.** In a recent survey aimed at consumers in nine EU Member States,[156]

[153] European Commission, Ipsos Survey, 2020. Large companies were represented significantly less than SMEs (44% as opposed to just above 50%).

[154] Note, however, that 54% also thought that regulation of AI inhibits innovation (TÜV, *Künstliche Intelligenz in Unternehmen*, 2020).

[155] McKinsey Global Institute, *Artificial Intelligence the next digital frontier?*,2017. Global survey to C-level executives, N=3.073

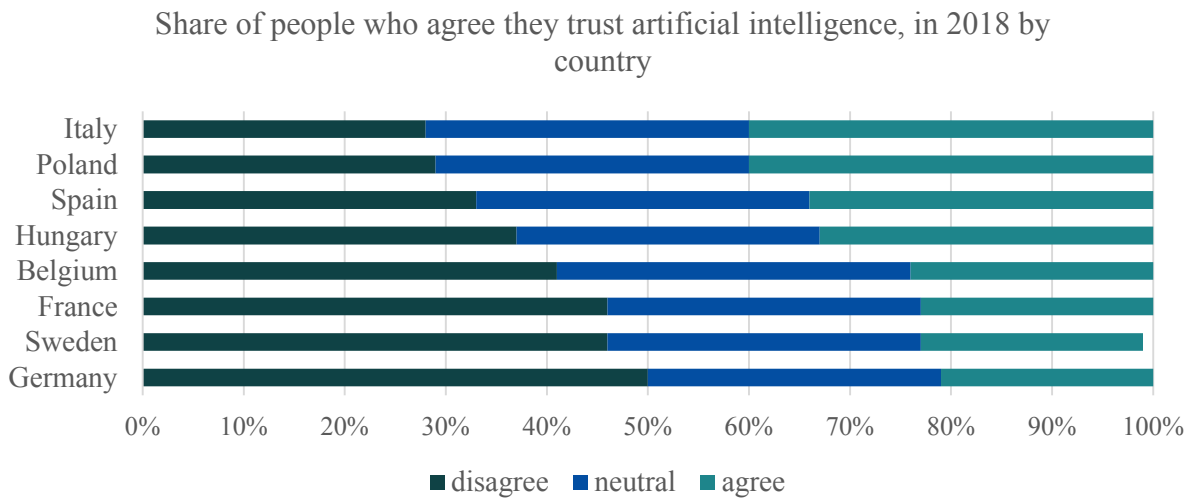
[156] BEUC, *Artificial Intelligence: what consumers say*, 2020.

around 56% of respondents said they had low trust in authorities to exert effective control over AI. Similarly, only 12% of the survey respondents in Sweden and Finland reported to trust private companies' abilities to address the ethical dilemmas AI brings. 77% thought that companies that are developing new AI solutions should be bound to ethical guidelines and regulation.[157]

Faced with reluctant private and business customers, **businesses will find it then more difficult to invest in and adopt AI** than if customers embrace AI. As a result, demand for new and innovative AI applications will be sub-optimal. This mistrust will hamper innovation because companies will be more hesitant in offering innovative services for which they first have to establish credibility, which will be particularly challenging in a context of growing public fears about the risks of AI. That is why in a recent study, 42% of EU executives surveyed see "guidelines or regulations to safeguard ethical and transparent use" as part of a policy to promote AI to the benefit of Europe.[158]

Similarly, a recent survey of European companies found that for 65% of respondents a lack of trust among citizens is an obstacle to the adoption of AI.[159] There is already a substantial share of companies not preparing for the AI-enabled future: 40% of the companies neither use any AI application whatsoever nor intend to do so.[160] The problem is particularly acute for SMEs that cannot rely on their brand to reassure customers that they are trustworthy. Significantly fewer large companies (34%) than SMEs (between 41% and 43% depending on company size) saw lack of trust as not an obstacle. The share of companies citing this as a major obstacle was relatively evenly distributed across company sizes (between 26% and 28%).

Figure 3: Trust in AI applications



Source: Ipsos survey 20-28 September 2018, quoted by Statista

Without a sound common framework for trustworthy AI, **Europe could lose out on the beneficial impact of AI on competitiveness.** Yet, the benefits of rapid adoption of AI are generally estimated as being very significant. McKinsey estimates that if Europe (EU28) on average develops and distributes AI according to its current assets and digital position relative to the world, it could add some €2.7 trillion, or 20%, to its combined economic output by 2030.[161] According to a European Value Added Assessment, prepared by the European Parliament, a common EU framework on the

[157] Tieto, *People in the Nordics are worried about the development of AI – personal data processing a major concern*, 2019. Number of respondents N=2648.
[158] McKinsey and DG CNECT, *Shaping the digital transformation in Europe*, 2020.
[159] European Commission, European *enterprise* survey on the use *of technologies based on artificial intelligence*, 2020.
[160] See above.
[161] McKinsey Global Institute, *Notes from the AI frontier: tackling Europe's gap in digital and AI*, 2019.

ethics of AI has the potential to bring the European Union € 294.9 billion in additional GDP and 4.6 million additional jobs by 2030.[162]

Thus, a slower adoption of AI would have **significant economic costs and would hamper innovation**. It would also mean **foregoing part of the wide array of the societal benefits** that AI is poised to bring in areas such as health, transport or pollution reduction. It would thus negatively affect not just businesses and the economy, but consumers and the society as well.

Problem 6: Fragmented measures create obstacles for a cross-border AI single market and threaten the Union's digital sovereignty

In the absence of a common European framework to address the risks examined before and build trust in AI technology, Member States can be expected to start taking action at a national level to deal with these specific challenges. While national legislation is within the Member States' sovereign competences,[163] there is a risk that **diverging national approaches will lead to market fragmentation** and could create obstacles especially for smaller companies to enter multiple national markets and scale up across the EU Single Market. Yet, as noted in section 1.2., AI applications are rapidly increasing in scale. Where advanced models work with billions of parameters, companies need to scale up their models to remain competitive. Since the high mobility of AI producers could lead to a race to the bottom where companies move to Member States with the lightest regulation and serve the entire EU market from there, other Member States may take measures to limit access from other Member States, leading to further market fragmentation.

That is why Member States in general support a common European approach to AI. In a recent position paper 14 Member States recognise the risk of market fragmentation and emphasise that **the 'main aim must be to create a common framework where trustworthy and human-centric AI goes hand in hand with innovation, economic growth and competitiveness'.**[164] Earlier, in its conclusions of 9 June 2020, the Council called upon the Commission 'to put forward concrete proposals, taking existing legislation into consideration, which follow a risk-based, proportionate and, if necessary, regulatory approach for artificial intelligence.'[165]

While waiting for a European proposal, **some Member States are already considering national legislative or soft-law measures to address the risks, build trust in AI and support innovation.** For example, the German Data Ethics Commission has called for a five-level risk-based system of horizontal regulation on AI that would go from no regulation for the most innocuous AI systems to a complete ban for the most dangerous ones.[166] Denmark has just launched the prototype of a Data Ethics Seal, whilst Malta has introduced a voluntary certification system for AI. Spain is in the process of adopting a Code of Ethics and considering certification of AI products and services. Finland issued recommendations for self-regulation and the development of responsibility standards for the private sector,[167] while Italy envisages certificates to validate and to monitor AI applications developed in an ethically sound way. Moreover, several Member States (e.g. Belgium, Sweden,

---

[162] European Parliamentary Research Service, *European added value assessment: European framework on ethical aspects of artificial intelligence, robotics and related technologies*, 2020.

[163] See, for example, CJEU Judgement of 14 October 2004, Omega , Case C-36/02, ECLI:EU:C:2004:614, where the EU Court of Justice has stated that Member States can take unilateral measures to restrict the free movements of services and goods if necessary and proportionate to ensure respect of fundamental rights guaranteed by the legal order of the Member States.

[164] *Non-paper - Innovative and trustworthy AI: two sides of the same coin*, Position paper on behalf of Denmark, Belgium, the Czech Republic, Finland, France Estonia, Ireland, Latvia, Luxembourg, the Netherlands, Poland, Portugal, Spain and Sweden, 2020.

[165] Council of the European Union, *Shaping Europe's Digital Future-Council Conclusions*, 8711/20, 2020.

[166] Datenethikkommission, *Opinion of the German Data Ethics Commission*, 2019.

[167] The AI Finland Project's ethics working group and the Ethics Challenge added emphasis on companies and self-regulation (AI Finland, 'Etiikkahaaste (Ethics Challenge)', *Tekoäly on uusi sähkö*. 2020).

Netherlands and Portugal) are considering the need for binding legislation on the legal and ethical aspects of AI.

In addition to this increasingly patchy national landscape, there is an **ongoing proliferation of voluntary international technical standards** for various aspects of 'Trustworthy AI' adopted by international standardisation organisations (e.g. IEEE, ISO/IEC, ETSI, ITU-T, NIST).[168] While these standards can in principle be very helpful in ensuring safe and trustworthy AI systems, there is also a growing risk of divergence between them since they are adopted by different international organisations. Moreover, these technical standards may not be fully compliant with existing EU legislation (e.g., on data protection or non-discrimination),[169] which creates additional liability risks and legal uncertainty for companies adhering to them. In addition, there is also a **proliferation of national technical standards on A**I that are adopted or being developed by a number of Member States and many other third countries around the world.[170] This means national standards risk not being fully interoperable or compatible with each other, which will create obstacles to cross-border movement and to the scaling up of AI-driven services and products across different Member States.

The impact of this increasing fragmentation is disproportionately affecting small companies. This is because large companies, especially global ones, can spread the additional costs for operating across an increasingly fragmented single market over their larger sales, especially when they already have established a dominant position in some markets. Meanwhile, SMEs and start-ups which do not have the market power or the same resources may be deterred from entering the markets of other Member States and thus fail to profit from the single market. This problem is further exacerbated since big tech players have not only a technological advantage but also exclusive access to large and quality data necessary for the development of AI. They may try to use this information asymmetry to seek economic advantages and further harm smaller companies. These dominant tech companies may also try to free ride on political efforts aiming to increase consumer protection by ensuring that the adopted standards for AI are in line with their own business practices to the detriment of newcomers and smaller players. This risk is significantly higher when the AI market is fragmented with individual Member States taking unilateral actions.

All these diverging measures **stand in the way of a seamless and well-functioning single market for trustworthy AI in the Union and pose particular legal barriers for SMEs and start-ups**. This in turn negatively affects the global competiveness of the EU industry, both regarding AI providers and the industries using AI, giving advantage to companies from third countries that are already dominant on the global market. Beyond the purely market dimension, there is a growing risk that **the 'digital sovereignty' of the Union and the Member States** might be threatened since such AI-driven products and services from foreign companies might not completely comply with Union values and/or legislation[171] or they might even pose security risks and make the European infrastructure more vulnerable. As stated by the President of the Commission von der Leyen, to ensure 'tech sovereignty', the EU should strengthen its capability to make its own choices, based on its own values, respecting its own rules.[172] Aside from strengthening the EU internal market, such

[168] See, for example, ISO, *ISO/IEC JTC 1/SC 42 Artificial intelligence*, 2017; Oceanis, *Repository*, 2020.

[169] Christofi, A., et.al. 'Erosion by Standardisation: Is ISO/IEC29134:2017 on Privacy Impact Assessment Up to GDPR Standard?', in M. Tzanou (ed.), *Personal Data Protection and Legal Developments in the European Union*, IGI Global, 2020.

[170] StandICT, *Standards Watch*, 2020.

[171] See, for example, the Clearview AI scandal where AI technology that is based on scraping of billions of images online can enter the Union market and be used by businesses and law enforcement agencies (Pascu, L., 'Hamburg data protection commissioner demands answers on biometric dataset from Clearview AI', *Biometric Update*, 2020.

[172] Ursula von der Leyen, *Shaping Europe's digital future: op-ed by Ursula von der Leyen President of the European Commission*, 2020.

tech sovereignty will also facilitate the development and the leverage of Union's tools and regulatory power to shape global rules and standards on AI.[173]

2.2. What are the main problem drivers?

The uptake of AI systems has a strong potential to bring benefits, economic growth and enhance EU innovation and global competitiveness.[174] However, as mentioned in the section above, in certain cases, the use of AI systems can also create problems for businesses and national authorities as well as new risks to safety and fundamental rights of individuals. The key cause explaining the analysed problems are the **specific characteristics of AI systems** which make them qualitatively different from previous technological advancements. Table 3 below explains what each characteristic means and why they can create problems for fundamental rights and safety.[175]

Table 3: Five specific characteristics of AI and Problem Drivers

CHARACTERISTICS	EXPLANATION (simplified)	WHY IS IT A PROBLEM?/ DRIVERS
Opacity/ (lack of transparency)	Limited ability of human mind to understand how certain AI systems operate	A lack of transparency (opacity)) in AI (due to complexity or how the algorithm was realized in code or how the application is realised) makes it difficult to monitor, identify and prove possible breaches of laws, including legal provisions that protect fundamental rights.
Complexity	Multiplicity of different components and processes of an AI system and their interlinks	The complexity of AI makes it difficult to monitor, identify and prove possible breaches of laws, including legal provisions that protect fundamental rights.
Continuous adaptation and unpredictability	Functional ability of some AI systems to continuously 'learn' and 'adapt' as they operate, sometimes leading to unpredictable outcomes.	Some AI systems change and evolve over time and may even change their own behaviour in unforeseen ways. This can give rise to new risks that are not adequately addressed by the existing legislation.
Autonomous behaviour	Functional ability of some AI systems to generate outputs such as 'decisions' with limited or no human intervention	The autonomous behaviour of AI systems can affect safety because of the functional ability of an AI system to perform a task with minimum or no direct human intervention.
Data	Functional dependence on data and the quality of data	The dependence of AI systems on data and their quality, the AI's 'ability' to infer correlations from data input and learn from data, including proxies, can reinforce systemic biases and errors, and exacerbate discriminatory and adverse results.

[173] European Council, *Special meeting of the European Council (1 and 2 October 2020) – Conclusions*, EUCO 13/20, 2020.

[174] See section 1. 'Introduction' of this impact assessment.

[175] For detailed analysis see Annex 5.2: Five specific characteristics of AI. Table 3 presents a simplified and non-technical explanation of the AI characteristics and their link to the problems. The main aim is to highlight main elements rather than provide a detail account of all elements of each characteristic, which is not possible due to space limitation.

Certain AI systems may include only some of the characteristics.[176] However, as a rule, the more specific characteristics a given AI system has, the higher the probability that it becomes a 'black box'. The term **'black box' reflects the limited ability of even most advanced experts to monitor an AI system**. This stands in considerable difference with the ability to monitor other technologies. The OECD report[177] explains this 'black box' effect with the example of neural networks as follows: "Neural networks iterate on the data they are trained on. They find complex, multi-variable probabilistic correlations that become part of the model that they build. However, they do not indicate how data would interrelate. The data are far too complex for the human mind to understand."[178] **'Black box' AI systems present new challenges for public policy compared to traditional and other technologies**.[179]

These specific characteristics of AI systems may create new (1) safety and security and (2) fundamental rights risks and accelerate the probability or intensity of the existing risks, as well as (3) make it hard for enforcement authorities to verify compliance with and enforce the existing rules. This set of problems leads in turn to (4) legal uncertainty for companies, (5) potentially slower uptake of AI technologies, due to the lack of trust, by businesses and citizens as well as (6) regulatory responses by Member States to mitigate possible externalities.[180] Consequently, problems triggered by specific characteristics of AI may lead to safety risks and breaches of fundamental rights and **challenge the effective application of and compliance with the EU legal framework for the protection of fundamental rights and safety**.

Table 4: Problem Tree

DRIVERS	PROBLEMS
The complexity and lack of transparency (opacity) of AI makes it difficult to monitor, identify and prove possible breaches of laws, including legal provisions that protect fundamental rights.	**(1)** Use of AI poses **increased risks to safety and security** of citizens
Some AI systems change and evolve over time and may even change their own behavior in unforeseen ways. It can give rise to new risks. The existing legislation is not adequately addressing these risks.	**(2)** Use of AI systems poses **increased risk of violations of citizens' fundamental rights and Union values**
Autonomy can affect the safety of the product, because it may alter a product's characteristics substantially, including its safety features.	**(3) Authorities do not have powers, procedural frameworks and resources to ensure and monitor compliance** of AI development and use with applicable rules
The dependence of AI systems on data and their quality, the AI's 'ability' to infer correlations from data input and learn from context data, including proxies, can reinforce systemic biases and errors, and exacerbate discriminatory and adverse results.	**(4) Legal uncertainty and complexity on how existing rules apply to AI systems** dissuade businesses from developing and using AI systems
	(5) Mistrust in AI would slow down AI development in Europe and reduce the global competitiveness of the EU economy
	(6) Fragmented measures create **obstacles for cross-border AI single market and threaten Union's digital sovereignty**

Rapid developments and uptake of AI systems increase this challenge. The Ipsos 2019 survey of European business indicates that 42% of enterprises currently use at least one AI technology, a quarter of them use at least two types, and 18% have plans to adopt AI technologies in the next two

[176] Furthermore, some AI systems may include mitigating mechanisms to reduce negative effects of some of the five characteristics.
[177] OECD, *Artificial Intelligence in Society*, 2019, p.23.
[178] OECD, *Artificial Intelligence in Society*, 2019.
[179] For analysis and reference to the supporting evidence see e.g. OECD, *Artificial Intelligence in Society*, 2019.
[180] See 'Problems' section 2 of this impact assessment.

years.[181] Moreover, the intensity of use of AI technology by business is expected to grow in the next two years. This data suggests that **in the next two years, it is likely that more than half of all EU businesses will be using AI systems.** Thus, AI systems already affect business and consumers in the EU on a large scale.

According to the European Commission Better Regulation Guidelines and accompanying Toolbox,[182] a public policy intervention may be justified, among other grounds, when regulations fail or when protection and fulfilment of fundamental rights afforded to citizens of the Union provide grounds for intervention.[183] Several factors may cause regulatory failures, including, when existing public intervention becomes "out of date as the world evolves." Protection and fulfilment of fundamental rights afforded to citizens of the Union may also provide important reasons for policy intervention 'because even a perfectly competitive and efficient economy can produce outcomes that are unacceptable in terms of equity'.[184]

2.3. How will the problem evolve?

Given the increasing public awareness of the potential for violation of safety and fundamental rights that AI has, it is likely that the proliferation of ethics principles will continue. Companies would adopt these principle unilaterally in an effort to reassure their potential customers. However, such non-binding ethical principles cannot build the necessary trust as they cannot be enforced by affected parties and no external party or regulator is actually empowered to check whether these principles are duly respected by the companies and the public authorities developing and using AI systems. Moreover, the multiplicity of commitments would require consumers to spend an extraordinary amount of time understanding which commitments apply to which application.

As a consequence, and given the significant commercial opportunities offered by AI solutions, 'untrustworthy' AI solutions could ensue, with a likely backlash against AI technology as a whole by citizens and businesses. If and when this happens, European citizens will lose out on the benefits of AI and European companies will be placed at a significant disadvantage compared to their overseas competitors with a dynamic home market.

Over time, technological developments in the fields of algorithmic transparency, accountability and fairness could improve the situation, but progress and impact will be uncertain and uneven across Europe. On the contrary, as AI develops, it can be implemented in more and more situations and sectors, so that the problems identified above apply to an ever-growing share of citizens' life.

It cannot be excluded that over the long run and after a sufficient number of incidents, consumers will prefer companies with a proven track record of trustworthy AI. However, apart from the damage done in the meantime, this would have the consequence of favouring large companies, who can rely on their brand image, over SMEs who will face increasing legal barriers to enter the market.

3. WHY SHOULD THE EU ACT?

3.1. Legal basis

The initiative constitutes a core part of the EU single market strategy given that artificial intelligence has already found its way into a vast majority of services and products and will only continue to do so in the future. EU action on the basis of Article 114 of the Treaty on the

---

[181] According to a global survey, the number of business using AI grew 270% in the past four years and just in the last year tripled, Gartner, *Gartner Survey Shows 37 Percent of Organizations Have Implemented AI in Some Form*, 2019.

[182] European Commission, *Commission Staff Working Document – Better Regulation Guidelines*, SWD (2017) 350.

[183] See above, specifically Toolbox 14.

[184] See above, Toolbox 14, pp. 89-90.

Functioning of the European Union can be taken for the purposes of the approximation of the provisions laid down by law, regulation or administrative action in the Member States when it has as its object the establishment and functioning of the internal market. The measures must be intended to improve the conditions for the establishment and functioning of the internal market and must genuinely have that object, actually contributing to the elimination of obstacles to the free movement of goods or services, or to the removal of distortions of competition.

Article 114 TFEU may be used as a legal basis to prevent the occurrence of these obstacles resulting from diverging national laws and approaches how to address the legal uncertainties and gaps in the existing legal frameworks applicable to AI.[185] The present initiative aims to improve the functioning of the internal market by setting harmonized rules on the development, placing on the Union market and the use of products and services making use of the AI technology or provided as stand-alone AI applications. Some Member States are already considering national rules to ensure AI is safe and is developed and used in compliance with fundamental rights obligations. This will likely lead to a further fragmentation of the internal market and increasing legal uncertainty for providers and users on how existing and new rules will apply to AI systems.

Furthermore, the Court of Justice has recognised that applying heterogeneous technical requirements could be valid grounds to trigger Article 114 TFEU.[186] The new initiative will aim to address that problem by proposing harmonised technical standards for the implementation of common requirements applicable to the design and development of certain AI systems before they are placed on the market. The initiative will also address the situation after AI systems have been placed on the market by harmonising the way in which ex-post controls are conducted.

Based on the above, Article 114 TFEU is the applicable legal basis for the present initiative.[187] In addition, considering that this Regulation contains certain specific rules, unrelated to the functioning of the internal market, restricting the use of AI systems for 'real-time' remote biometric identification by the law enforcement authorities of the Member States, which necessarily limits the processing of biometric data by those authorities, it is appropriate to base this Regulation, in as far as those specific rules are concerned, on Article 16 of the Treaty.

3.2. Subsidiarity: Necessity of EU action

The intrinsic nature of AI which often relies on large and varied datasets and which might be embedded in any product or service circulating freely within the internal market mean that the objectives of the initiative cannot effectively be achieved by Member States alone. An emerging patchy framework of potentially divergent national rules will hamper the seamless provision of AI systems across the EU and is ineffective in ensuring the safety and protection of fundamental rights and Union values across the different Member States. Such an approach is unable to solve the problems of ineffective enforcement and governance mechanisms and will not create common conditions for building trust in the technology across all Member States. National approaches in addressing the problems will only create additional legal uncertainty, legal barriers and will slow market uptake of AI even further. Companies could be prevented from seamlessly expanding into other Member States, depriving consumers and other users of the benefits of their services and products and affecting negatively the competitiveness of European companies and the economy.

---

[185] CJEU Judgment of the Court (Grand Chamber) of 3 December 2019, *Czech Republic v European Parliament and Council of the European Union*, Case C-482/17, paras. 35.

[186] CJEU Judgment of the Court (Grand Chamber) of 2 May 2006, *United Kingdom of Great Britain and Northern Ireland v European Parliament and Council of the European Union*, Case C-217/04, paras. 62-63.

[187] Article 114 TFEU as a legal basis for EU action on AI was also suggested by the European Parliament in its 2017 resolution on civil law rules on robotics and in its 2020 resolution on ethical framework for AI, robotics and related technologies. See European Parliament resolution 2020/2012(INL).

3.3. Subsidiarity: Added value of EU action

The objectives of the initiative can be better achieved at Union level so as to avoid a further fragmentation of the Single Market into potentially contradictory national frameworks preventing the free circulation of goods and services embedding AI. In their positions to the White paper on AI all Member States support a coordinated action at EU level to prevent the risk of fragmentation and create the necessary conditions for a single market of safe, lawful and trustworthy AI in Europe. A solid European regulatory framework for trustworthy AI will also ensure a level playing field and protect all European citizens, while strengthening Europe's competitiveness and industrial basis in AI.[188] A common EU legislative action on AI could boost the internal market and has great potential to provide European industry with a competitive edge at the global scene and economies of scale that cannot be achieved by individual Member States alone. Setting up the governance structures and mechanisms for a coordinated European approach to AI across all sectors and Member States will enhance safety and the respect of fundamental rights, while allowing businesses, public authorities and users of AI systems to capitalise on the scale of the internal market and use safe and trustworthy AI products and services. Only common action at EU level can also protect Union's tech sovereignty and leverage its tools and regulatory powers to shape global rules and standards.

4. OBJECTIVES: WHAT IS TO BE ACHIEVED?

The problems analysed in section 2 above are complex and cannot be fully addressed by any single policy intervention. This is why the Commission proposes to address emerging problems related specifically to AI systems gradually. The objectives of this initiative are defined accordingly.

Table 5: General/Specific objectives

GENERAL OBJECTIVE	SPECIFIC OBJECTIVES
The general objective of the intervention is to **ensure the proper functioning of the single market** by creating the conditions for the **development and use of trustworthy artificial intelligence** in the Union.	**Ensure that AI systems** placed on the market and used are **safe and respect existing law on fundamental rights and Union values**. **Ensure legal certainty** to facilitate **investment and innovation in AI**. **Enhance governance and effective enforcement** of existing law on fundamental rights and safety requirements applicable to AI systems. **Facilitate** the development of a **single market for lawful, safe and trustworthy AI applications** and prevent market fragmentation.

4.1.General objectives

The general objective of the intervention is to **ensure the proper functioning of the single market** by creating the conditions for the development and use of trustworthy artificial intelligence in the Union.

4.2.Specific objectives

The **specific objectives** of this initiative are as follows:

[188] For the analysis of the European added value of the EU action see also European Parliamentary Research Service, *European added value assessment: European framework on ethical aspects of artificial intelligence, robotics and related technologies*, 2020.

- set requirements specific to AI systems and obligations on all value chain participants in order to ensure that AI systems placed on the market and used are safe and respect the existing law on fundamental rights and Union values;
- ensure legal certainty to facilitate investment and innovation in AI by making it clear what essential requirements, obligations, as well as conformity and compliance procedures must be followed to place, or use an AI system in the Union market;
- enhance governance and effective enforcement of the existing law on fundamental rights and safety requirements applicable to AI systems by providing new powers, resources and clear rules for relevant authorities on conformity assessment and ex post monitoring procedures and the division of governance and supervision tasks between national and EU levels;
- facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation by taking EU action to set minimum requirement for AI systems to be placed and used in the Union market in compliance with the existing law on fundamental rights and safety.

4.2.1. Ensure that AI systems placed on the market and used are safe and respect the existing law on fundamental rights and Union values

Safeguarding the safety and fundamental rights of EU citizens is a cornerstone of European values. However, the emergence of AI creates new challenges regarding safety and protection of fundamental rights and hinder the enforcement of these rights, due to the specific features of this technology (see section 2.2.). However, the same rights and rules that apply in the analogue world should also be respected when AI systems are used. The first specific objective of the initiative is, therefore, to ensure that AI systems that are developed, placed on the market and/or used in the Union are safe and respect the existing law on fundamental rights and Union values by setting requirements specific to AI systems and obligations on all value chain participants. The ongoing review of the sectoral safety legislation will pursue a similar objective to ensure safety of products embedding AI technology, but focusing on the overall safety of the whole product and the safe integration of the AI system into the product.[189]

4.2.2. Ensure legal certainty to facilitate investment and innovation in AI

Due to the specific characteristics of AI, businesses using AI technology are also facing increasing legal uncertainty and complexity on how to comply with existing legislation. As long as the challenges and risks to safety and fundamental rights have not been addressed, companies must also calculate the risk that legislation or other requirements will be introduced, without knowing what this will imply for their business models. Such legal uncertainty is detrimental for investment and especially for innovation. The second objective is therefore to promote investment and innovation by creating single market-wide legal certainty and an appropriate framework that stimulates innovation by making it clear what essential requirements, obligations as well as conformity and compliance procedures must be followed to place or use an AI system in the Union market. The complementary initiative on liability rules will also aim to increase legal certainty in the use of AI technology, but by ensuring a high-level of protection of victims who have suffered harms caused by certain AI systems.[190]

[189] For the interaction between the AI initiative and revision of the product safety legislation see section 8 (preferred option) and Annex 5.3.

[190] For the interaction between the AI initiative and the initiative on liability see section 8 (preferred option).

4.2.3. Enhance governance and effective enforcement of the existing law on fundamental rights and safety requirements applicable to AI systems

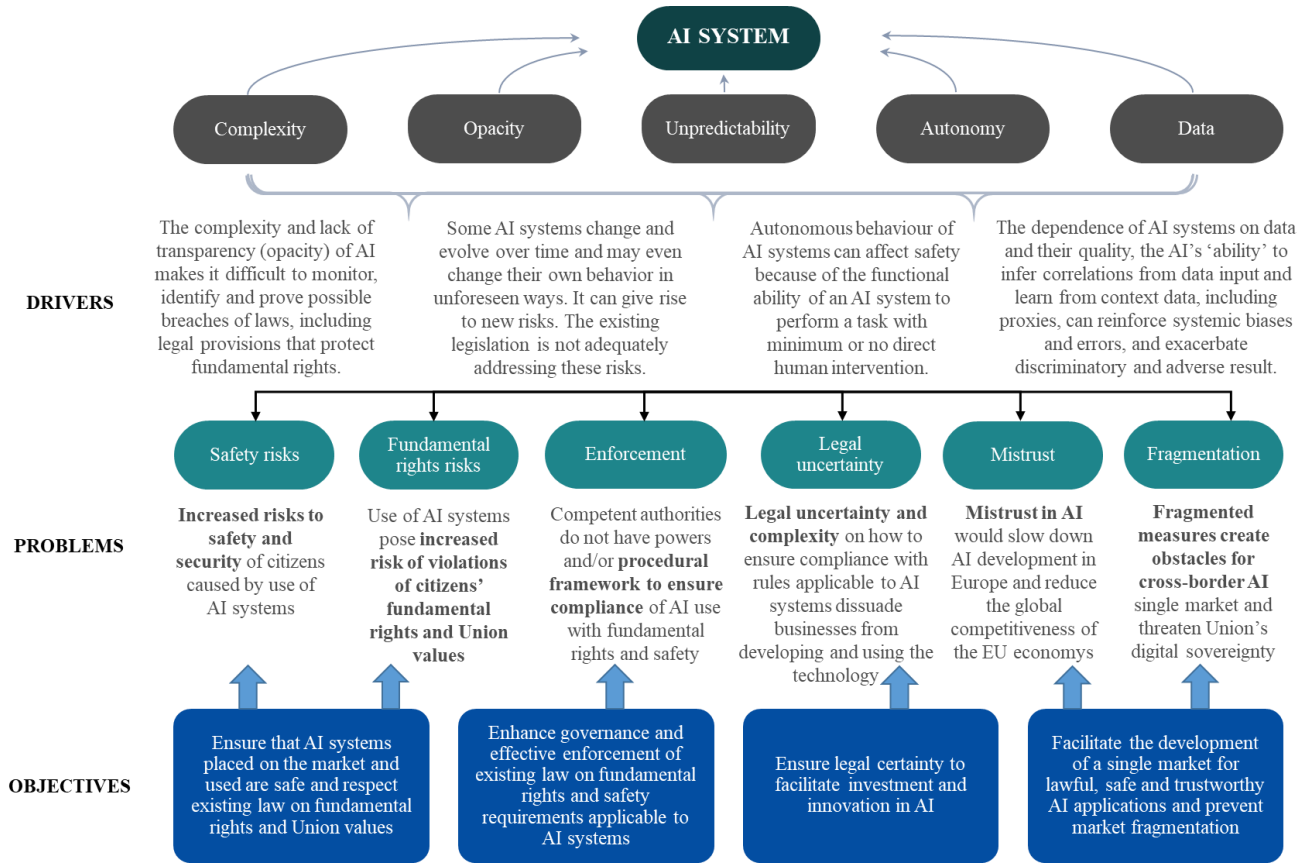
The technological features of AI such as opacity and autonomous behaviour might cause violations of safety rules and the existing law on fundamental rights that may not even be noticed by the concerned person, and that, even when they are noticed, are often difficult to prove. Existing competent authorities might also face difficulties in auditing the compliance of certain AI systems with the existing legislation due to the specific technological features of AI. They might also lack powers to intervene against actors who are outside their jurisdiction or lack sufficient resources and a mechanism for cooperation and joint investigations with other competent authorities. The enforcement and governance system needs to be adapted to these new challenges so as to ensure that possible breaches can be effectively detected and sanctioned by enforcement authorities and those affected. The third objective is therefore to improve the governance mechanism and effective enforcement of the existing law on fundamental rights and safety requirements applicable to AI by providing new powers, resources and clear rules for relevant authorities on conformity assessment and ex post monitoring procedures and the division of governance and supervision tasks between national and EU levels.

4.2.4. Facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation

The safety and fundamental rights risks posed by AI may lead citizens and consumers to mistrust this technology, inciting in turn Member States to address these problems with national measures which may create obstacles to cross border sales, especially for SMEs. The fourth objective is, hence, to foster trustworthy AI, which will reduce the incentives for national and potentially mutually incompatible legislations and will remove legal barriers and obstacles to cross border movement of products and services embedding the AI technology by taking EU action to set minimum requirements for AI systems to be placed and used in the Union market in compliance with the existing law on fundamental rights and safety. The complementary initiative on liability rules would also aim to increase trust in the AI technology, but by ensuring a high-level of protection of victims who have suffered harms caused by certain AI systems.

4.3. Objectives tree/intervention logic.

Figure 4: Intervention logic



The specific characteristics of certain AI systems (opacity, complexity, autonomous behaviour, unpredictability and data dependency) may create (1) safety and security and (2) fundamental rights risks and (3) make it hard for enforcement authorities to verify compliance with and enforce the existing rules. This set of problems in turn leads to other problems causing (4) legal uncertainty for companies, (5) potentially slower uptake of AI technologies, due to the lack of trust, by businesses and citizens as well as (6) unilateral regulatory responses by Member States to mitigate possible externalities.

Firstly**, current EU law does not effectively ensure protection for safety and fundamental rights risks specific to AI systems,** as shown in the problem definition (Problems 1 and 2). Particularly, risks caused by opacity, complexity, continuous adaptation, autonomous behaviour and the data dependency of AI systems (drivers) are not fully covered by the the existing law. Accordingly, this initiative sets out the specific objective to set requirements specific to AI systems and obligations on all value chain participants in order to ensure that AI systems placed or used in the Union market are safe and respect the existing law on fundamental rights and Union values (specific objective 1).

Secondly, under current EU law, **competent authorities do not have sufficient powers**, resources and/or procedural frameworks in place to effectively ensure and monitor compliance of AI systems with fundamental rights and safety legislation (**problem 3**). The specific characteristics of AI systems (drivers) often make it hard to verify how outputs/decisions have been reached where AI is used. As a consequence, **it may become impossible to verify compliance with existing EU law** meant to guarantee safety and protection of fundamental rights. Competent authorities also do not have sufficient powers and resources to effectively inspect and monitor these systems. To address these problems, the initiative sets the objective to enhance governance and effective enforcement of the existing law on fundamental rights and safety requirements applicable to AI systems by providing new powers, resources and clear rules for relevant authorities on conformity assessment and ex post monitoring procedures and the division of governance and supervision tasks between national and EU levels (**specific objective 3**).

Thirdly, current EU legislation does provide certain requirements related to safety and protection of fundamental rights that apply to new technologies, including AI systems. However, those **requirements are not specific to AI systems, they lack legal certainty or standards on how to be implemented and are not consistently imposed on different actors across the value chain**. Considering the specific characteristics of AI (drivers), providers and users do not have clarity as to how existing obligations should be applied to AI systems for these systems to be considered safe, trustworthy and in compliance with the existing law on fundamental rights (**problem 4**). Furthermore, the lack of clear distribution of obligations across the AI value chain also contributes to problems 1 and 2. To address those problems, the initiative sets the objective to clarify what essential requirements, obligations, as well as conformity and compliance procedures actors must follow to place, or use an AI system in the Union market (**specific objective 2**).

Finally, in the absence of an EU legislation on AI that addresses the new specific risks to safety and fundamental rights, **businesses and citizens distrust the technology (problem 5)**, while Member States' unilateral action to address that **problem risks to create obstacles for a cross-border AI single market** and threatens the Union's digital sovereignty (problem 6). To address these problems, the proposed initiative has the objective to facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation by taking EU action to set minimum requirements for AI systems to be placed and used in the Union market in compliance with the existing law on fundamental rights and safety (**specific objective 4**).

5. WHAT ARE THE AVAILABLE POLICY OPTIONS?

The analysed policy options are based on the following main dimensions: a) The nature of the EU legal act (no EU intervention/ EU act with voluntary obligations/ EU sectoral legislation/ horizontal EU act); b) Definition of AI system (voluntary/ ad hoc for specific sectors/ one horizontal definition); c) Scope and content of requirements and obligations (voluntary/ ad hoc depending on the specific sector/ risk-based/ all risks covered); d) Enforcement and compliance mechanism (voluntary/ ex ante or ex post only/ ex ante and ex post); e) Governance mechanism (national, national and EU, EU only).

The policy options, summarized in Table 6 below, represent the spectrum of policy options based on the dimensions outlined above.

Table 6: Summary of the analysed policy options

	Option 1 **EU Voluntary labelling scheme**	**Option 2** **Ad hoc sectoral approach**	**Option 3** **Horizontal risk-based act on AI**	**Option 3+** **Codes of conduct**	**Option 4** **Horizontal act for all AI**
NATURE OF ACT	An EU act establishing **a voluntary labelling scheme**	**Ad hoc sectoral acts** (revision or new)	A single binding **horizontal act on AI**	Option 3 + code of conducts	A single binding **horizontal** act on AI
SCOPE/ DEFINITION OF AI	One definition of AI, however applicable only on a voluntary basis	**Each sector can adopt a definition** of AI and determine the riskiness of the AI systems covered	**One horizontally applicable AI definition** and methodology for determination of high- risk **(risk-based)**	Option 3 + industry-led codes of conduct for non-high-risk AI	**One horizontal AI definition,** but no methodology/or gradation (**all risks covered**)
REQUIREMENTS	Applicable only for **voluntarily labelled AI systems.**	Applicable only **for sector specific AI systems** with possible additional safeguards/ limitations for specific AI use cases per sector	Risk-based horizontal requirements **for prohibited and high risk AI systems** + min. information requirements for certain other AI	Option 3 + industry-led codes of conduct for non-high-risk AI	For **all AI systems** irrespective of the level of the risk

			systems		
OBLIGATIONS	**Only for providers** who adopt voluntary scheme and no obligations for users of certified AI systems	**Sector specific obligations for providers** and **users** depending on the use case	**Horizontal obligations** for providers and users of high-risk AI systems	Option 3 + commitment to comply with codes of conduct for non-high-risk AI	Same as Option 3, but applicable to all AI (irrespective of risk)
EX ANTE ENFORCEMENT	**Self-assessment** and an ex ante check by national competent authorities responsible for monitoring compliance with the EU voluntary label	Depends on the enforcement system **under the relevant sectoral acts**.	**Conformity assessment** for providers of high-risk systems (3rd party for AI in a product and other systems based on internal checks) + **registration in an EU database.**	Option 3 + self-assessment for compliance with codes of conduct for non- high-risk AI	Same as Option 3, but applicable to all AI (irrespective of risk)
EX POST ENFORCEMENT	Monitoring by authorities responsible for EU voluntary label	Monitoring by competent authorities under the relevant sectoral acts	Monitoring of **high-risk systems** by market surveillance authorities	Option 3 + unfair commercial practice in case of non-compliance with codes	Same as Option 3, but applicable to all AI (irrespective of risk)
GOVERNANCE	National competent authorities designated by Member States responsible for the EU label + a light EU cooperation mechanism	Depends on the sectoral acts at national and EU level; no platform for cooperation between various competent authorities.	At the national level but reinforced with cooperation between Member States authorities and with the EU level (AI Board)	Option 3 + without EU approval of the codes of conduct	Same as Option 3, but applicable to all AI (irrespective of risk)

5.1. What is the baseline from which options are assessed?

Under the **baseline scenario**, there would be no specific legislation at European level comprehensively addressing the issues related to AI discussed above. Ongoing revisions of other existing legislations, such as the review of the Machinery Directive 2006/42/EC and the General Product Safety Directive 2001/95/EC would continue. Both directives are technology neutral and their review will address aspects that are related to new digital technologies, not only specific to AI systems.[191] In other areas, in particular with regard to use of automated tools, including AI, by online platforms, the rules proposed in the Digital Services Act and the Digital Markets Act (once adopted) would establish a governance system to address risks as they emerge and ensure a sufficient user-facing transparency and public accountability in the use of these systems. [192]

[191] The revision of the Machinery Directive will address risks emerging from new technologies and problems related to software with a safety function and placed independently on the market, human-robot collaboration, the loss of connection of a device and cyber risks, transparency of programming codes, risks related to autonomous machines and lifecycle related requirements. The revision of the General Product Safety Directive might address cybersecurity risks when affecting safety, mental health, evolving functionalities and substantive modifications of consumer products.

[192] The proposal of the Digital Services Act, for examples, include obligations to maintain a risk management system, including annual risk assessments for determining how the design of intermediary service, including their algorithmic processes, as well as the use (and misuse) of their service contribute or amplify the most prominent societal risks posed by online platforms. It would also include an obligation to take proportionate and reasonable measures to mitigate the detected risks, and regularly subject the risk management system to an independent audit.

In parallel to these revisions, the EU would also promote industry-led initiatives for AI in an effort to advance 'soft law' but would not establish any framework for such voluntary codes. Currently, an increasingly large number of AI principles and ethical codes has already been developed by industry actors and other organisations.[193] In the Union, the HLEG developed a set of Ethics Guidelines for Trustworthy AI with an assessment list aimed at providing practical guidance on how to implement each of the key requirements for AI. The 'soft law' approach could build upon existing initiatives and consist of reporting on the voluntary compliance with such initiatives based on self-reporting (without any involvement of public supervisory authorities or other accredited organisations); encouraging industry-led coordination on a single set of AI principles; awareness raising among AI systems developers and users around the existence and utility of existing initiatives; monitoring and encouraging the development of voluntary standards that could be based on the non-binding HLEG ethical guidelines.

In the absence of a regulatory initiative on AI, the risks identified in section 2 would remain unaddressed. EU legislation on the protection of fundamental rights and safety would remain relevant and applicable to a large number of emerging AI applications. However, increased violations of fundamental rights and a higher exposure to safety risks including problems with enforcement of existing EU law may grow as AI continues to develop.

In the baseline scenario, there is also a large offer of forecasts of the AI market which all assume an unopposed development of AI and significant growth with projections for the EU market in 2025 between €32 billion and €66 billion. However, by not considering the possibility of backlashes, the forecasts may prove over-optimistic in the absence of regulation. As an example of such a backlash, in March 2020 one major forecaster predicted a compound annual growth rate for the facial recognition market of 14.5% from 2020 of 2027.[194] Yet in June 2020, following claims that facial recognition systems were of discriminatory nature, one market leader (IBM) stopped developing and selling these systems, while two other major players (Amazon and Microsoft) decided to suspend their sales to a major customer (the law enforcement sector).[195] Similar developments cannot be excluded in other AI use cases, especially where claims of discrimination have already led to pressure from public opinion (e.g. recruitment software,[196] sentencing support[197] or financial services).[198] Similarly, the use of CT scans for COVID diagnosis has not been rolled out as quickly as possible due to the reluctance of hospitals to use uncertified technologies.

Consequently, the lack of any decisive policy action by the EU could lead to increased fragmentation due to interventions at Member State level, as public opinion would put pressure on politicians and law-makers to address the concerns described above. As a result of national approaches the single market for AI products and services would be further fragmented with different standards and requirements that will create obstacles to cross border movement. This would reduce the competitiveness of European businesses and endanger Europe's digital autonomy.

Furthermore, enhanced transparency and reporting obligations with regard to content moderation and content amplification are proposed. Finally, proposal of the Digital Services Act also envisages user-facing transparency obligation of content recommender systems, enabling users to understand why, and influence how information is being presented to them, as well as far-reaching data access provisions for regulators and vetted researchers, and strong enforcement and sanctions powers including at EU level.

[193] See Fundamental Rights Agency, *AI Policy Initiatives 2016-2020*, 2020; Jobin, A., M. Ienca and E. Vayena, '*The global landscape of AI ethics guidelines*', *Nature Machine Intelligence* Volume 1, pp. 389–399, 2019.

[194] Grand View Research, *Facial Recognition Market Size.* Industry Report, 2020.

[195] Hamilton I.A., *Outrage over police brutality has finally convinced Amazon, Microsoft, and IBM to rule out selling facial recognition tech to law enforcement. Here's what's going on.* Business Insider, 13/06/2020.

[196] Dastin J., Amazon *scraps secret AI recruiting tool that showed bias against women,* Reuters, 11/11/ 2020.

[197] Larson J., et al., *How We Analyzed the COMPAS Recidivism Algorithm,* Propublica, 23/05/2016.

[198] Vigdor N., *Apple Card Investigated After Gender Discrimination Complaints,* The New York Times, 10/11/2019.

***Stakeholders views*:** In the public consultation on the White Paper on AI, 16% of SMEs saw current legislation as sufficient to address concerns related to AI. 37% saw gaps in current legislation, and 40% the need for new legislation. Among large businesses too, a majority of respondents said that current legislation was insufficient. Academic and research institutions overwhelmingly came to the conclusion that current legislation was not sufficient. Only 2% said otherwise, while 48% saw a need for new legislation and 35% saw gaps in existing legislation. Almost no civil society organisation deemed current legislation sufficient. Among those stakeholders claiming that current legislation was sufficient, EU citizens, large companies (both 25%), and business associations (22%) were the largest groups. For large companies and business associations, this was around double their share in the overall sample of respondents. For SMEs (13%), this was also true.

5.2. Option 1: EU legislative instrument setting up a voluntary labelling scheme

Under this option, an EU legislative instrument would establish an **EU voluntary labelling scheme** to enable providers of AI applications **certify their AI systems' compliance with certain requirements for trustworthy AI and obtain an EU-wide label**. While participation in the scheme would be voluntary, the instrument would envisage an appropriate enforcement and governance system to ensure that providers who subscribe comply with the requirements and take appropriate measures to monitor risks even after the AI system is placed on the market. Given the voluntary character of the initiative aimed at the AI system's certification, the instrument would not impose certification obligations on users of labelled AI systems since these would be impractical and not voluntary in nature.

Table 6.1. Summary Option 1: EU Voluntary labelling scheme

Nature of act	An EU act establishes **a voluntary labelling scheme**, which becomes binding once adhered to
Scope	OECD definition of AI; adherence possible irrespective of the level of risk, but certain risk differentiation amongst the certified AI systems also possible
Content	**Requirements for labelled AI systems:** data, transparency and provision of information, traceability and documentation, accuracy, robustness and human oversight (to be ensured by providers who choose to label their AI system)
Obligations	**Obligations for providers (**who voluntarily agree to comply) for quality management, risk management and ex post monitoring **No obligations for users** of certified AI systems (impractical given the voluntary character of the label aimed at certification of specific AI systems)
Ex ante enforcement	Self-assessment and ex ante check by national competent authorities responsible for monitoring compliance with the EU voluntary label
Ex post enforcement	Ex post monitoring by national competent authorities responsible for monitoring compliance with the EU voluntary label
Governance	National competent authorities designated by Member States as responsible for the EU label + a light EU cooperation mechanism

5.2.1. Scope of the EU voluntary labelling scheme

Given the voluntary nature of the EU voluntary labelling scheme, this would be applicable **regardless of the level of risk** of the AI system, but certain risk differentiation amongst the certified AI systems could also be envisaged. The instrument would build on the internationally recognized **OECD definition of AI,**[199] because it is technology neutral and future proof. This

[199] AI will be defined in the legal act as 'a machine-based system that can, for a given set of human-defined objectives, generate output such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy' based on the OECD definition (OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 2019). To cover a broader range of 'AI outputs' (e.g. deep fakes and other content), the OECD definition has been slightly adapted referring to 'AI outputs such as predictions, recommendations or decisions'.

policy choice is justified because a technology-specific definition[200] could cause market distortions between different technologies and quickly become outdated. The OECD definition is also considered sufficiently broad and flexible to encompass all problematic uses as identified in the problem section. Last but not least, it would allow a consensus with the international partners and third countries and ensure that the proposed scheme would be compatible with the AI frameworks adopted by the EU's major trade partners.

Stakeholders views: In the Public Consultation on the White Paper on AI there were some disagreements between stakeholder groups regarding the exact definition of AI, proposed as comprising 'data' and 'algorithms'. At least 11% of large companies and 10% of SMEs found this definition too broad. Only 2% of large companies and no SMEs said it was too narrow. This likely reflects concerns about too many AI systems falling under potential future requirements, thus creating an additional burden on companies. On the other hand, the civil society organisations tended to find it too narrow. Furthermore, at least 11% of large companies and 5% of SMEs said that the definition was unclear and would need to be refined.

5.2.2. Requirements for Trustworthy AI envisaged in the EU voluntary labelling scheme

The voluntary labelling scheme would impose certain **requirements for Trustworthy AI** which would aim to address the main sources of risks to safety and fundamental rights during the development and pre-market phase and provide assurance that the AI system has been properly tested and validated by the provider for its compliance with existing legislation.

These requirements for trustworthy AI would be limited to the necessary minimum to address the problems and include the following 5 requirements, identified in the White Paper: a) data governance and data quality; b) traceability and documentation; c) algorithmic transparency and provision of information; d) human oversight, and e) accuracy, robustness and security.

Figure 5: Requirements for Trustworthy AI systems

[200] For example, focusing only on machine learning technology.

Requirement	Description	Outcome in relation to drivers and problems
Data governance & quality datasets	▪ Design and develop the AI systems based on data governance procedures (e.g. on collection, labelling etc.) ▪ Datasets should be sufficiently representative in view of the intended use, including the vis-à-vis the affected people and the specific European context of application	▪ Address the data dependency of the AI system (driver) ▪ Improve the reliability and accuracy of AI systems ▪ Prevent risks of potentially inaccurate, biased or discriminatory decisions or safety risks (problems 1 and 2)
Traceability and Documentation	▪ Ensure that the AI's decision-making process is traceable and that relevant documentation is kept, including data used	▪ Address the AI's complexity, opacity and unpredictability (drivers) ▪ Demonstrate compliance with existing legislation and facilitate investigating breaches of fundamental rights and safety obligations (problems 1, 2 and 3)
Transparency & provision of information	▪ Strengthen algorithmic transparency and provide information to users about the purpose and key characteristics of the system	▪ Ensure users can effectively control the system even in case of unpredictable and/or autonomous behavior (drivers) ▪ Users properly informed and can take all necessary measures to minimize the residual risks (problems 1 and 2) ▪ AI systems' outputs can be verified and justified (problems 2 and 3)
Human oversight	▪ Ensure the AI system can be overseen by humans through appropriate measures (e.g. indications for abnormal performance, stop button). ▪ Other mitigating measures should be identified by the provider and implemented by the user	▪ Address autonomous behaviour of the AI systems and in certain instances their unpredictability and continuous adaptation (drivers) ▪ Ensure that humans always remain in control and are responsible for the operation of systems even when they learn (problem 1 and 2)
Accuracy, robustness and security	▪ Ensure that the AI system meets the level of accuracy, robustness and security intended by the provider and suitable for the intended purpose	▪ Address the complexity, autonomous behavior and unpredictability (drivers) ▪ Ensure AI systems are safe and secure to external attacks and perform accurately and reliably and do not violate fundamental rights (problems 1 and 2)

The 5 requirements above are the result of two years of preparatory work and derived from the Ethics Guidelines of the HLEG,[201] piloted by more than 350 organisations.[202] They are also largely consistent with other international recommendations and principles[203] which would ensure that the proposed EU framework for AI would be compatible with those adopted by the EU's international trade partners. The EP also proposed similar requirements,[204] but it was decided not to include some of the EP proposals or the HLEG principles as requirements. This is because they were considered either too vague, too difficult to operationalize,[205] or already covered by other legislation.[206]

---

[201] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019.

[202] They were also endorsed by the Commission in its 2019 Communication on human-centric approach to AI.

[203] For example, the OECD AI Principles endorsed also by G20, the Council of Europe Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems. April 2020, the U.S. President's Executive Order from 3 December 2020 on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government etc.

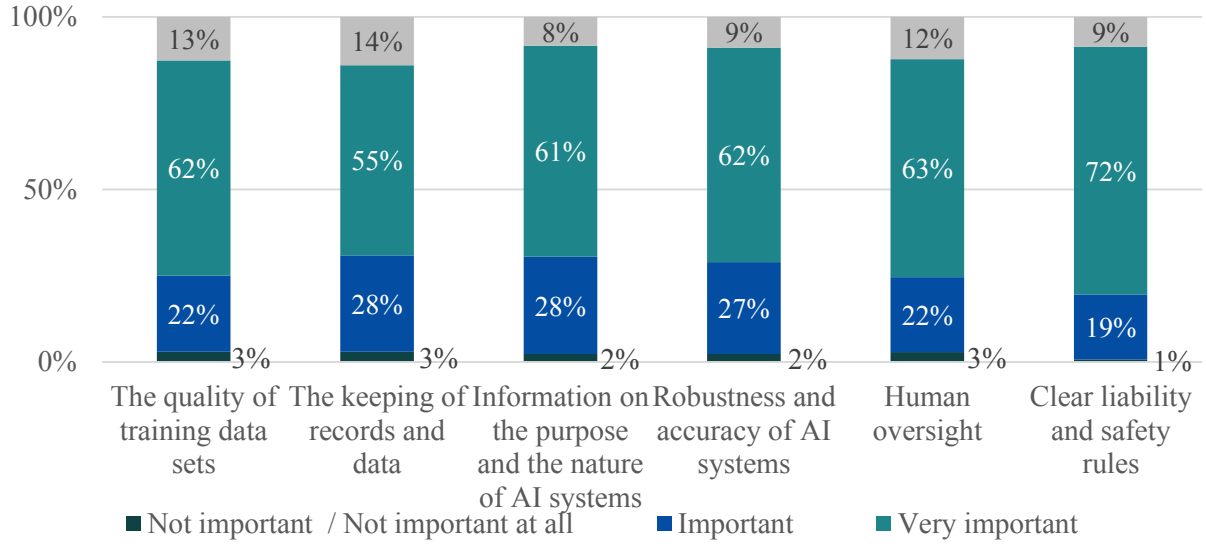
[204] The EP has proposed the following requirements for high-risk AI: human oversight; safety, transparency and accountability; non bias and non-discrimination; social responsibility and gender equality; environmental sustainability; privacy, and the right to seek redress. The requirements for accountability is for example operationalised through the documentation requirements and obligations, while non-discrimination through the requirements for data and data governance.

[205] Environmental and social well-being are aspirational principles included in the HLEG guidelines, the OECD and G20 principles, the draft UNICEF Recommendation of the ethics of AI as well as the EP position. They have not been included in this proposal, because they were considered too vague for a legal act and too difficult to operationalise. The same applies for social responsibility and gender equality proposed by the EP.

[206] For example, the requirement for privacy is already regulated under the GDPR and the Law Enforcement Directive. The requirement for effective redress proposed by the EP would be covered by the separate liability initiative on AI (planned for Q4 2021) or under existing legislation (e.g. rights of data subjects under the GDPR to challenge a fully automated decision).

The proposed minimum requirements are already state-of-the-art for many diligent business operators and they would ensure **minimum degree of algorithmic transparency and accountability** in the development of AI systems. Requiring that AI systems are developed with **high quality datasets** that reflect the specific European context of application and intended purpose would also help to ensure that these systems are reliable, safe and minimize the risk of discrimination once deployed by users. These requirements have also been largely supported by stakeholders in the consultation on the White Paper on AI.[207]

Figure 6: Stakeholder consultation results on the requirements for AI



Source: Public Consultation on the White Paper on Artificial Intelligence

To prove compliance with the requirements outlined above, providers who choose to subscribe to the voluntary labelling scheme would also have to establish within their organisation an **appropriate quality management and risk management system**, including with prior testing and validation procedures to detect and prevent unintended consequences and minimize the risks to safety and fundamental rights.[208] To take into account the complexity and the possibility for continuous adaptation of certain AI systems and the evolving risks, providers would also have to **monitor the market ex post** and take any corrective action, as appropriate (problems 1, 2 and 3).

5.2.3. Enforcement and governance of the EU voluntary labelling scheme

While participation in the labelling scheme would be voluntary, providers who choose to participate would have to comply with these requirements (in addition to existing EU legislation) to be able to display a quality 'Trustworthy AI' label. The label would serve as an indication to the market that the labelled AI application is trustworthy, thus addressing partially the mistrust problem for those certified AI applications (problem 5).

The scheme would be enforced **through ex ante self-assessment[209] and ex-post monitoring by competent authorities designated by the Member States**. This is justified by the need to improve governance and enforceability of the requirements specific to AI systems (problem 3) as well as for practical reasons. On the one hand, competent authorities would first have to register the

[207] For a detailed breakout of the views of the various stakeholder groups on these issues, see Annex 2.

[208] Council of Europe Recommendation CM/Rec(2020)1 also states that risk-management processes should detect and prevent the detrimental use of algorithmic systems and their negative impacts. Quality assurance obligations have also been introduced in other regulatory initiatives in third countries such as the Canada's Directive on Automated Decision-Making.

[209] Possibly based on the ALTAI self-assessment tool.

commitment of the provider to comply with the AI requirements and check if this is indeed the case. On the other hand, ex post supervision would be necessary to ensure that compliance is an ongoing process even after the system has been placed on the market. Where relevant, this would also aim to address the evolving performance of the AI system due to its continuous adaptation or software updates.

The instrument would also establish **appropriate sanctions** for providers participating in the scheme that have claimed compliance with the requirements, but are found to be non-compliant. The sanctions would be imposed following an investigation with a final decision issued by the competent national authority responsible for the scheme at national level.[210]

A light mechanism for EU cooperation is also envisaged with a network of national competent authorities who would meet regularly to exchange information and ensure uniform application of the scheme. An alternative would be not to envisage any cooperation at EU level, but this would compromise the European character and uniform application of the European voluntary labelling scheme.

Despite some inherent limitations of the voluntary labelling scheme in ensuring legal certainty and achieving the development of a truly single market for trustworthy AI (problems 4 and 6), this option would still have some positive effects to increase trust and address AI challenges by means of a more gradual regulatory approach, so it should not be excluded a priori from the detailed assessment.

5.3. Option 2: A sectoral, 'ad-hoc' approach

This option would tackle specific risks generated by certain AI applications **through ad-hoc legislation or through the revision of existing legislation on a case by case basis**.[211] There would be no coordinated approach on how AI is regulated across sectors and no horizontal requirements or obligations. The sector specific acts adopted under this option would include sector specific requirements and obligations for providers and users of certain risky AI applications (e.g. remote biometric identification, deep fakes, AI used in recruitments, prohibition of certain AI uses etc.). Their content would depend on the specific use case and would be enforced through different enforcement mechanisms without a common regulatory framework or platform for cooperation between national competent authorities. The development and use of all other AI systems would remain unrestricted subject to the existing legislation.

Table 6.2. Summary Option 2: Ad hoc sectoral approach

Nature of act	**Case-by-case binding sectoral acts** (review of existing legislation or ad hoc new acts)
Scope	Different sectoral acts could adopt different definitions of AI that might be inconsistent. Each sectoral act will determine the risky AI applications that should be regulated.
Content	**Sector specific requirements for AI systems** (could be similar to Option 1, but adapted to sectoral acts) + **Additional safeguards for specific AI use cases:** - Prohibition of certain harmful AI practices - Additional safeguards for permitted use of remote biometric identification

[210] Sanctions will include 1) suspension of the label and 2) imposition of fines proportionate to the size of the company in case of serious and/or repeated infringements for providing misleading or inaccurate information. In case of minor and first time infringements, only recommendations and warnings can be issued for imposing possible future sanctions in case the non-compliance persists.

[211] De facto, this is already happening in some sectors, for example, how drones are regulated under the EU Regulations 2019/947 and 2019/945 for the safe operation of drones in European skies or the specific rules for trading algorithms under the MiFID II/MiFIR financial legislation.

	(RBI) systems, deep fakes, chatbots.
Obligations	**a. Sector specific obligations for providers** (could be similar to Option 1, but adapted to ad hoc sectoral acts) b. **Sector specific obligations for users** depending on the use case (e.g. human oversight, transparency in specific cases etc.)
Ex ante enforcement	Would depend on the enforcement system under the relevant sectoral acts For use of remote biometric identification (RBI) systems at publicly accessible spaces (when permitted): prior authorisation required by public authorities
Ex post enforcement	Ex post monitoring by competent authorities under the relevant sectoral acts
Governance	Would depend on the existing structures in the sectoral acts at national and EU level; no platform for cooperation between various competent authorities.

5.3.1. Scope of the ad-hoc sectoral acts

It would be for each ad-hoc piece of legislation to determine what constitutes risky AI applications that require regulatory intervention. These different acts might also adopt different definitions of AI which would increase legal uncertainty and create inconsistencies across sectors, thus failing to address effectively problems 4 and 6.

To address problems 1 and 2, the ad hoc approach would target both risks to fundamental rights and safety and cover the following sectoral initiatives:

- With regard to **AI which are safety components of products** covered by new approach or old-approach safety legislations, this option would entail the review of those legislations so as to include dedicated requirements and obligations addressing safety and security risks and, to the extent appropriate, fundamental right risks related to the AI safety components of those products which are considered high-risk.[212]
- **With regard to other AI with mainly fundamental rights implications,** the sectoral approach would exclude integration of the new specific requirements for AI in the data protection legislation, because it is designed as a technology-neutral legislation covering automated personal data processing in general (i.e. automated and non-automated processing). This means that each specific AI use case posing high risks to fundamental rights would have to be regulated through new ad-hoc initiatives or integrated into existing sectoral legislation to the extent that such exist.[213]

[212] Based on up-to date analysis, concerned NLF legislations would be: Directive 2006/42/EC on machinery (which is currently subject to review), Directive 2009/48/EU on toys, Directive 2013/53/EU on recreational craft, Directive 2014/33/EU of the European Parliament and of the Council of 26 February 2014 on lifts and safety components for lifts, Directive 2014/34/EU on equipment and protective systems intended for use in potentially explosive atmospheres, Directive 2014/53/EU on radio-equipment, Directive 2014/68/EU on pressure equipment, Directive 2014/90/EU on marine equipment, Regulation (EU) 2016/424 on cableway installations, Regulation (EU) 2016/425 on personal protective equipment, Regulation (EU) 2016/426 on gas appliances, Regulations (EU) 745/2017 on medical devices and Regulation (EU) 746/2017 on in-vitro diagnostic medical devices. The concerned old-approach legislation would be Regulation (EU) 2018/1139 on Civil Aviation, Regulation 858/2018 on the approval and market surveillance of motor vehicles, Regulation (EU) 2019/2144 on type-approval requirements for motor vehicles, Regulation (EU) 167/2013 on the approval and market surveillance of agricultural and forestry vehicles, Regulation (EU) 168/2013 on the approval and market surveillance of two- or three-wheel vehicles and quadricycles, Directive (EU) 2016/797 on interoperability of railway systems.

[213] An example of such sectoral ad hoc legislation targeting specific use case of non-embedded AI with mainly fundamental rights implications could be the recent proposal of the New York City Council proposal for a regulation on automated hiring tools. In addition to employment and recruitment, Option 3 and Annex 5.4 identify other high risk use cases also for remote biometric identification systems in publicly accessible places, use of AI for determining access to educational institutions and evaluations; to evaluate the eligibility for social security benefits and services; creditworthiness, predictive policing as well as some other problematic use cases in law enforcement, judiciary, migration and asylum and border controls.

5.3.2. Ad hoc sector specific AI requirements and obligations for providers and users

The content of the ad hoc initiatives outlined above would include: a) ad hoc sector specific requirements and obligations for providers and users of certain risky AI systems; b) additional safeguards for the permitted use of remote biometric identification systems in publicly accessible places; and c) certain prohibited harmful AI practices.

a) Ad hoc sector specific AI requirements and obligations for providers and users

Firstly, the ad hoc sectoral acts would envisage **AI requirements and obligations for providers similar to those in Option 1, but specifically tailored to each use case**. This means they may be different from one use case to another, which would allow consideration of the specific context and the sectoral legislation at place. This approach would also encompass the full AI value chain with some **obligations placed on users, as appropriate for each use case.** Examples include obligations for users to exercise certain human oversight, prevent and manage residual risks, keep certain documentation, inform people when communicating with an AI system, in case the latter believe they are interacting with a human,[214] and label deep fakes, if not used for legitimate purposes so as to prevent the risk of manipulation.[215]

However, this ad hoc approach would also lead to sectoral market fragmentation and increase the risk of inconsistency between the new requirements and obligations, in particular where multiple legal frameworks apply to the same AI system. All these potential inconsistencies could further increase legal uncertainty and market fragmentation (problems 4 and 6). The high number of pieces of legislation concerned would also make the timelines of the relevant initiatives unclear and potentially very long with the mistrust in AI further growing (problem 5).

b) Additional safeguards for the permitted use of remote biometric identification systems in publicly accessible places

One very specific and urgent case that requires regulatory intervention is the use of **remote biometric identification systems in publicly accessible spaces**.[216] 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. In addition, a dedicated ad hoc instrument would **prohibit certain uses of remote biometric identification systems in publicly accessible spaces** given their unacceptable adverse impact on fundamental rights, while other uses of such systems would be considered high-risk because they pose significant risks to fundamental rights and freedoms of individuals or whole groups thereof.[217]

---

[214] The draft UNICEF recommendation on AI also emphasizes the need to protect the right of users to easily identify whether they are interacting with a living being, or with an AI system imitating human or animal characteristics.

[215] The EP similarly requested in a recent report that an obligation for labelling of deep fakes should be introduced in a legislation. This is in line with actions taken in some states in the U.S. and also considered in the UK to require labelling of deep fakes or prohibit their use in particular during election campaigns or for person's impersonation. See also a recent report of Europol and United Nations Interregional Crime and Justice Research Institute on the Malicious Uses and Abuses of Artificial Intelligence, 2020 which identifies deep fakes as an emerging threat to public security.

[216] See problem section 2.1.2.1.

[217] This is overall consistent with the EP position in its resolution on the ethics of AI that the use and gathering of biometric data by private entities for remote identification purposes in public areas, such as biometric or facial recognition, would not be allowed. Only Member States' public authorities may carry out such activities under strict conditions, notably regarding its scope and duration. The Council of Europe has also proposed certain prohibitions and safeguards in the use of facial recognition technology, see Consultative Committee of The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data Convention 108 Guidelines on Facial Recognition, 28 January 2021, T-PD(2020)03rev4.

For these high-risk uses, in order to balance risks and benefits, in addition to the requirements and obligations placed on the provider before the system is placed on the market (as per option 1), the ad hoc instrument would also impose **additional safeguards and restrictions on the use of such system.** In particular, uses of remote facial recognition systems in publicly accessible places that are not prohibited would require submission of a data protection impact assessment by the user to the competent data protection authority to which the data protection authority may object within a defined period. Additional safeguards would ensure that such use for legitimate security purposes is limited only to competent authorities. It would have to comply with strict procedural and substantive conditions justifying the necessity and proportionality of the interference in relation to the people who might be included in the watchlist, the triggering events and circumstances that would allow the use and strict limitations on the permitted geographical and temporal scope. All these additional safeguards and limitations would be on the top of the existing data protection legislation that would continue to apply by default.

An alternative policy option requested by some civil society organisations is to prohibit entirely the use of these systems in publicly accessible spaces, which would however prevent their use in duly justified limited cases for security purposes (e.g., in the presence of an imminent and foreseeable threat of terrorism or for identifying offenders of a certain numerus clausus of serious crimes when there is a clear evidence that they are likely to occur in a specific place at a given time). Another option would be not to impose any further restrictions on the use of remote biometric identification in publicly accessible places and apply only the requirements for Trustworthy AI (as per option 1). However, this policy choice was also discarded as it would not effectively address the high risks to fundamental rights posed by these systems and the current potential for their arbitrary abuse without an effective oversight mechanism and limitations on the permitted use (problems 2 and 3).

***Stakeholders views*:** In the public consultation on the White Paper on AI, 28% of respondents supported a general ban of this technology in public spaces, while another 29.2% required a specific EU guideline or legislation before such systems may be used in public spaces. 15% agreed with allowing remote biometric identification systems in public spaces only in certain cases and under conditions and another 4.5% asked for further requirements (on top of the 6 requirements for high-risk applications proposed in the white paper) to regulate such conditions. Only, 6.2% of respondents did not think that any further guidelines or regulations are needed. Business and industry were more likely to have no opinion on the use of remote biometric identification, or to have a slightly more permissive stance: 30.2% declared to have no opinion on this issue, 23.7% would allow biometric identification systems in public spaces only in certain cases and under conditions, while 22.4% argued for a specific EU guideline or legislation before such systems may be used in public spaces. On the other hand, civil society was more likely to call for bans (29.5%) or specific EU guidelines/legislation (36.2%). 55.4% of the respondent citizens were most likely to call for a ban, while academia (39%) were more supportive of specific EU guidelines/legislation.

The Commission has also registered a European Citizens' Initiative entitled 'Civil society initiative for a ban on biometric mass surveillance practices'.

c) Prohibition of certain harmful AI practices

Finally, to increase legal certainty and set clear red lines when AI cannot be used (problems 2 and 4), the ad-hoc approach would also introduce dedicated legislation to **prohibit certain other particularly harmful AI practices** that go against the EU values of democracy, freedom and human dignity, and violate fundamental rights, including privacy and consumer protection.[218] Alternatively, these could be integrated into relevant existing laws once reviewed.[219]

[218] Prohibition of certain particularly harmful AI practices has been requested by more than 60 NGOs who sent an open letter to the Commission.

[219] The prohibition of the manipulative practice could possibly be integrated into the Unfair Commercial Practice Directive, while the prohibitions of the general purpose social scoring of citizens could be possibly included in the General Data Protection Regulation.

Evidence and analysis in the problem definition suggest that exiting legislation does not provide sufficient protection and there is a need for the prohibitions of i) certain manipulative and exploitative AI systems, and ii) general purpose social scoring:

i. AI systems that **manipulate humans** through subliminal techniques beyond their consciousness or exploit vulnerabilities of specific vulnerable groups such as children in order to materially distort their behaviour in a manner that is likely to cause these people psychological or physical harm. As described in the problem section, this prohibition is justified by the increasing power of algorithms to subliminally influence human choices and important decisions interfering with human agency and the principle of personal autonomy. This prohibition is consistent with a number of recommendations of the Council of Europe[220] and UNICEF.[221]
ii. AI systems used for **general purpose social scoring of natural persons** done by public authorities defined as large scale evaluation or classification of the trustworthiness of natural persons based on their social behaviour in multiple contexts and/or known or predicted personality characteristics that lead to detrimental treatment in areas unrelated to the context in which the information was collected, including by restricting individual's fundamental rights or limiting their access to essential public services. This prohibition is justified because such mass scale citizens' scoring would unduly restrict individuals' fundamental rights and be contrary to the values of democracy, freedom and the principles that all people should be treated as equals before the law and with dignity. A similar prohibition of social scoring was requested by more than 60 NGOs and also recommended by the HLEG.[222] The EP has also recently requested a prohibition of intrusive citizens' mass scale social scoring in one of its reports on AI.[223]

Other manipulative and exploitative practices enabled by algorithms that are usually identified as harmful (e.g., exploitative profiling and micro-targeting of voters and consumers) were considered as potential candidates for prohibition but discarded, since these problems have been specifically examined and targeted by the recent proposal for the Digital Services Act.[224] To a large extent, they are also already addressed by existing Union legislation on data protection and consumer protection that impose obligations for transparency, informed consent/opt out and prohibit unfair commercial practices, thus guaranteeing the free will and choice of people when AI systems are used. Furthermore, other prohibitions requested by NGOs (e.g., in relation to predictive policing, use of AI for allocation of social security benefits, in border and migration control and AI-enabled individualised risk assessments in the criminal law)[225] were also considered, but eventually discarded. That is because the new requirements for trustworthy AI proposed by the sectoral ad hoc

[220] Council of Europe, Declaration on the manipulative capabilities of algorithmic processes, 13 February 2019, Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems also states that experimentation designed to produce deceptive or exploitative effects should be explicitly prohibited. With regard to children who are vulnerable group, this prohibition is also consistent with Recommendation CM/Rec(2018)7 on Guidelines to respect, protect and fulfil the rights of the child in the digital environment that advocates for a precautionary approach and for taking measures to prevent risks and practices adversely affecting the physical, emotional and psychological well-being of a child and to protect their rights in the digital environment.

[221] UNICEF, Policy guidance on AI for children, September, 2020.

[222] HLEG Policy and Investment Recommendations For Trustworthy AI, June 2019. For a similar social credit system introduced in China, see Chinese State Council Notice concerning Issuance of the Planning Outline for the Construction of a Social Credit System (2014-2020) GF No. (2014)21. Systems where people are socially scored with discriminatory or disproportionately adverse treatment have also been put in place in some Member States such as the Gladaxe system in Denmark or the SyRI system in the Netherlands.

[223] See EP report on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice (2020/2013(INI)).

[224] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final.

[225] See an open letter to the Commission from EDRi and more than 60 other NGOs.

acts under this option would aim to address these problematic use cases and ensure that the AI systems used in these sensitive contexts are sufficiently transparent, non-discriminatory, accurate and subjected to meaningful human oversight without the need to prohibit outright the use of AI in these contexts that could also be beneficial, if subjected to appropriate safeguards.[226]

5.3.3. Enforcement and governance of the ad hoc sectoral acts

The **enforcement mechanism** would depend on each sector and legislative act, and could vary according to the specific class of applications. For example, in the context of safety legislation, enforcement of the new requirements and obligations would be ensured through the existing ex ante conformity assessment procedures and the ex post market surveillance and monitoring system. The enforcement of the prohibited practices would take place through ex post monitoring by the competent data protection or consumer protection authorities. The new rules on remote biometric identification systems would be enforced by the data protection authorities responsible for the authorisation of their use subject to the conditions and limitations in the new instrument and the existing data protection legislation.

As regards the **governance system**, the sectoral national authorities under each framework would be responsible for supervising the compliance with the new requirements and obligations. There would be no platform for cooperation, nor possibilities for joint investigations between various competent authorities responsible for the implementation of the ad hoc sectoral legislation applicable to AI. The cooperation at EU level would be limited to the existing mechanisms and structures under each sectoral act. Competent authorities would still be provided with new powers, resources and procedures to enforce the sectoral rules applicable to certain specific AI systems, which could partially address problem 3.

5.4. Option 3: Horizontal EU legislative instrument establishing mandatory requirements for high-risk AI applications

Option 3 would envisage a **horizontal EU legislative instrument** applicable to all AI systems placed on the market or used in the Union following a **proportionate risk-based approach.** The horizontal instrument would establish a single definition of AI (section 5.4.1) and harmonised horizontal requirements and obligations to address in a proportionate, risk-based manner and limited to the strictly necessary the risks to safety and fundamental rights specific to AI (section 5.4.2). A common system for enforcement and governance of the new rules would also be established applicable across the various sectors (section 5.4.3) complemented with specific measures to support innovation in AI (section 5.4.4).

The rules would be uniform and harmonised across all Member States which would address the problems of legal uncertainty and market fragmentation and help citizens and companies build trust in the AI technology placed on the Union market (problems 4 to 6).

Table 6.3. Summary of Option 3: Horizontal risk-based act on AI

Nature of act	A single **binding horizontal act** following a **risk-based** approach
Scope	OECD definition of AI (reference point also for other sectoral acts); clear methodology and criteria how to determine what constitutes a high-risk AI system
Content	Risk-based approach: a. **Prohibited AI practices and additional safeguards for the permitted use of remote biometric identification systems in publicly accessible spaces** (as

[226] The use of these AI systems in the public sector could also be beneficial if subjected to appropriate safeguards as it would help public authorities to be more effective in the allocation of scarce public resources, thus potentially improving the access to these services and even reducing discrimination in individual human decisions that might also be biased.

	per Option 2) b. **Horizontal requirements** as per Option1, but binding **for high-risk AI** and operationalized through harmonised standards c. **Minimal transparency for non-high-risk AI** (inform when using chatbots and deep fakes as per Option 2) +**Measures to support innovation** (sandboxes etc.)
Obligations	Binding **horizontal obligations for all actors** across the value chain: a. **Providers of high-risk AI systems** as per Option 1 + conformity (re-) assessment, reporting of risks/breaches etc. b. **Users of high-risk AI systems** (human oversight, monitoring, minimal documentation)
Ex ante enforcement	**Providers:** a. Third party conformity assessment for high-risk AI in products (under sectoral safety legislation) b. Mainly ex ante assessment through internal checks for other high-risk AI systems + registration in a EU database **Users:** Prior authorisation for use of Remote biometric identification in publicly accessible spaces (as per Option 2)
Ex post enforcement	Ex post monitoring by market surveillance authorities designated by Member States
Governance	Governance at national level with a possibility for joint investigations between different competent authorities + cooperation at EU level within an AI Board

5.4.1. A single definition of AI

Like option 1, the horizontal instrument would build on the internationally recognized OECD definition of an AI system, because it is technology neutral and future proof. To provide legal certainty, the broad definition may be complemented with a list of specific approaches and techniques that can be used for the development of AI systems with some flexibility to change the list to respond to future technological developments.[227] The definition of AI in the horizontal act would act as a reference point for other sectoral legislation and would ensure consistency across the various legislative frameworks applicable to AI, thus enhancing legal certainty for operators and reducing the risk of sectoral market fragmentation (problems 4 and 6).

5.4.2. Risk-based approach with clear and proportionate obligations across the AI value chain

The horizontal instrument would follow a risk-based approach where AI applications would be regulated only where strictly necessary to address the risks and with the minimum necessary regulatory burden placed on operators. The risk-based approach would have the following elements: a) prohibited AI practices and additional safeguards for the permitted use of remote biometric recognition systems in publicly accessible spaces; b) a consistent methodology for identification of high-risk AI systems; c) horizontal requirements for high-risk AI systems and clear and

[227] The EP has called for an instrument equally applying to AI, but also covering robotics and related technologies. 'Related technologies' is defined by EP as 'software to control with a partial or full degree of autonomy a physical or virtual process, technologies capable of detecting biometric, genetic or other data, and technologies that copy or otherwise make use of human traits' would be covered by the OECD definition to the extent that this concerns technologies that enable software to control with a partial or full degree of autonomy a physical or virtual process. 'Technologies capable of detecting biometric, genetic or other data, and technologies that copy or otherwise make use of human traits' are covered only to the extent that they use AI systems as defined by OECD. The rest is excluded, because any technology that is able to detect and process 'other data' would qualify as AI, which is considered excessively broad and beyond AI. While cognitive robotics would be included in the list of AI approaches and techniques, other robots are not related to the same AI characteristics as described in section 2.2. and do not pose specific fundamental rights risks, so these are already sufficiently covered by the existing product safety legislation.

proportionate obligations for providers and users of these systems; d) minimal transparency requirements for certain low-risk AI systems.

a) Prohibited AI practices and additional safeguards for the permitted use of remote biometric recognition systems in publicly accessible spaces;

Firstly, the instrument would prohibit some harmful AI practices with a view to increasing legal certainty and setting clear red-lines when AI cannot be used (problems 2 and 4). These would include the same practices as envisaged in Option 2 (e.g., manipulative and exploitative AI and general-purpose social scoring of citizens). This option would also integrate the same prohibitions of certain uses of remote biometric identification systems in publicly accessible spaces and additional safeguards of the use of such systems when permitted as per Option 2.

b) A consistent methodology for identification of high-risk AI systems

Secondly, the instrument would introduce **clear horizontal rules for a number of 'high-risk' AI use cases**[228] with demonstrated high-risks for safety and/or fundamental rights (problems 1, 2 and 4). The list of applications considered 'high-risk' would be identified on the basis of **common criteria and a risk assessment methodology** specified in the legal act as follows:

1. **AI systems that are safety components of products** would be high-risk if the product or device in question undergoes third-party conformity assessment pursuant to the relevant new approach or old approach safety legislation.[229-230]
2. **For all other AI systems**,[231] it would be assessed whether the AI system and its intended use generates a high-risk to the health and safety and/or the fundamental rights and freedom of persons on the basis of a number of criteria that would be defined in the legal proposal.[232] These criteria are objective and non-discriminatory since they treat similar AI systems similarly, regardless of the origin of the AI system (EU or non-EU). They also focus on the probability and severity of the harms to the health and safety and/or the fundamental rights, taking into account the specific characteristics of AI systems of opacity, complexity etc.

---

[228] The definition of high-risk used in the context of a horizontal framework may be different from the notion of high-risk used in sectoral legislation because of different context of the respective legislations. The qualification of an AI system as high-risk under the AI horizontal instrument does not necessarily mean that the system should be qualified as high-risk under other sectoral acts.

[229] This is irrespective of whether the safety components are placed on the market independently from the product or not.

[230] NLF product legislation may also cover some AI systems which are to be considered products by themselves (e.g., AI devices under the Medical Device Regulations or AI safety components placed independently on the market which are machinery by themselves under the Machinery Directive).

[231] This can include standalone AI systems not covered by sectoral product safety legislation (e.g., recruitment AI system) or AI systems being safety components of products which are not covered by sectoral product safety legislation under point 1 and which are regulated only by the General Product Safety Directive. An initial list of high-risk AI systems covered by this point is detailed in Annex 5.4.

[232] These criteria include: a) the extent to which an AI system has been used or is about to be used; b) the extent to which an AI system has caused any of the harms referred to above or has given rise to significant concerns around their materialization; c) the extent of the adverse impact of the harm; d) the potential of the AI system to scale and adversely impact a plurality of persons or entire groups of persons; e) the possibility that an AI system may generate more than one of the harms referred to above; f) the extent to which potentially adversely impacted persons are dependent on the outcome produced by an AI system, for instance their ability to opt-out of the use of such an AI system; g) the extent to which potentially adversely impacted persons are in a vulnerable position vis-à-vis the user of an AI system; h) the extent to which the outcome produced by an AI system is reversible; i) the availability and effectiveness of legal remedies; j) the extent to which existing Union legislation is able to prevent or substantially minimize the risks potentially produced by an AI system.

Although evidence for individual legal challenges and breaches of fundamental rights is growing,[233] **robust and representative evidence** for harms inflicted by the use of AI is scarce due to **lack of data** and mechanisms to monitor AI as a set of emerging technology. To address these limitations, the initial assessment for the level of risk of the AI systems is based on the risk assessment methodology above[234] and on several other sources listed in Annex 5.4.

Based on the evidence in the problem definition, the sources and the methodology outlined above **Annex 5.3** includes the list of all sectoral product safety legislation that would be affected by the new initiative (new and old approach) and explains how the AI horizontal regulation would interplay with existing product safety legislation. For other AI systems that are mainly having fundamental rights implications, **a large pool of AI use cases has been screened[235]** by applying the criteria above with **Annex 5.4** identifying the initial list of high-risk AI systems proposed to be annexed to the horizontal instrument.[236] This classification of high-risk AI systems is largely consistent with the position of the EP with certain exceptions.[237]

Some flexibility would be provided to ensure that the list of high-risk AI systems is future proof and can respond to technological and market developments by **empowering the Commission within preliminary circumscribed limits to amend the list of specific use cases through delegated acts**.[238] Any change to the list of high-risk AI use cases would be based on the solid methodology described above, supporting evidence and expert advice.[239] To ensure legal certainty, future amendments would also require impact assessment following broad stakeholder consultation and there would always be a sufficient transitional period for adaptation before any amendments become binding for operators.

In contrast to the risk-based approach presented above, an alternative could be to place the assessment of the risk as a burden on the provider of the AI system and foresee in the legislation only general criteria for the risk assessment. This approach could make the risk assessment more

---

[233] See problem section 2.1.2.

[234] As an additional criteria, it could be envisaged that broader sectors are identified to select the high-risk AI use cases, as proposed in the White Paper and by the EP. The EP report proposes to base the risk assessment on exhaustive and cumulative high-risk sectors and of high-risk uses or purposes. Risky sectors would comprise employment, education, healthcare, transport, energy, public sector, defence and security, finance, banking and insurance. The Commission considers that broad sectors are not really helpful to identify specific high risk use cases. Applications may be low-risk even in high risk sectors (i.e. document management in the justice sector) or high-risk in sectors which are classified as low risk. On the other hand, more specific fields of AI applications could be envisaged to circumscribe the possible change in the use cases as another alternative.

[235] Final Draft of ISO/IEC TR 24030 identifies a list of 132 AI Use Cases that have been screened as a starting point by applying the risk assessment criteria and the methodology specified above. Other sources of use cases have been also considered such as those identified as high-risk in the EP report, in the public consultation on the White paper and based on other sources presented in Annex 5.4.

[236] See more details in Annex 5.4 on how the methodology has been applied and what are the identified high-risk use cases.

[237] The EP has identified as high-risk uses the following: recruitment, grading and assessment of students, allocation of public funds, granting loans, trading, brokering, taxation, medical treatments and procedures, electoral processes and political campaigns, some public sector decisions that have a significant and direct impact on the rights and obligations of natural or legal persons, automated driving, traffic management, autonomous military systems, energy production and distribution, waste management and emissions control. The proposed list by the Commission largely overlaps, it is summarised in Annex 5.4. It does not include algorithmic trading because this is regulated extensively by the Commission Delegated Regulation (EU) 2017/589. Use of AI for exclusive military purposes is considered outside the scope of this initiative given the implications for the Common Foreign and Security Policy regulated under Title V of TEU. Electoral processes and political campaigns are considered covered by the proposal for a Digital Services Act and the proposal for e-Privacy regulation. Brokering, taxation and emission controls were considered sufficiently covered by existing legislation and there is no sufficient evidence for harms caused by AI, but it could not be excluded that these might be included at a later stage with future amendments.

[238] The EP has also proposed amendments to the list of high risk uses cases via Commission's delegated acts.

[239] An expert group would support the work of the European AI Board and would regularly review the need for amendment of the list of high-risk AI systems based on evidence and expert assessment.

dynamic and capture high-risk use cases that the initial assessment proposed by the Commission may miss. This option has been, however, discarded because the economic operators would face significant legal uncertainty and higher burden and costs for understanding whether the new rules would apply in their case.

Stakeholders views: During the public consultation on the White Paper on AI, the limitation of requirements to high-risk AI applications was supported by 42.5% while 30.6% doubted such limitation. A majority (51%) of SMEs favoured limiting new compulsory requirements to high-risk applications, 21% opposed this. With regard to large businesses, a clear majority also favoured such an approach as well as the academic/research institutions. The stance of most civil society organisations differed from this view: more organisations opposed rather than supported this approach. At the same time, several organisations advocated fundamental or human rights impact assessments and cautioned against creating loopholes, for example regarding data protection, for low-risk applications. Of those stakeholders opposing the idea of limiting new requirements to high-risk AI applications, almost half were EU citizens (45%), with civil society and academic and research institutions being the second-largest groups (18% and 15%, respectively). For all these groups, this was higher than their share in the composition of the overall sample.

c) Horizontal requirements for high-risk AI systems and obligations on providers and users

The instrument would define **horizontal mandatory requirements for high-risk AI systems** that would have to be fulfilled for any high-risk AI system to be permitted on the Union market or otherwise put into service. The same requirements would apply regardless of whether the high-risk AI system is a safety component of a product or a stand-alone application with mainly fundamental rights implications (systems covered by both Annex 5.3. and Annex 5.4).

The requirements would be the same as in option 1 (incl. data, algorithmic transparency, traceability and documentation etc.), but operationalized by means of voluntary **technical harmonized standards**. In line with the principles of the New Legislative Framework, these standards would provide a legal presumption of conformity with requirements and constitute an important means to facilitate providers in reaching and demonstrating legal compliance.[240] The standards would improve consistency in the application of the requirements as compared to the baseline and ensure compatibility with Union values and applicable legislation, thus contributing to all 4 specific objectives. The reliance on harmonised standards would also allow the horizontal legal framework to remain sufficiently agile to cope with technological progress. While the legal framework would contain only high-level requirements setting the objectives and the expected outcomes, technological solutions for implementation would be left to more flexible market-driven standards that are updated on a regular basis to reflect technological progress and state-of-the-art. The governance mechanism of the European standardisation organisations who are usually mandated to produce the relevant harmonised standards would also ensure full consistency with ongoing and future standardisation activities at international level.[241]

Furthermore, the instrument would place **clear, proportionate and predictable horizontal obligations on providers of 'high-risk' AI systems** placing such system on the Union market as

[240] The AI legislation would be built as a New Legislative Framework (NLF) type legislation that is implemented through harmonised technical standards. The European Standardisation Organisations (CEN/CENELEC and ETSI) will adopt these standards on the basis of a mandate issued by the Commission and submit them to the Commission for possible publication in the Official Journal.

[241] While CEN/CENELEC and ETSI, as European Standardisation Organisations, are the addressees of Commission's Standardisation Requests in accordance with Regulation 2012/1025/EU, the Vienna Agreement signed between CEN and ISO in 1991 recognizes the primacy of international standards and aims at the simultaneous recognition of standards at international and European level by means of improved exchange of information and mutual representation at meetings. This usually ensures the full coordination between international and European process for standardisation. Moreover, other important international standardisation organisations, IEEE, and CEN/CENELEC have recently engaged in upscaling their level of collaboration and mutual cooperation.

well as on **users.**[242] Considering the various sources of risks and the AI specific features, responsibility would be attributed for taking reasonable measures necessary as the minimum to ensure safety and respect of existing legislation protecting fundamental rights throughout the whole AI lifecycle (specific objectives 1, 2, 3 and 4).

Figure 7: Obligations for providers of high-risk AI systems

Providers' Obligations	Description
Ensure compliance with the AI requirements	• Do conformity assessment to demonstrate compliance with AI requirements before the system is placed on the market (problem 1 to 5) • Re-assess the conformity in case of substantial modification to take into account the continuous learning capabilities (driver)
Registration	▪ Register AI systems (not safety components of products) in a public EU database that would improve legal certainty, enforceability of the rules and build public trust (problems 3 to 5)
Quality & risk management	▪ Implement a quality management system for achievement and maintenance of compliance. ▪ Test and validate the AI systems, assess and monitor risks and take appropriate mitigating measures (problems 1 to 5)
Post-market monitoring	▪ Implement a post-market monitoring system (incl. collect relevant data) ▪ Taking corrective and preventive action (incl. recalling or withdrawing the system from the market) (problems 1 to 4)
Reporting to competent authorities	▪ Report to authorities when a high-risk AI system presents a risk or serious incidents and breaches of fundamental rights obligations they become aware of (problems 1 to 3)

Figure 8: Obligations for users of high-risk AI systems

[242] Except where users use the high-risk AI system in the course of a personal (non-business) or transient activity e.g. travellers from third countries could use for example their own self-driving car and do not comply with the new obligations, while they are in Europe.

Users Obligations	Description
Human oversight	▪ Follow the instructions by the provider for human oversight and take all indicated measures to minimise residual risks (problems 1 to 2) ▪ Continuous monitoring of the AI system (driver)
Documentation	▪ Keep minimal documentation with input data in case of self-learning systems and automatically generated trails logs to allow for potential investigations (problems 3 and 4)
Data Protection Impact Assessment	▪ Use the information given by the provider as an input to perform a Data Protection Impact Assessment (when required by data protection law) (problem 2 to 4)

These clear and predictable requirements for high-risk AI systems and obligations placed on all AI value chain participants are mostly common practice for diligent market participants and would ensure **minimum degree of algorithmic transparency and accountability** in the development and use of AI systems. Without creating new rights, these rules would help to ensure that reasonable and proportionate measures are taken to avoid and mitigate the risks to fundamental rights and safety specific to AI systems and ensure that the same rules and rights in the analogue world apply when high-risk AI systems are used (problems 1 and 2). The requirements for algorithmic transparency and accountability, and trustworthy AI would be enforceable and effectively complied with (problem 3) and businesses and other operators would also have legal certainty on who does what and what are the good practice and state-of-the-art technical standards to demonstrate compliance with the legal obligations (problem 4). These harmonised rules across all sectors would also help to increase the trust of citizens and users that AI use is safe, trustworthy and lawful (problem 5) and prevent unilateral Member States actions that risk to fragment the market and to impose even higher regulatory burdens on operators developing or using AI systems (problem 6).

d) Minimal transparency obligations for non-high-risk AI systems

For all other non-high risk AI systems, the instrument would not impose any obligations or restrictions except for some minimal transparency obligations in two specific cases where people might be deceived (problem 2) which are not effectively addressed by existing legislation[243]. This would include:

- Obligation to inform people when interacting with an AI system (chatbot) in cases where individuals might believe that they are interacting with another human being;
- Label deep fakes except when these are used for legitimate purposes such as to exercise freedom of expression and subject to appropriate safeguards for third parties' rights.

These minimal transparency obligations would apply irrespective of whether the AI system is embedded in products or not. All other non-high-risk AI systems would be shielded from potentially diverging national regulations which would stimulate the creation of a single market for trustworthy AI and prevent the risk of market fragmentation for this substantial category of non-high-risk AI systems (problems 4, 5 and 6).

[243] Other use cases involving the use of AI that merit transparency requirements have also been considered (e.g., when a person is subject to solely automated decisions or micro-targeted), but these were discarded. This is because relevant transparency obligations already exist in data protection legislation (Articles 13 and 14 of the GDPR), in consumer protection law as well as in the proposals for the e-Privacy Regulation COM/2017/010 final - 2017/03 (COD) and the proposal for the Digital Services Act (COM/2020/825 final).

5.4.3. Enforcement of the horizontal instrument on AI

For the enforcement of the horizontal instrument, there are three options: a) ex post system; b) ex ante system; or c) a combination of ex ante and ex post enforcement.

a) Ex post enforcement of the horizontal instrument

Firstly, enforcement could rely exclusively on an **ex-post system for market surveillance** and supervision to be established by national competent authorities designated by the Member States.[244] Their task would be to control the market and investigate compliance with the obligations and requirements for all high-risk AI systems already placed on the market. Market surveillance authorities would have all powers under Regulation (EU) 2019/1020 on market surveillance, including *inter alia* powers to:

- follow up on complaints for risks and non-compliance;
- make on-site and remote inspections and audits of the AI systems;
- request documentations, technical specifications and other relevant information from all operators across the value chain, including access to source code and relevant data;
- request remedial actions from all concerned operators to eliminate the risks, or where the non-compliance or the risk persists - prohibit or order its withdrawal, recall from the market or immediate suspension of its use;
- impose sanctions for non-compliance with the obligations with proportionate, dissuasive and effective penalties;[245]
- benefit from the EU central registration database established by the framework as well as from the EU RAPEX system for the exchange of information among authorities on risky products.

Member States would have to ensure that all national competent authorities are provided with sufficient financial and human resources, expertise and competencies in the fields of AI, including fundamental rights and safety risks related to AI to effectively fulfil their tasks under the new instrument. The minimal transparency obligations for low-risk AI and the prohibited AI practices would also be enforced ex post. In order to avoid duplications, for high-risk AI systems which are safety components of products, covered by sectoral safety legislations, the ex-post enforcement of the horizontal instrument would rely on existing market surveillance authorities designated under those legislations (see more details in Annex 5.3).

The governance system would also enable cooperation between market surveillance authorities and other competent authorities supervising enforcement of existing Union and Member State legislation (e.g., equality bodies, data protection) as well as with authorities from other Member States. The mechanism for cooperation would also include new opportunities for exchange of information and joint investigations at national level as well as in cross border cases. All these new powers and resources for market surveillance authorities and mechanisms for cooperation would aim to ensure effective enforcement of the new rules and the existing legislation on safety and fundamental rights (problem 3).

b) Ex ante enforcement of the horizontal instrument

---

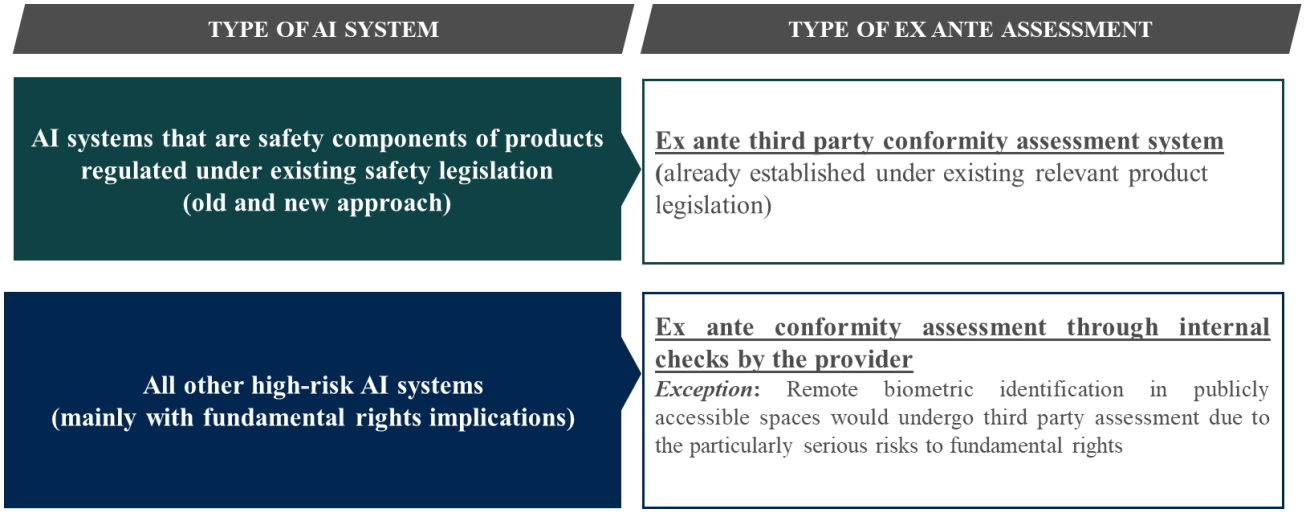
[244] To ensure consistency in the implementation of the new AI instrument and existing sectoral legislation, Member States shall entrust market surveillance activities for those AI systems to the national competent authorities already designated under relevant sectoral Union legislation, where applicable (e.g. sectoral product safety, financial service).

[245] Thresholds and criteria for assessment would be defined in the legal act to ensure effective and uniform enforcement of the new rules across all Member States. Fines would be in particular imposed for supplying incorrect, incomplete or false information and non-compliance with the obligations for ex ante conformity assessment and post market monitoring, failure to cooperate with the competent authorities etc.

Secondly, **ex ante conformity assessment procedures** could be made mandatory for high-risk AI systems in consistency with the procedures already established under the existing New-Legislative Framework (NLF) product safety legislation. After the provider has done the relevant conformity assessment, it should register stand-alone AI system with mainly fundamental rights implications[246] in an EU database that would be managed by the Commission. This would allow competent authorities, users and other people to verify if the high-risk AI system complies with the new requirements and to ensure enhanced oversight by the public authorities and the society over these systems (problems 3 to 5).

The ex-ante verification (through internal checks or with the involvement of a third-party) could be split according to the type of risks and level of interplay with existing EU legislation on product safety.

Figure 9: Types of ex ante conformity assessment procedures



In any of the cases above, **recurring re-assessments of the conformity** would be needed in case of substantial modifications to the AI systems (changes which go beyond what is pre-determined by the provider in its technical documentation and checked at the moment of the ex-ante conformity assessment).[247] In order to operationalise this approach for continuously learning AI systems and keep the administrative burden to a minimum, the instrument would clarify that: 1) a detailed description of pre-determined algorithm changes and changes in performance of the AI systems during their lifecycle (with information for solutions envisaged to ensure continuous compliance), should be part of the technical documentation and evaluated in the context of the ex-ante conformity assessment; 2) changes that have not been pre-determined at the moment of the initial conformity assessment and are not part of the documentation would require a new conformity assessment.

Harmonised standards to be adopted by the EU standardisation organisations would play a key role in facilitating the demonstration of compliance with the ex-ante conformity assessment obligations. For remote biometric identification systems or where foreseen by sectoral product safety legislation,[248] providers **could replace the third-party conformity assessment with an ex-ante conformity assessment through internal checks**, provided that harmonised standards exist and

[246] See footnote 231 and Annex 5.4.

[247] This approach is fundamentally in line with the idea of "pre-determined change control plan" developed and proposed by the US Food and Drug Administration (FDA) in the field of AI-based software in the medical field in a discussion paper produced in 2019. The effectiveness of the described approach is further reinforced by the fact that an obligation for post-market monitoring is set for providers of high-risk AI system, obliging them to collect data about the performance of their systems on a continuous basis after deployment and monitor it.

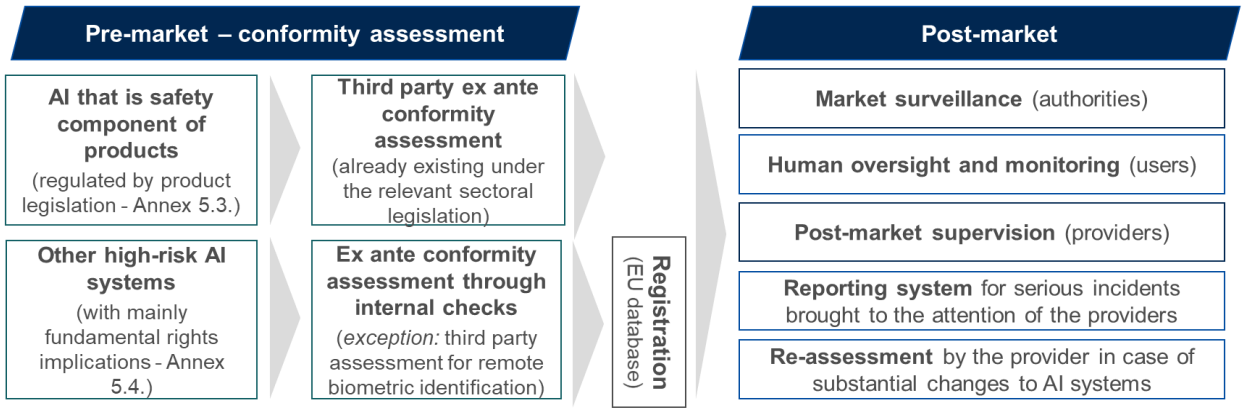
[248] More details on the interaction between the ex-ante enforcement system envisaged in the horizontal act and its interplay with product safety legislation can be found in section 8 of this impact assessment and Annex 5.3.

they have complied with those standards. Work on AI standardisation is already ongoing, many standards, notably of foundational nature, have already been produced and the preparation of many others is ongoing. The assumption of the Commission is that a large set of relevant harmonised standards could be available within 3-4 years from now that would coincide with the timing needed for the legislative adoption of the proposal and the transitional period envisaged before the legislation becomes applicable to operators.

c) Combination of an ex ante and ex post system of enforcement

As a third option, the **ex-ante enforcement could be combined with an ex-post system** for market surveillance and supervision as described above. Since this option most effectively addressed problems 1, 2 and 3, the combination between ex post and ex ante enforcement system has been chosen and the other alternatives discarded.[249]

Figure 10: Compliance and enforcement system for high-risk AI systems



d) Alternative policy choices for the system of enforcement and obligations

Four alternative policy choices for the ex-ante obligations and assessment have also been considered: i) distinction between safety and fundamental rights compliance; ii) ex-ante conformity assessment through internal checks, or third party conformity assessment for all high-risk AI systems; iii) registration of all high-risk AI systems in the EU database or no registration at all, or iv) additional fundamental rights/algorithmic impact assessment.

i) Distinction between safety and fundamental rights compliance

A first alternative approach could be to apply an NLF-type ex-ante conformity assessment only to AI systems with safety implications, while **for AI systems posing fundamental rights risks** (Annex 5.4.) the instrument could envisage documentation and information requirements for providers and **more extensive risk-management obligations for users**. This approach was however discarded. Given the importance of the design and development of the AI system to ensure its trustworthiness and compliance with both safety and fundamental rights, it is appropriate to place responsibilities for assessment on the providers. This is because users are already bound by the fundamental rights legislation in place, but there are gaps in the material scope of the existing legislation as regards the obligations of producers as identified in problem 3 of this impact assessment.

ii) Ex-ante conformity assessment through internal checks or third party conformity assessment for all high-risk AI systems

[249] This choice is also consistent with the EP position which envisages ex ante certification and ex post institutional control for compliance.

A second alternative would be to apply ex-ante conformity assessment through internal checks for all high-risk AI systems or to foresee third-party involvement for all high-risk AI systems. On the one hand, a comprehensive **ex-ante conformity assessment through internal checks,** combined with a strong ex-post enforcement, could be an effective and reasonable solution given the early phase of the regulatory intervention and the fact the AI sector is very innovative and expertise for auditing is just about to be accumulated. The assessment through internal checks would require a full, effective and properly documented ex ante compliance with all requirements of the regulation and compliance with a robust quality and risk management systems and post-market monitoring. Equipped with adequate resources and new powers, market surveillance authorities would also ensure ex officio enforcement of the new rules through systematic ex-post investigations and checks; request remedial actions or withdrawal of the risky or non-compliant AI systems from the market and/or impose financial sanctions. On the other hand, high-risk AI systems that are safety components of products are, by definition, already subject to the **third party conformity assessment** foreseen for the relevant product under sectoral product safety legislation (see more details in Annex 5.3), so the new horizontal initiative should not disrupt but rather integrate into that system.

In conclusion, the combination of the two alternatives reflects regulatory and practical considerations and results in an appropriate mix of enforcement tools to deal respectively with safety and fundamental right risks.[250]

iii) Require registration of all high-risk AI systems in the EU database or no registration at all

As to the registration obligation applicable only to stand-alone AI systems with mainly fundamental rights implications (Annex 5.4), an alternative policy choice would be to require registration of any high-risk AI system, including systems that are safety components of products or devices. However, this option was discarded because this latter category of AI systems might already be subject to registration according to the existing product safety legislation (e.g. medical device database) and duplication of databases should be avoided. Furthermore, in the scenario where sectoral safety legislation does not establish a registration obligation for the products, the registration in a central database of high-risk AI systems that are components of products would prove to be of limited value for the public and the market surveillance authorities given that the product as a whole is not subject to central registration obligations.

A second alternative would be not to require registration even for the high-risk AI systems with fundamental rights implications, but this policy choice was also discarded. The reason is that without such a public database the specific objectives of the initiative would be compromised, particularly in relation to increasing public trust and the enforceability of the existing law on fundamental rights (problems 3 and 5). Keeping the registration obligation for these systems with fundamental rights implications is thus justified given the need for increased transparency and public oversight over these systems.[251]

iv) Require an additional fundamental rights/algorithmic impact assessment

Another alternative for high-risk AI systems with fundamental rights implications would be to require a **fundamental rights impact assessments/algorithmic impact assessments** as implemented in Canada and the U.S. and recommended by some stakeholders, the Council of

[250] Nevertheless, the conformity assessment rules of many existing relevant sectorial legislations would allow providers of high-risk AI systems that are safety components of products to carry out a conformity assessment through internal checks if they have applied harmonised standards.

[251] This would also contribute to the principle of societal well-being endorsed by the OECD and the HLEG and follows the recommendation of the Council of Europe CM/Rec(2020)1 to increase transparency and oversight for AI systems having significant fundamental rights implications.

Europe[252] and the Fundamental Rights Agency.[253] However, this was also discarded, because users of high-risk AI systems would normally be obliged to do a Data Protection Impact Assessment (DPIA) that already aims to protect a range of fundamental rights of natural persons and which could be interpreted broadly, so new regulatory obligation was considered unnecessary.

Stakeholders views: During the public consultation on the White Paper on AI, 62% of respondents supported a combination of ex-post and ex-ante market surveillance systems. 3% of respondents support only ex-post market surveillance. 28% supported third party conformity assessment of high-risk applications, while 21% of respondents support ex-ante self-assessment. While all groups of stakeholders had the combination of ex-post and ex-ante market surveillance systems as their top choice, industry and business respondents preferred ex-ante self-assessment to external conformity assessment as their second best choice.

5.4.4. Governance of the horizontal instrument on AI

The governance system would include enforcement at national level with a cooperation mechanism established at EU level.

At national level, Member States would designate competent authorities responsible for the enforcement and supervision of the new rules and the ex post market surveillance. As explained in details above, they should be provided with competences to fulfil their tasks in an effective manner, ensuring that they have adequate funding, technical and human resource capacities and mechanisms to cooperate given that a single AI system may fall within the sectoral competences of a number of regulators or some AI systems may be currently not supervised at all. The new reporting obligations for providers to inform competent authorities in case of incidents and breaches of fundamental rights obligations of which they have become aware would also significantly improve the effective enforcement of the rules (problem 3).

At EU level, coordination would be ensured through a mechanism for cross-border investigations and consistency in implementation across Member States and the establishment of a dedicated EU body (e.g. EU Board on AI)[254] responsible for providing uniform guidance on the new rules.[255] The establishment of an AI Board is justified by the need to ensure a smooth, effective and uniform implementation of the future AI legislation across the whole EU. Without any governance mechanism at EU level, Member States could interpret and apply very differently the new rules and would not have a forum to reach consistency and cooperate. This would fail to enhance governance and effective enforcement of fundamental rights and safety requirements applicable to AI systems (problem 3). Eventually, the divergent and ineffective application of the new rules would also lead to mistrust and lower level of protection, legal uncertainty and market fragmentation that would also endanger specific objectives 1, 3 and 4. Since there is no other body at EU level that encompasses the full range of competences to regulate AI across all different sectors in relation to both fundamental rights and safety, establishing a new EU body is justified.

5.4.5. Additional measures to support innovation

In line with specific objective 3, Option 3 would also envisage additional measures to support innovation including: a) AI regulatory sandboxing scheme and b) other measures to reduce the regulatory burden and support SMEs and start-ups.

a) AI regulatory sandboxing scheme

[252] Council of Europe, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems.* 2020.

[253] European Agency for Fundamental Rights, *Getting The Future Right – Artificial Intelligence and Fundamental Rights*, 2020.

[254] The AI Board would be an independent EU 'body' established under the new instrument. Its status would be similar to the European Data Protection Board.

[255] This has also been requested by the European Parliament resolution of 20 October 2020 (2020/2012(INL)).

The horizontal instrument would provide the possibility to create **AI regulatory sandboxes** by one or more competent authorities from Member States at national or EU level. The objective would be to enable **experimentation and testing of innovative AI technologies, products or services for a limited time before their placement on the market** and pursuant to a specific testing plan under the direct supervision by competent authorities ensuring that appropriate safeguards are in place.[256] **Through direct supervision and guidance by competent authorities,** participating providers would be assisted in their efforts to reach legal compliance with the new rules, benefitting from increased legal certainty on how the rules should apply to their concrete AI project (problem 4). This would be without prejudice to the powers of other supervisory authorities who are not associated to the sandboxing scheme.

The instrument would set clear limits for the experimentation. **No derogations or exemptions from the applicable legislation would be granted**, taking into account the high risks to safety and fundamental rights and the need to ensure appropriate safeguards.[257] Still, the competent authorities would have certain flexibility in applying the rules within the limits of the law and within their discretionary powers when implementing the legal requirements to the concrete AI project in the sandbox.[258] Any significant safety risks or adverse impact on fundamental rights identified during the testing of such systems should result in immediate rectification and, failing that, in the suspension of the system until such rectifications can take place.[259]

The regulatory sandboxes would foster innovation and increase legal certainty for companies and other innovators giving them a quicker access to the market, while minimising the risks for safety and fundamental rights and fostering effective compliance with the legislation through authoritative guidance given by competent authorities (problems 1, 2, 3 and 4). They would also provide regulators with new tools for supervision and hands-on experience to detect early on emerging risks and problems or possible need for adaptations to the applicable legal framework or the harmonised technical standards (problem 3). Evidence from the sandboxes would also help national authorities identify new high-risk AI use cases that would further inform the regular reviews by the Commission of the list of high-risk AI systems to amend it, as appropriate.

b) Other measures to reduce the regulatory burden and support SMEs and start-ups

To further reduce the regulatory burden on SMEs and start-ups, the **national competent authorities** could envisage additional measures such as provision of priority access to the AI regulatory sandboxes, specific awareness raising activities tailored to the needs of the SMEs and start-ups etc.

Notified bodies should also take into account the specific interests and needs of SMEs and start-ups when setting the **fees for conformity assessment** and reduce them proportionately.

---

[256] See for a similar definition Council Conclusions on regulatory sandboxes and European Commission, TOOL #21. Research & Innovation, Better Regulation Toolbox; European Commission; 6783/20 (COM (2020)103).

[257] See also Council Conclusions on regulatory sandboxes which emphasize the need to always respect and foster the precautionary principle and ensure existing levels of protection are respected. While under certain regulatory sandboxes there is a possibility to provide complete derogations or exemptions from the existing rules, this is not considered appropriate in this context given the high risks to safety and fundamental rights. Similar approach has also been followed by other competent authorities establishing sandboxes in the financial sector where the sandbox is rather used as a tool to apply flexibility permitted by law and help reach compliance in an area of legal uncertainty instead of disapplication of already existing Union legislation. See in this sense ESMA, EBA and EIOPA Report FinTech: Regulatory sandboxes and innovation hubs, 2018.

[258] For example, how to determine when the AI specific application is sufficiently accurate, robust or transparent for its intended purpose, whether the established risk management system and quality management systems are proportionate, in particular for SMEs and start-ups, etc.

[259] See in this sense also Council of Europe Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies).

As part of the implementing measures, **Digital Innovation Hubs and Testing Experimentation Facilities** established under the Digital Europe Programme could also provide support as appropriate. This could be achieved by providing, for example, by relevant training to providers on the new requirements and upon request, providing relevant technical and scientific support as well as testing facilities to providers and notified bodies in order to support them in the context of the conformity assessment procedures.

Stakeholders views: Stakeholders suggested different measures targeted at fostering innovation in the public consultation on the White Paper. For example, out of the 408 position papers that were submitted, at least 19 discussed establishing regulatory sandboxes as one potential pathway to better allow for experimentation and innovation under the new regulatory framework. Generally, at least 19 submissions cautioned against creating regulatory burdens that are too heavy for companies and at least 12 submissions highlighted the benefits of AI as a factor to be taken into account when contemplating new regulation, which might create obstacles in reaping those benefits. At least 12 Member States supported regulatory sandboxes in their national strategies. The Council Conclusion on regulatory sandboxes (13026/20) also highlight that regulatory sandboxes are increasingly used in different sectors, can provide an opportunity for advancing regulation through proactive regulatory learning and support innovation and growth of all businesses, especially SMEs.

5.5. Option 3+: Horizontal EU legislative instrument establishing mandatory requirements for high-risk AI applications + voluntary codes of conduct for non-high risk applications

Option 3+ would combine mandatory requirements and obligations for high-risk AI applications as under option 3 with voluntary codes of conduct for non-high risk AI.

Table 6.4. Summary of Option 3+

Nature of act	Option 3 + code of conducts non high-risk AI
Scope	Option 3 + voluntary codes of conduct non-high-risk AI
Content	Option 3 + industry-led codes of conduct for non-high-risk AI
Obligations	Option 3 + commitment to comply with codes of conduct for non-high-risk AI
Ex ante enforcement	Option 3 + self- assessment for compliance with codes of conduct for non-high-risk AI
Ex post enforcement	Option 3 + unfair commercial practice in case of non-compliance with codes
Governance	Option 3 + without EU approval of the codes of conduct

Under this option, the Commission would encourage industry associations and other representative organisations to adopt voluntary codes of conduct so as to allow providers of all non-high-risk applications to voluntarily comply with similar requirements and obligations for trustworthy AI. These codes could build on the existing self-regulation initiatives described in the baseline scenario and adapt the mandatory requirements for Trustworthy AI to the lower risk of the AI system. It is important to note that the obligatory minimal transparency obligations for non-high-risk AI systems under option 3 would continue to apply simultaneously with the voluntary codes of conduct.

The proposed system of voluntary codes of conduct would be light for companies to subscribe and not include a 'label' or a certification of AI systems. Combination with a voluntary labelling scheme for low-risk AI was discarded as an option because it could be still too complex and costly for SMEs to comply with. Furthermore, a separate label for trustworthy AI may create confusion with the CE label that high-risk AI systems would obtain under option 3. Under a voluntary labelling scheme, it would also be very complex and lengthy to create standards suitable for a potentially very high number of non-high-risk AI systems. Last but not least, such a voluntary labelling scheme for non-high-risk AI also received mixed reactions in the stakeholder consultation.

Stakeholders views: During the public consultation on the White Paper on AI, 50.5% of respondents found voluntary labelling useful or very useful for non-high-risk application, while another 34% of respondents did not agree with that approach. Public authorities, industry and business and private citizens were more likely to agree, while non-governmental organisations were divided. The Council conclusions of 9 June 2020 specifically called upon the Commission to include a 'voluntary labelling scheme that boosts trust and safeguards security and safety'. In a recent non-paper, representatives from ministries of 14 Member States call for a 'voluntary European labelling scheme' and in its recent resolution the European Parliament also envisaged it for non-high risk AI systems.

To reap the benefits of a voluntary framework for non-high risk AI, Option 3+ proposes instead that the Commission would encourage the providers of non-high risk AI systems to subscribe to and implement codes of conduct for Trustworthy AI developed by industry and other representative associations in Member States.

These industry-led codes of conduct for trustworthy AI could integrate and operationalise the main principles and requirements as envisaged under Options 1 and 3. The codes of conduct may also include other elements for Trustworthy AI that have not been included in the requirements and the compliance procedures under option 1 and 3 (e.g., proposed by HLEG, EP or Council of Europe in relation to diversity, accessibility, environmental and societal-well-being, fundamental rights or ethical impact assessments etc.). Before providers could give publicity to their adherence to a code of conduct, they should undergo the self-assessment procedure established by the code of conduct to confirm compliance with its terms and conditions. False or misleading claims that a company is complying with a code of conduct should be considered unfair commercial practices.

The Commission would not play any active role in the approval or the enforcement of these codes and they would remain entirely voluntary. As part of the review clause of the horizontal instrument, the Commission would evaluate the proposed scheme for codes of conduct for non-high risk AI and, building on the experience and the results, propose any necessary amendments.

5.6. Option 4: Horizontal EU legislative instrument establishing mandatory requirements for all AI applications, irrespective of the risk they pose

Under this option, the same requirements and obligations as the ones for option 3 would be imposed on providers and users of AI systems, but this would be applicable for all AI systems irrespective of the risk they pose (high or low).

Table 6.5. Summary of Option 4: Horizontal act for all AI systems

Nature of act	A single binding **horizontal act**, **applicable to all AI**
Scope	OECD definition of AI; applicable to all AI systems without **differentiation between the level of risk**
Content	Same as Option 3, but applicable to all AI systems (irrespective of risk)
Obligations	Same as Option 3, but applicable to all AI systems (irrespective of risk)
Ex ante enforcement	Same as Option 3, but applicable to all AI systems (irrespective of risk)
Ex post enforcement	Same as Option 3, but applicable to all AI systems (irrespective of risk)
Governance	Same as Option 3, but applicable to all AI systems (irrespective of risk)

5.7. Options discarded at an early stage

No options were discarded from the outset. However, in analysing specific policy options certain policy choices were made (i.e. sub-options within the main option. This selection of sub-options is summarized in Table 7 below.

Table 7: Summary of selected and discarded sub-options

POLICY OPTIONS	SELECTED SUB-OPTION	DISCARDED ALTERNATIVE SUB-OPTIONS
Relevant for Option 1, 3, 3+	**OECD Definition of AI** (technology neutral) – pp. 40 and 48	Technology-specific definition (e.g. Machine learning)

and 4		
Relevant for all Options	**5 requirements** (proposed in the White Paper on AI) – p. 40	Other discarded requirements: • Environmental and societal well-being • Social responsibility and gender equality • Privacy • Effective redress
Relevant for Option 2, 3, 3+ and 4	Prohibitions of certain use of **remote biometric identification in public spaces** + additional safeguards and limitations for the permitted use (p. 45)	Other discarded alternatives: • Complete prohibition of remote biometric identification systems in publicly accessible spaces • Application of the requirements for trustworthy AI (as per option 1) without additional restrictions on the use
Relevant for Option 2, 3, 3+ and 4	**Prohibition of other harmful AI practices:** • Manipulative and exploitative AI • General purpose social scoring (p. 46)	Complete prohibition of other AI uses: • Other manipulative and exploitative AI uses (e.g. profiling and micro-targeting of voters, consumers etc.) • Predictive policing • AI used for allocation of social security benefits • AI used in border and migration control • Individualised risk assessments in the criminal law context
Relevant for Option 3 and 3+	**List of high-risk AI systems** identified by the legislator (pp.49-50)	Each provider is obliged to assess if its AI system is high-risk or not on the basis of criteria defined by the legislator
Relevant for Option 3 and 3+	**AI systems included in the list of high-risk AI** (Annex 5.4)	A larger pool of AI use cases has been screened and discarded (drawing from EP proposals, ISO report, stakeholder consultation and additional research)
Relevant for Option 3 and 3+	**Transparency requirements for non-high-risk AI** in relation to chatbots and labelling of deep fakes	Other AI uses such as use of automated decision affecting people, profiling and micro-targeting of individuals
Relevant for Option 1, 3, 3+ and 4	**Ex ante and ex post enforcement** (p.42 and pp.53-55)	Only ex ante or only ex post enforcement
Relevant for Option 3, 3+ and 4	**Ex ante conformity assessment** (split between assessment though internal checks and third party conformity assessment) + **registration in an EU database** of high-risk AI systems with fundamental rights implications (p. 54)	Discarded alternatives: • Distinguish between safety and fundamental rights ex ante assessments • Ex ante assessment through internal checks for all high-risk AI systems or third party conformity assessment for all high-risk AI systems • Registration in the EU database of all high-risk AI systems or no database at all • Additional fundamental rights/algorithmic impact assessment
Relevant for Option 3, 3+ and 4	Option1: **Governance system** with national competent authorities + light mechanism for EU cooperation (p. 42) Option 3, 3+ and 4: Governance system with national competent authorities + European AI Board (p. 57)	No cooperation at EU level
Relevant for Option 3+	Option 3 + voluntary codes of conduct for non-high-risk AI (pp. 59-60)	Option 3 + voluntary labelling for non-high-risk AI

6. WHAT ARE THE IMPACTS OF THE POLICY OPTIONS?

The policy options were evaluated against the following economic and societal impacts, with a particular focus on impacts on fundamental rights.

6.1. Economic impacts

6.1.1. Functioning of the internal market

The impact on the internal market depends on how effective the regulatory framework is in preventing the emergence of obstacles and fragmentation by mutually contradicting national initiatives addressing the problems set out in section 2.1.2.4.

Option 1 would have a limited impact on the perceived risks that AI may pose to safety and fundamental rights. A labelling scheme would give information to businesses that would wish to deploy AI and consumers when purchasing or using AI applications, thus redirecting some demand from non-labelled products to labelled products. However, the extent of this shift - and hence the incentive for both AI suppliers to adopt the voluntary label - is uncertain. Therefore, at least in some Member States, public opinion is expected to continue to put pressure towards a legislative solution, possibly leading to at least partial fragmentation.

Option 2 would address the risk of fragmentation for those classes of applications for which specific legislation is introduced. Since these are likely to be the ones where concerns have become most obvious and most urgent, it is possible that Member States will refrain from additional legislation. Where they see a need for supplementary action, they could bring it to the attention of the EU to make further EU-wide proposals. However, some Member States may consider that a horizontal approach is also needed and pursue such an approach at national level.

Options 3 and 3+ effectively address the risks set out in section 2.2. in all application areas which are classified sensitive or 'high-risk, with option 3 in addition also ensuring a European approach for low-risk applications.[260] Hence, no Member State will have an incentive to introduce additional legislation. Where Member States would wish to classify an additional class of applications as high-risk, they have at their disposal a mechanism (the possibility to amend the list in the future) to include this class into the regulatory framework. Only if a Member State wishes to include an additional class of applications, but fails to convince the other Member States, could there be a potential risk for unilateral action. However, since the most risky application fall within the scope of the regulatory framework, it is unlikely that Member States would take such a step for a class of applications at the margin of riskiness.

Option 4 addresses the risks created by AI in all possible applications. Thus, Member States are unlikely to take unilateral action.

6.1.2. Impact on uptake of AI

Currently, in the European Union the share of companies use AI at a scale is 16% lower than in the US, where growth continues.[261] There is thus ample scope to accelerate uptake of AI in the EU. Faster uptake by companies would yield significant economic benefits. As an example, by 2030, companies rapidly adopting AI are predicted to gain about 122% in economic value (economic

[260] With regard to so-called 'old approach' products like cars the introduction of specific requirements for AI will require changes in sectorial legislation. Such modifications should follow the principles of the horizontal legislation. As these sectorial legislations are regularly updated, a timely insertion of the specific AI requirements can be expected. Therefore, as well in this area, MS should not be in need to legislate unilaterally.

[261] McKinsey Global Institute, *Notes from the AI frontier: tackling Europe's gap in digital and AI*, 2019.

output minus AI-related investment and transition costs). In contrast, companies only slowly adapting AI could lose around 23% of cash flow compared with today.[262]

The regulatory framework can enhance the uptake of AI in two ways. On the one hand, by increasing users' trust it will lead to a corresponding increase in the demand by AI using companies. On the other hand, by increasing legal certainty it will make it easier for AI suppliers to develop new attractive products which users and consumers appreciate and purchase.

Option 1 can increase users' trust for those AI systems that have obtained the label. However, it is uncertain how many applications will apply, and hence the increase in user's trust remains uncertain. Also, users will have more trust when they can rely on legal requirements, which they can enforce in courts if need be, than if they have to rely on voluntary commitments. Regarding legal certainty, option 1 does provide this neither to AI suppliers nor to AI using companies, since it has no direct legal effect.

Option 2 would enhance users' trust in those types of AI applications to which regulations apply. However, regulation, whether new or amended existing legislation, would only occur once concerns have emerged, and may thus be delayed. Moreover, it would provide AI suppliers and AI using companies with legal certainty only regarding these particular classes of applications and mightlead to inconsistencies in the requirements imposed by sectorial legislations, hampering uptake.

Option 3 would enhance users' trust towards the high-risk cases, which are those where trust is most needed. Hence, its positive effect on uptake would be precisely targeted. Moreover, it would not allow a negative reputation to build up in the first place, but ensure a positive standing from the outset. Option 3+ would in addition allow further trust building by AI suppliers and AI using companies and individuals where they see fit. Option 4 would have the same effect on trust for the high-risk cases but would in addition increase trust for many applications where this would have marginal effect. For options 3, 3+ and 4 legal certainty would rise, enabling AI suppliers to bring new products more easily to market.

6.1.3. Compliance costs and administrative burdens[263]

Stakeholders views: With regard to costs arising due to regulation, more than three quarters of companies did not explicitly mention such costs. However, at least 14% of SMEs and 13% of large companies addressed compliance costs as a potential burden resulting from new legislation in their position papers. Further at least 10% and 9%, respectively, also mentioned additional administrative burdens tied to new regulation in this context.

The costs are calculated relative to the baseline scenario not taking into account potential national legislation. However, **if the Commission does not take action, Member States would be likely to legislate against the risks of artificial intelligence. This could lead to similar or even higher costs if undertakings were to comply with distinct and potential mutually incompatible national requirements.**[264]

[262] McKinsey Global Institute, *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*, 2018.

[263] Administrative burdens mean the costs borne by businesses, citizens, civil society organizations and public authorities as a result of administrative activities performed to comply with information obligations included in legal rules; compliance costs the investments and expenses that are faced by businesses and citizens in order to comply with substantive obligations or requirements contained in a legal rule. (Better Regulation tool 58)

[264] For the estimates related to the European Added Value see e.g. European Parliamentary Research Service, European added value assessment: European framework on ethical aspects of artificial intelligence, robotics and related technologies, 2020. This analysis suggest that a common EU framework on ethics (as compared to fragmented national actions) has the potential to bring the European Union €294.9 billion in additional GDP and 4.6 million additional jobs by 2030.

The estimates in this section are taken from Chapter 4 *"Assessment of the compliance costs generated by the proposed regulation on Artificial Intelligence*" of the *Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe"*.[265]

The costs estimations are based on a Standard Cost Model, assessing required time (built on a reference table from Normenkontrollrat (2018)) and evaluating costs by the reference hourly wage rate indicated by Eurostat for the Information and Communication sector (Sector J in the NACE rev 2 classification). The cost estimation is built upon time expenditures of activities induced by the selected requirements for an average AI unit of an average firm in order to arrive at meaningful estimates. In practice, AI systems are very diverse, ranging from very cheap to very expensive systems.

This methodology is regularly used to estimate the costs of regulatory intervention. It assumes that businesses need to adopt measures to comply with every requirement set out. Thus, it represents the theoretical maximum of costs. **For companies that already fulfil certain specific requirements, the corresponding cost for these specific requirements would in practice be zero**, e.g. if they already ensure accuracy and robustness of their system, the costs of this requirement would be zero.

Option 1 would create comparable compliance costs to those of the regulatory approach (see option 3), assuming the voluntary labelling contains similar requirements. The administrative burden per AI application would likely be lower, as the documentation required without conformity assessment will be lighter. The aggregate costs would then depend on the uptake of the labelling, i.e. what share of AI applications would apply to obtain the label and can therefore range from 0 to a theoretical maximum of around €3 billion (see calculations in option 3). Presumably, only those applications would apply to obtain the label that would benefit from user trust or those that process personal data (thus excluding industrial applications), so it would be less than 100%. At the same time, there are applications that would benefit from user trust but are not high-risk applications as in option 3. Hence, the share would be above the share of option 3. However, this estimate depends on the success of the label, i.e. a general recognition of the label. Note that companies would only accept the administrative burden if they considered the costs lower than the benefits.

For **option 2**, the compliance costs and administrative burden would depend on the specific subject regulated and are thus impossible to estimate at this point in time. Since only one class of applications will fall in the scope of each regulation, the share of total AI applications covered will be much smaller even than the share of high-risk applications. It is quite possible that the requirements may be more stringent, since these will be the most controversial applications. In any case, a business developing or using several of these classes of applications (e.g. remote biometric identification in publicly accessible spaces and deep fakes), would not be able to exploit synergies in the certification processes.

In **option 3** there would be five sets of requirements, concerning data, documentation and traceability, provision of information and transparency, human oversight and robustness and accuracy. As a first step, it is necessary to identify the maximum costs of the measures necessary to fulfil each of these requirements and adding them up gives total compliance costs per AI application (Table 8). However, **economic operators would already take a certain number of measures even without explicit public intervention**. In particular, they would have to still ensure that their product actually works, i.e. robustness and accuracy. This cost would therefore only arise for companies not following standard business procedures (Table 8a). For the other requirements, operators would also take some measures by themselves, which would however not be sufficient to comply with the legal obligations. Here, they may build additional measures on top of the existing in order to achieve full compliance. In a second step, it is therefore necessary to estimate which share of these costs would be additional expenditure due to regulatory requirements. In addition, it

[265] ISBN 978-92-76-36220-3

should be noted that human oversight represents overwhelmingly an operating cost which arises to the user, if at all (depending on the use case)(Table 8b).

Table 8: Maximum fixed compliance costs and administrative burden for AI suppliers

Compliance costs regarding data	€2 763
Administrative burden regarding documentation and traceability	€4 390
Administrative burden regarding provision of information	€3 627

Table 8a: Additional costs for companies not following state-of-the-art business procedures

Compliance costs regarding robustness and accuracy	€10 733

Table 8b: Operating costs for AI users

Compliance costs regarding human oversight	€7 764

In **option 3**, the theoretical maximum compliance costs and administrative burden of **algorithmic transparency and accountability** per AI application development (the sum of the three sets of requirements for AI suppliers) amount to around €10 000 for companies following standard business procedures.

In accounting for the share of costs which correspond to a normal state-of-the-art business operation ("business-as-usual factor"), one can expect a steep learning curve, as companies will integrate the actions they take to fulfil the requirements with the actions they take for business purposes (for instance add a testing for non-discrimination during the regular testing of an AI application). As a result the adjusted maximum costs taking into account the business-as-usual can be estimated at around two thirds of the theoretical costs, i.e. on average at around €6 000 - €7 000 by 2025.

Since the average development cost of an AI system is assumed in the Standard Cost Model to be around €170 000 for the purposes of the cost calculation, this would amount to roughly 4-5%. These costs are for the software alone; when AI is embedded with hardware, the overall costs increase significantly as the project becomes much more expensive and the AI compliance costs as a share of total costs become correspondingly smaller.

It is useful to relate the estimated costs of this option with compliance costs and administrative burdens of other recent initiatives.[266] Although these costs are not strictly comparable (AI costs are per product, GDPR costs are for the first year, VAT costs are per year), they nevertheless give an idea of the order of magnitude. For example, regarding GDPR, a recent study[267] found that 40% of SMEs spent more than €10 000 on GDPR compliance in the first year, including 16% that spent more than €50 000. Another report[268] found that an average organization spent 2,000-4,000 hours in meetings alone preparing for GDPR. Another benchmark are VAT registration costs, which amount to between €2500 and €4000 annually per Member State, resulting in €80-90 000 for access to the entire EU market.

[266] Given the assumption of the cost estimation that an average AI application costs €170 000 to develop, it is reasonable to assume that most SMEs only produce one or maximum two AI applications per year

[267] GDPR.EU, *Millions of small businesses aren't GDPR compliant, our survey finds*. Information website, 2019.

[268] Datagrail, *The Cost of Continuous Compliance, Benchmarking the Ongoing Operational Impact of GDPR and CCPA*, 2020.

With estimates for AI investment in the EU by 2025 in the range of €30 billion to €65 billion, 4-5% of the upper estimate of €65 billion translate into a maximum estimate of aggregate compliance costs for all AI applications of about €3 billion in 2025. However**,** since **option 3** only covers high-risk applications, one has to estimate the share of AI applications which would fall under the scope of the obligations, and adjust the costs accordingly. At this stage, it is not possible to estimate the costs precisely, since the legislator has not yet decided the list of high-risk applications. Nevertheless, given that in this option high-risk applications are based on exceptional circumstances, one could estimate that no more than 5% to 15% of all applications should be concerned by the requirements. Hence the corrected maximum aggregate compliance costs for high-risk AI applications would be no more than 5% to 15% of the maximum for all applications, i.e. €100 million to €500 million.

However, in practice the compliance costs and administrative burden for high-risk applications are likely to be lower than estimated. That is because the business-as-usual factor mentioned above has been calculated for an average AI application. For high-risk applications, companies would in any case have to take above-average precautions. Indeed, faced with sceptical or hostile parts of public opinion, companies will have to pay attention to issues like data representativeness regardless of legal obligations. As a result, the additional costs generated by the legislation would in practice be smaller than the estimated maximum.

For AI users, the costs for documentation would be negligible, since it will mostly rely on in-built functions such as use logs that the providers have installed. In addition, there would be the annual cost for the time spent on ensuring human oversight where this is appropriate, depending on the use case. This can be estimated at €5000 – €8000 per year (around 0.1 FTE).

In **option 3+** the additional aggregate costs would depend on how many companies submit their applications to a code of conduct; if the requirements of the code of conduct were the same as for high-risk applications, the maximum aggregate compliance costs and administrative burden of option 3+ would lie between €100 million to €500 million and again a theoretical maximum of €3 billion. However, it is likely that the codes of conduct will have fewer requirements, since they cover less risky applications. Aggregate compliance costs are thus likely to be lower.

In **option 4**, since all AI applications have to comply with the obligations, 4-5% per AI application with an upper estimate of €65 billion would correspond to a maximum estimate of aggregate compliance costs for of about €3 billion in 2025.

Verification costs

In addition to meeting the requirements, costs may accrue due to the need to demonstrate that the requirements have been met.

For **option 1**, ex-ante conformity assessment through internal checks would be combined with ex-post monitoring by the competent authorities. The internal checks would be integrated into the development process. No external verification costs would therefore accrue.

Under **option 2**, rules for verification would be laid down in the specific legislative acts and are likely to be different from one use case to another. Thus, one cannot estimate them at this stage.

In **option 3**, for AI systems that are safety components of products under the new legislative approach, the requirements of the new framework would be assessed as part of the already existing conformity assessments which these products undergo. For remote biometric identification systems in publicly accessible places, a new ex-ante third party conformity assessment procedure would be created. Provided that harmonised standards exist and the providers have applied those standards, they could replace the third-party conformity assessment with an ex-ante conformity assessment through internal checks applying the same criteria. All other high-risk applications would equally be assessed via ex-ante conformity assessments through internal checks applying the same criteria.

Third-party conformity assessment for AI applications comes in two elements: the assessment of a quality management system the provider would have to implement, which is already a common feature in product legislation, and the assessment of the technical characteristics of the individual AI system itself (so-called EU technical documentation assessment).

Companies supplying products that are third-party conformity assessed already have a quality management system in place. Companies supplying remote biometric identification systems in publicly accessible places, which is a very controversially discussed topic, can equally be presumed to have a quality management system, since no customer would want to risk their reputation by using such a system that hasn't been properly quality controlled. After adapting to the AI requirements, the quality management system has to be audited by the notified body and be proven compliant with the standards and the regulation. The initial audit costs between €1 000 and €2 000 per day, and the amount of days will depend on the number of employees. The audits need to be re-audited yearly, which will take less time and a correspondingly smaller costs. These costs could be further reduced when companies make use of existing standards as described above.[269] Moreover, the Regulation foresees that Notified Bodies, when setting their fees, shall take into account the needs of SMEs.

In addition, for each individual product the notified body will have to review documentation meant to prove that the product complies with the AI regulation to ensure that it is indeed compliant with the requirements. Such a review is expected to take between one and two and a half days. This amounts to a range of €3,000-7,500 for the notified body to monitor compliance with the documentation requirements.

With an assumed average cost for an AI system of €170 000, this amounts to between 2% and 5%. Applied to a maximum investment volume of €65 billion, aggregate costs would be between €1 billion and €3 billion if all AI systems were thus tested. However, only 5% to 15% of all AI applications are estimated to constitute a high risk, and only a subset of these (AI systems that are safety components of products and remote biometric identifications systems in publicly accessible places) would be subject to third-party conformity assessment. Hence, taking 5% as a reasonable estimate, aggregate costs would amount to around €100 million.

Analogue to the discussion above for compliance costs, when AI is embedded the total development costs are much higher than the software alone and the share of AI verification costs is correspondingly smaller. The share of 2% to 5% of total development costs would thus only apply to non-embedded AI applications that nevertheless need to undergo new third-party conformity assessment, i.e. remote biometric identification in publicly accessible spaces.

Again, to put these figures into perspective, the costs of other recent initiatives help. For example, for the Cybersecurity Act**,** in France the Certification Sécuritaire de Premier Niveau (CSPN) costs about €25,000 – €35,000 while in the Netherlands the Baseline Security Product Assessment (BSPA) costs on average €40,000[270]. Similarly, the conformity assessment for a laptop is estimated at around € 25 000.[271] The average cost of a conformity assessment under the Machinery Directive is € 275 000.[272]

In **option 3+,** additional aggregate verification costs would consist in random checks of companies having introduced a code of conduct, financed by fees from participating companies, if the code of conduct foresees such random checks. This would amount to a fraction of the costs of verification of the high-risk application. Total aggregate costs would thus lie slightly above option 3 (around € 100 million).

---

[269] Such as ISO 9001/2015 'general), IEC13485 (medical devices), ISO/IEC 25010 (software).
[270] European Commission, Impact Assessment Accompanying the Cybersecurity Act, SWD(2017) 500 final
[271] Centre for Strategy & Evaluation Services (CSES): Evaluation of Internal Market Legislation for Industrial Products
[272] ResearchGate, *Calculating average cost per company of annual conformity assessment activities*.

Under **option 4**, the assessment costs for each application would be identical, but all AI applications would be covered, resulting in aggregate costs between €1 billion and €3 billion.

Table 9: Overview: Estimated maximum aggregate compliance costs and administrative burden by 2025

	COMPLIANCE + ADMIN COSTS	VERIFICATION COSTS
Option 1	Between €0 and €3 billion (all voluntary)	€0
Option 2	n/a	n/a
Option 3	Between €100 million and €500 million	Around €100 million
Option 3+	Between €100 / €500 million and €3 billion (voluntary above €100 / €500 million)	Slightly above €100 million
Option 4	Around €3 billion in 2025	Between €1 billion and €3 billion

Notabene: does not include the compliance costs for human oversight, which accrue to the user, not the AI supplier

6.1.4. SME test

Under **option 1**, SMEs would only sign up to the voluntary labelling scheme if the benefits in terms of credibility outweigh the costs. Thus, proportionality is ensured by design. **Option 2** limits the requirements to specific well-defined cases if and when problems arise or can be anticipated. Each ad-hoc regulation will thus only concern a small share of SMEs. SMEs working on several classes of applications subject to ad-hoc regulation, would, however, have to comply with multiple specific sets of requirement, increasing administrative burden.

Regarding SMEs, the approach proposed in **options 3 and 3+**, precisely targeting only a narrow set of well-defined high-risk AI applications and imposing only algorithmic transparency and accountability requirements, keeps costs to a minimum and ensures that the burden is no more than proportionate to the risk. For example, users' record-keeping will be done automatically through system logs, which providers will be required to make available. By establishing clear requirements and procedures to follow at horizontal level, it also keeps administrative overhead as low as possible.

The vast majority of SMEs would not be affected at all, since obligations would be introduced only for high-risk applications. These non-affected SMEs would benefit from additional legal certainty, since they could be sure that their applications are not considered high-risk and will therefore not be subject to additional compliance costs or administrative burdens. The AI supplying SMEs concerned would however have to bear the limited costs just as large companies. Indeed, due to the high scalability of digital technologies, small and medium enterprises can have an enormous reach, potentially impacting millions of citizens despite their small size. Thus, when it comes to high risk applications, excluding SMEs from the application of the regulatory framework could seriously undermine the objective of increasing trust. However, they would benefit from a single set of requirements, streamlining compliance across applications. Under option 3+ SMEs that are not covered by the scope could invest in additional trust by adopting a code of conduct, if they see an economic advantage in doing so.

As with all regulations, the AI supplying SMEs concerned would in principle be more affected than large companies for several reasons. Firstly, in so far as large companies produce more AI applications, they can distribute the one-off costs of familiarising themselves (including legal advice if necessary) over more applications and would also experience a faster learning curve. Nevertheless, most of the additional fixed compliance costs generated by the legislation occur for every new application and thus do not provide economies of scope to the larger companies. Secondly, in so far as their applications find more customers they can distribute the fixed costs of

regulation (such as the testing for non-discrimination effects) over more customers (economies of scale). However, many AI applications are bespoke developments for specific customers where this will not be possible, since the fixed costs may have to be incurred again (e.g. training data is likely to be different for customised applications). Thirdly, SMEs financial capacity to absorb additional burdens is much more limited. SMEs produce an average annual value added of €174 000, going as low as €69 000 for micro-enterprises (less than ten employees), compared to €71.6 million for large enterprises.[273] This compares with estimated compliance costs of €6 000 - €7 000 for those SMEs that would develop or deploy high-risk AI applications.

SMEs are also expected to benefit significantly from a regulatory framework. Firstly, small and therefore generally not well-known companies will benefit more from a higher overall level of trust in AI applications than large established companies, who already have large bases of established and trusting customers (e.g. an increase in trust in AI is less likely to benefit large platform operators or well-known e-Commerce companies, whose reputation depends on many other factors). This applies especially to the companies using AI in the business-to-consumer market, but also to the AI suppliers in the business-to-business market, where customers will value the reassurance that they are not exposed to legal risks from the application they are purchasing or licensing. Secondly, legal uncertainty is a bigger problem for SMEs than for large companies with their own legal department. Thirdly, for small enterprises seamless market access to all Member States is more important than for large companies, which are better able to handle different regulatory requirements. SMEs also lack the scale to recoup the costs of adapting to another set of regulatory requirements by sufficiently large sales in other Member States. As a result, SMEs profit more from the avoidance of market fragmentation. Thus, the legislation would reduce the existing disadvantages of SMEs on the markets for high-risk AI applications.

Options 3 and 3+ also foresee to implement regulatory sandboxes allowing for the testing of innovative solutions under the oversight of the public authorities. These sandboxes would allow proportionate application of the rules to the SMEs as permitted in the existing legislation and thus allow a space for experimentation under the new rules and the existing legal framework. This will support the SMEs in reaching compliance in the pre-market phase that will ultimately facilitate their entry into the market. The regulatory oversight shall give guidance to providers how to minimize the associated risks and allow competent authorities to exercise their margin of discretion and flexibility as permitted by the applicable rules. Before an AI system can be placed on the market or put into service into a 'live' environment, the provider should ensure compliance with the applicable standards and rules for safety and fundamental rights and complete the applicable conformity assessment procedure. Direct guidance from the competent authorities will minimise the legal risk of non-compliance and thus reduce the compliance costs for SMEs participating in the sandboxing scheme, for example by reducing the need for legal or technical advice from third parties. Moreover, it will allow SMEs to bring their products and services to market faster.

SMEs, like any other AI provider, will also be able to rely on harmonized standards that will guide them in the implementation of the new requirements based on standardized good practices and procedures. This would alleviate the SMEs from the burden of developing these standards and good practices on their own and help them to build trust in their products and services, which is key not only for consumers, but also businesses customers across the value chain.

As a result, the foreseen regulatory requirements would not create a barrier to market entry for SMEs. One should also recall that notified bodies are bound to take the size of the company into account when setting their fees, so that SMEs will have lower costs than large companies for conformity assessment.

[273] Estimates for 2018 produced by DIW Econ, based on 2008-2016 figures from the Structural Business Statistics Database (Eurostat, *Structural business statistics overview*, 2020).

Option 4 would lead to SMEs being exposed to the regulatory costs when developing or using any AI application, no matter whether the application poses risks or not, or whether consumer trust is an important sales factor for this application. Despite the limited costs, it would thus expose SMEs as well as large companies to disproportionate expenditures. Regulatory sandboxes analogue to options 3 and 3+ could be foreseen, but in order to have a similar effect, there would have to be many more of them, since many more applications would be in the scope of the regulatory framework.

In addition, **all options** would envisage measures to support SMEs, including through the AI resources and services made available by the AI-on-demand platform[274] and through the provision of model compliance programmes and guidance and support through the Digital Innovation Hubs and the Testing and Experimentation Facilities.[275]

The combined effect of the regulation on those SMEs providing AI will depend on how effective the support measures are in offsetting cost increases generated by the new legal requirements. The additional fixed compliance costs per AI system have been estimated at €6 000 - €7 000 for an average high-risk AI system of € 170 000, with another €3 000 to € 7 500 for conformity assessment (see section 6.1.3), while the monetary value of the support measures can not be determined with accuracy.

Access to free advice in the framework of regulatory sandboxes will be especially valuable to SMEs, at the beginning in particular, since they not only save on legal fees, but also receive guidance, reducing legal uncertainty to a minimum. Nevertheless, familiarisation with the requirements only accounts for a small part of the compliance costs. Access to the experimentation facilities of the Digital Innovation Hubs (DIH) and Testing and Experimentation Facilities (TEF) can be very valuable for SMEs thanks to their free services, although this may vary across sectors. For some sectors with little hardware requirements, cost savings will be smaller. For others, testing will require considerable physical infrastructure and free access to testing facilities is thus more beneficial. Contrary to large companies, SMEs cannot amortise costs for their own facilities over a large number of products. Note that cost savings provided by access to DIHs and TEFs may reduce both costs that are due to the regulation and costs which are not linked to the regulatory requirements. Finally, reduced costs for conformity assessment can partially compensate disadvantages SMEs may face due to the smaller scale of their operations. Moreover, by providing a focal point, regulatory sandboxes, DIHs and TEFs also facilitate partnering with complementary enterprises, knowledge acquisition and links to investors. For example, in a financial sector sandbox, a recent study found that entry into the sandbox was followed by an average increase in capital raised of 15% over the following two years.[276]

As a result, with the support measures the cost of regulatory requirements to SMEs are smaller than without such measures, but the costs are not completely offset. Whether the additional costs can at the margin discourage some SMEs from entering into certain markets for high-risk AI applications will depend on the competitive environment for the specific application and its technical specificities.

6.1.5. Competitiveness and innovation

To date the European AI market currently accounts for roughly a fifth of the total world market and is growing fast. It is thus highly attractive not just for European firms but also for competitors from third countries.

[274] https://www.ai4eu.eu
[275] Established under the Digital Europe Programme (currently under negotiation).
[276] Inside the Regulatory Sandbox: Effects on Fintech Funding

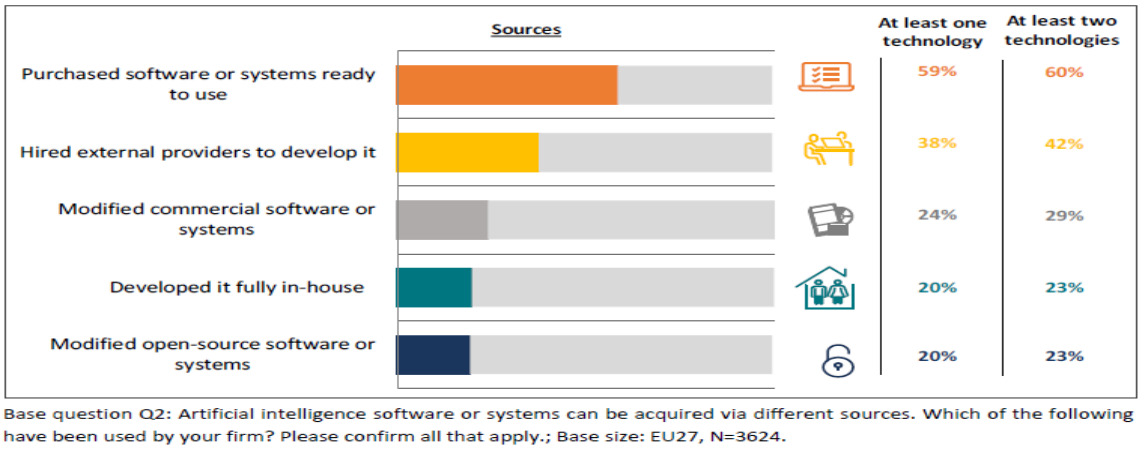
Table 10: AI investment estimates (€ million) from 2020 to 2025 [277]

AI INVESTMENTS	2020	2021	2022	2023	2024	2025
Global (Grand View)	48 804	70 231	101 064	145 433	209 283	301 163
EU (Grand View)	10 737	15 451	22 234	31 995	46 042	66 256
Global (Allied Market)	15 788	24 566	38 224	59 476	92 545	144 000
EU (Allied Market)	3 473	5 404	8 409	13 085	20 360	31 680

Source: Contractor' interpolation based on Allied Market Research, Grand View Research, and Tractica

The international competition is particularly tough because companies develop or adapt AI application in-house only to a smaller extent, and to a larger extent purchase them from external providers, either as a ready-to-use system or via hired external contractors. Thus, AI providers companies from outside the EU find it relatively easy to win market share, and as a result supply chains are often international.

Figure 11: Most common AI sourcing strategies for enterprises



Notabene: Applies to enterprises using at least one or two technologies.

Source: European enterprise survey on the use of technologies based on artificial intelligence, European Commission 2020 (Company survey across 30 European countries, N= 9640)

Against this background, the impact of the options on competitiveness and innovation is crucial. In principle, the impact of a regulatory framework on innovation, competitiveness and investment depends on two contradicting factors. On the one hand, the additional compliance costs and administrative burdens (see section 6.1.3.) make AI projects more expensive and hence less attractive for companies and investors. From an economic point of view, whether the obligations are imposed on the user or on the developer is irrelevant, since any costs the developer has to bear will eventually be passed on to the user. On the other hand, the positive impact on uptake (see section 6.1.2.) is likely to increase demand even faster, and hence make projects more attractive for companies and investors. The overall impact will depend on the balance of these two factors.

Under Option 1, companies will only undergo the additional costs if they consider that the increased uptake of their products and services will outweigh the additional costs. It will thus not negatively affect innovation and thus the competitiveness of European providers of AI applications.

Under Option 2, only a small number of specific applications would have to undergo the additional costs. A positive effect on uptake is possible, but less likely for revisions of existing legislation than

[277] Assuming a constant European share of the global AI market at 22%, based on its share in the AI software market in 2019 (Statista, *Revenues from the artificial intelligence software market worldwide from 2018 to 2025, by region*, 2019).

for ad-hoc legislation addressing a specific issue, since there would be no publicity effect. Innovation would become more expensive only for the specific applications regulated. Where regulation already exists, e.g. for many products, the impact will be lower, since companies are already equipped to deal with requirements.

However, under Option 3 increased costs and increased uptake through higher trust would be limited to a small subset of particularly high-risk applications. It is possible that AI providers would therefore focus investment on applications that do not fall in the scope of the regulatory framework, since the additional costs of the requirements would make innovations in non-covered AI applications relatively more attractive. Option 3+ would have similar effects, insofar as applications outside the scope would not be obliged to undergo the additional costs. Option 4 would see no such shift of supply but would see a much larger overall increase in cost, thus dampening innovation across all AI applications.

For options 3, 3+ and 4, there is no reason why investment into the development of 'high risk' use cases of AI would move to third countries in order to serve the European market, because for AI suppliers the requirement are identical on all markets. On the EU market foreign competitors would have to fulfil the same requirements; on third country markets, EU companies would not be obliged to fulfil the criteria (if they sell only to these markets). However, there is a theoretical risk that certain high-risk applications could not be sold profitably on the EU market or that the additional costs for users would make them unprofitable in use. For example, at the margin a recruitment software could be profitable for the provider if sold without respecting the requirements, but not if the provider would have to prevent it from discriminating against women. In those cases the choice implicit in the regulation would be that the respect of the fundamental right in question (in this case: non-discrimination) prevails over the loss of economic activity. Nevertheless, given the size of the EU market, which in itself accounts for 20% of the world market, it is very unlikely that the limited additional costs of algorithmic transparency and accountability would really prevent the introduction of this technology to the European market.

In addition, for Options 2, 3, 3+ and 4 there will be in addition the positive effect of legal certainty. By fulfilling the requirements, both AI providers and users can be certain that their application is lawful and do not have to worry about possible negative consequences. Hence, they will be more likely to invest in AI and to innovate using AI, thus becoming more competitive.

6.2. Costs for public authorities

Under options 1, 3, 3+ and 4, in-house conformity assessment as well as third-party conformity assessment would be funded by the companies (through fees for the third party mechanism). However, Member States would have to designate a supervisory authority in charge of implementing the legislative requirements and/or the voluntary labelling scheme, including market monitoring. Their supervisory function could build on existing arrangements, for example regarding conformity assessment bodies or market monitoring, but would require sufficient technological expertise. Depending on the pre-existing structure in each Member States, this could amount to 1 to 25 Full Time Equivalent (FTE) per Member State.[278] The resource requirement would be fairly similar, whether ex-ante enforcement takes place or not. If it does, there is more work to supervise the notified bodies and/or the ex-ante conformity assessment through internal checks of the companies. If it doesn't, there will be more incidents to deal with.

Options 1, 3, 3+ and 4 would benefit from a European coordination body to exchange best practices and to pool resources. Such a coordination body would mainly work via regular meetings of national competent authorities, assisted by secretarial support at EU level. This could amount to 10

[278] As a comparison, Data Protection Authorities in small Member States usually have between 20 and 60 staff, in big Member States between 150 and 250 (Germany is the outlier with 700; Brave, *Europe's Governments are failing the GDPR*, 2020).

FTE at EU level. As an additional option, the board could be supported by external expertise, e.g. in the form of a group of experts; the expertise would have to paid when needed and the cost would depend on the extent to which such expertise would be required. In addition, the EU would have to fund the database of high-risk AI applications with impacts mainly for fundamental rights. The additional costs at EU level should be more than offset by the reduction in expertise needed at national level, especially regarding the selection of applications for regulation and the gathering of a solid evidence base to support such a selection, which would be carried out at European level. Indeed, one of the key reasons why a European coordination mechanism is needed is that the pooling of resources is more efficient than a purely national build-up of expertise. Which costs option 2 would cause to public authorities would depend on the specific legislations. It is thus impossible to estimate at this stage. For the ad-hoc modifications of existing legislation with existing enforcement and supervisory structures, the costs to public authorities would be incremental, and European coordination could rely on existing structures as well. For ad-hoc legislation on new issues, e.g. remote biometric identification in publicly accessible spaces, Member States would have to build up new enforcement and supervisory structures with a more significant cost, including the building up of European coordination structures where they do not exist yet.

6.3. Social impact

All options, by increasing trust and hence uptake of AI applications, will lead to additional labour market impacts of AI. Generally, increasing uptake of AI applications is considered to cause a loss of some jobs, to create some others, and to transform many more, with the net balance uncertain. The increase in the use of AI and its facilitation also has important implications for skills, both in terms of requiring high-level AI skills and expertise and in ensuring that people can effectively use and interact with AI systems across the breadth of applications.[279] The High-Level Group High-Level Expert Group on the Impact of the Digital Transformation on EU Labour Markets[280] and the Special Report requested by former Commission president Juncker[281] have recently analysed these effects in depth or the Commission. Increasing uptake of AI will therefore reinforce these effects, and the stronger one option increases uptake, the stronger this effect will be.

By setting requirements for training data in high-risk applications, options 3 and 3+ would contribute to reducing involuntary discrimination by AI systems, for example used in recruiting and career management, thus improving the situation of disadvantaged groups and leading to greater social cohesion. Option 4 would have the same impact on a larger set of applications; however, since the additional applications are not high risk, the marginal impact of reducing discrimination is less significant. Option 2 would only have this effect where the classes of applications that was subject to ad-hoc regulation was prone to unfair discrimination. Similarly, option 1 would only have this effect for the applications obtaining the label and only in so far as these applications were high risk and prone to unfair discrimination.

Given the effect of AI applications to enable efficiencies, expand and improve service delivery across sectors, advancing the uptake of AI will also accelerate the development of socially beneficial applications, such as in relation to education, culture or youth. For example, by enabling new forms of personalised education, AI could improve education overall, and in particular for individuals that do not learn well under a one-size-fits-all approach. Similarly, by enabling new forms of collaboration, new insights and new tools, it allows young people to engage in creative activities. It could also be used to improve accessibility and provide support to persons with disabilities for example through innovative assistive technologies.

---

[279] OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 2019.

[280] High-Level Expert Group on The Impact of the Digital Transformation on EU Labour Markets, *Final Report with Recommendation*, 2018.

[281] Servoz, M. *AI – the future of work? Work of the future!*, 2019.

The potential for health improvement by AI applications in terms of better prevention, better diagnosis and better treatment, is widely recognised. Here, option 3 would address the most pertinent applications. However, since trust is so important in this sector, it would be very beneficial to give other AI applications as well the opportunity to prove their trustworthiness, even if they are not strictly high-risk. Option 3+ would therefore be highly relevant. The benefits of option 1 would be limited in this field of applications, since voluntary commitments do not yield the same level of confidence. Option 2 would well address the issue of AI applications for health, since the health sector already has a well-developed regulatory system.

6.4. Impacts on safety

All options aim to fill gaps in relation to the specific safety and security risks posed by AI-embedded in products in order to minimize the risks of death, injury and material damages.

While option 2 primarily concerns amendments to existing legislation for AI embedded in products but no new regulations for AI in services or as a stand-alone application, options 3, 3+ and 4 extend the scope of the horizontal framework to AI used in services and decision-making processes (for example software used for automatically managing critical utilities with severe safety consequences).

Compared to option 2, benefits of options 3 and 4 are generated by several factors. First of all, the risks to safety from the introduction of AI applications would decrease since a larger scope of AI systems posing risks to safety would be subject to AI-specific requirements. In particular, these requirements would concern AI components that are integrated into both products and services. With regard to the AI safety components of products already covered by option 2, option 3, 3+ and 4 would have greater benefits in terms of legal certainty, consistency and harmonised implementation of requirements aimed at tackling risks which are inherent to AI systems. This is because options 3, 3+ and 4 will avoid sectoral approach to tackling AI risks and regulate them in a harmonised and consistent manner. A horizontal instrument under options 3, 3+ and 4 would also provide harmonized requirements for managing the evolving nature of risks which will help to ensure that products are continuously safe during their lifecycle. This would be of particular value to AI providers and users who often operate in several sectors.

Moreover, under option 3, 3+ and 4, the process of development and adoption of harmonised standards on AI systems would be significantly streamlined, with the production of a consistent and comprehensive set of horizontal and vertical standards in the field. This would very much support providers of AI and manufacturers of AI-driven products in demonstrating their compliance with relevant rules. In addition, the integration of the requirements for AI embedded in products into conformity assessment procedures foreseen under sectoral legislation minimises the burden on sector-specific providers and, more generally, sector-specific operators.

While option 3 will impose new safety requirements only for high-risk AI systems, the positive safety-related benefits for society under option 4 are expected to be higher since all AI systems will have to comply with the new requirements for safety, security, accuracy and robustness and be accordingly tested and validated before being placed on the market. Option 3+ is fundamentally the same as option 3 in terms of binding legal requirements, while it introduces a system of codes of conduct for companies supplying or using low-risk AI. This voluntary system could be however a tool to push market operators to engage in ensuring a higher safety baseline for their products even if they are low-risk.

6.5. Impacts on fundamental rights

Strengthening the respect of EU fundamental rights and effective enforcement of the existing legislation is one of the main objectives of the initiative.

All options will have some positive effects on the fundamental rights protection, although their extent will largely depend on the intensity of the regulatory intervention. While option 1 voluntary

labelling may marginally facilitate compliance with fundamental rights legislation by setting common requirements for trustworthy AI, these positive effects will be only for providers of AI systems who voluntarily decide to subscribe to the scheme. By contrast, binding requirements under options 2 to 4 will significantly strengthen the respect of fundamental rights for the AI systems covered under the different options.

A sectoral 'ad-hoc' approach under option 2 will provide legal certainty and fill certain gaps in or complement the existing non-discrimination, data protection and consumer protection legal frameworks, thus addressing risks to specific rights, covered by these frameworks. However, option 2 might lead to delays, inconsistencies and will be limited to the scope of application of each sectoral legislation.

A horizontal framework under options 3 to 4 will ensure consistency and address cross-cutting issues of key importance for the effective protection of the fundamental rights. Such a horizontal instrument will establish common requirements for trustworthy AI applicable across all sectors and will prohibit certain AI practices considered as contravening the EU values. Options 3 to 4 will also impose specific requirements relating to the quality of data, documentation and traceability, provision of information and transparency, human oversight, robustness and accuracy of the AI systems which are expected to mitigate the risks to fundamental rights and significantly improve the effective enforcement of all existing legislation. Users will also be better informed about the risks, capabilities and limitations of the AI systems, which will place them in a better position to take the necessary preventive and mitigating measures to reduce the residual risks.

An ex ante mechanism for compliance with these requirements and obligations will ensure that providers of AI systems take measures to minimize the risks to the fundamental rights by design since otherwise they will not be allowed to place their AI systems on the Union market. Conformity assessment through independent third party notified bodies would be more effective than ex ante conformity assessment through internal checks as an enforcement mechanism in this respect to ensure the effective protection of the fundamental rights. In particular, documentation and transparency requirements will be important to ensure that fundamental rights can be duly enforced before judicial or other administrative authorities. In addition, the ex post market surveillance and supervision by competent authorities should ensure that any violation of fundamental rights can be investigated and sanctioned in a proportionate, effective and dissuasive manner. Authorities will also have stronger powers for inspection and joint investigations. The obligations placed on providers to report to the competent national authorities serious breaches of obligation under Union and Member State law intended to protect fundamental rights will further improve the detection and sanctioning of these infringements.

The positive effect on the fundamental rights will be different depending on whether option 3, 3+ or 4 is chosen. While option 4 envisages horizontal regulatory intervention for all AI systems irrespective of the risk, option 3 targets only systems posing 'high risks' that require regulatory action because of their expected severity and high risks for the fundamental rights and safety. Given the larger scope of applications to all AI systems, option 4 might therefore lead to better protection of all fundamental rights examined in the problem definition section. However, the regulatory burden placed on so many economic operators and users and the impact on their freedom to conduct a business can actually prevent the development of many low-risk AI applications that can benefit fundamental rights (for instance AI used for bias detection, detection of security threats etc.).

Option 3+, which combines option 3 with codes of conduct for non-high risk, might be thus most suitable to achieve an optimal level of protection of all fundamental rights. This is expected to enhance the trust in the AI technology and stimulate its uptake, which can be very beneficial for the promotion of a whole range of political, social and economic rights, while minimizing the risks and addressing the problems identified in section 2.

In addition to these overall positive benefits expected for all fundamental rights, certain fundamental rights are likely to be specifically affected by the intervention. These are analysed in Annex 5.5.

6.6. Environmental impacts

The environmental impact depends on how effective the regulatory framework is in increasing trust and hence uptake, balanced against the resources needed for compliance and against the positive effects from increased uptake.

The environmental impact of option 1 would depend on how widespread the adoption of the label would be. If the adoption were to be sufficiently large to create a public perception that AI development has become more trustworthy than previously, it would increase uptake and hence energy and resource consumption, to be balanced by efficiency gains obtained through AI applications.

The environmental impact of option 2 would vary with the specific problem addressed. However, it would not create widespread trust in AI as such, but only in the class of applications regulated. Thus, it would reduce energy and resource consumption by limiting certain applications, but increase adoption of this particular class of applications. Given the horizontal usability of AI, the impact of regulating a single class of applications would be negligible from a society-wide point of view.

In options 3 and 3+, the direct environmental impacts that can be expected from the proposed measures are very limited. On the one hand, this options prevents the development of applications on the black list, and it limits the deployment of remote biometric identification systems in publicly accessible spaces. All of this will reduce energy and resource consumption and correspondingly CO_2 output.

On the other hand, the requirements do impose some additional activities with regard to testing and record-keeping. However, while machine learning is energy-intensive and rapidly becoming more so, the vast majority of the energy consumption occurs during the training phase. A significant increase in energy consumption would only take place if retraining were to be necessary on a large scale. However, whilst this may occur initially, developers will quickly learn how to make sure that their systems avoid retraining, given the enormous and rapidly increasing costs associated.

The indirect environmental impacts are more significant. On one hand, by increasing trust the measures will increase uptake and hence development and thus use of resources. It should be pointed out that this effect will not be limited to high-risk applications only – through cross-fertilization between different AI applications and re-use of building blocks, the increase in trust will also foster development in lower or no risk applications. On the other hand, many of the AI applications will be beneficial to the environment because of their superior efficiency compared to traditional (digital or analogue) technology. AI systems used in process optimisation by definition make processes more efficient and hence less wasteful, e.g. reducing the amounts of fertilizers and pesticides needed, decreasing the water consumption at equal output, etc.). AI systems supporting improved vehicle automation and traffic management contribute to the shift towards cooperative, connected and automated mobility, which in turn can support more efficient and multi-modal transport, lowering energy use and related emissions.

In addition, it is also possible to purposefully direct AI applications to improve the environment. For example, they can help pollution control and modelling the impact of climate change mitigation or adaptation measures. Finally, AI applications will minimise resource usage and energy consumption if policies encourage them to do so. Technical solutions include more efficient cooling systems, heat reuse, the use of renewable energy to supply data centres, and the construction of these data centres in regions with a cold climate. In the context of the Coordinated Plan on Artificial Intelligence with Member States, the Commission will consider options to encourage and promote

AI solutions that have a neutral/positive impact on climate change and environment. This will also reduce potential environmental impacts of the present initiative.

In option 4 the direct impacts would be very similar to those in option 3. The only difference is that more testing would take place, and hence consume more energy. The indirect impacts would be identical, except that the increase in uptake could be higher if some applications which require trust but are not considered 'high-risk' are more readily accepted by citizens.

7. HOW DO THE OPTIONS COMPARE?

7.1. Criteria for comparison

The following criteria are used in assessing how the options would potentially perform, compared to the baseline:

- Effectiveness in achieving the specific objectives:
  - ensure that AI systems placed on the market and used are safe and respect fundamental rights and Union values;
  - ensure legal certainty to facilitate investment and innovation;
  - enhance governance and effective enforcement of fundamental rights and safety requirements applicable to AI;
  - facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.
- Efficiency: cost-benefit ratio of each policy options in achieving the specific objectives;
- Coherence with other policy objectives and initiatives;
- Proportionality: whether the options go beyond what is a necessary intervention at EU level in achieving the objectives.

Table 11: Summary of the comparison of options against the four criteria

	EFFECTIVENESS				EFFICIENCY (cost-effectiveness)	COHERENCE	PROPOR-TIONALITY
	Objective 1	Objective 2	Objective 3	Objective 4			
Baseline scenario	0	0	0	0	0	0	0
Option 1: Voluntary labelling	0	0	0	+	++	+	+
Option 2: Ad-hoc legislation	+	+	+	+	++	+	+
Option 3: High risk only	++	++	++	++	++	+++	+
Option 3+: High risk + Codes of conduct	++	++	++	+++	++	+++	+
Option 4: All AI	+++	++	+++	++	0	+++	0

Notabene: table annotations should only be read in vertical; in the table, for options 3, 3+ and 4 it is assumed that ex-ante third party conformity assessments are mandatory for AI systems that are safety components of products and for remote biometric identification in publicly accessible spaces; "0" means same as baseline, "+" means partially better than baseline, "++" means better than baseline, "+++" means much better than baseline

7.2. Achievement of specific objectives

7.2.1. <u>First specific objective:</u> Ensure that AI systems placed on the market and used are safe and respect the existing law on fundamental rights and Union values

Option 1 would limit the risks for individuals regarding applications that have obtained the label, since companies would face sanctions if they claimed the label but did not actually respect the associated obligations. There would be a shift of demand to applications with the label, depending on how much attention consumers paid to this label. There is a chance that the label would become so widespread that it could set a standard that all market participants are forced to meet, but this is by no means certain. As a result, there is no guarantee that all or even most of high-risk applications would apply for the label and individuals would remain exposed to the risks identified earlier. Hence, option 1 would not be more effective than the baseline in achieving this objective.

Option 2 would effectively limit the risks for individuals, but only for those cases where action has been taken, assuming that the ad-hoc legislations will appropriately define the obligations for AI applications. Indeed, since the obligations can be precisely tailored to each use case, it will probably limit risks for the cases that are covered better than a horizontal framework. However, this effectiveness will only apply to the issues addressed in the separate legislations, leaving individuals unprotected against potential risks by other AI applications. Such an ad-hoc approach will also not be able to distribute obligations across the full AI value chain and will be limited to the material and personal scope of application of each sectorial legislation, which is likely to be more often for safety reasons than for fundamental rights. Option 2 is hence very effective for a number of cases but not comprehensive, and overall thus only be partially more effective than the baseline in achieving this objective.

Option 3 would effectively limit the risks to individuals for all applications that have been selected because the combination of the likelihood of violations and impact of such violations means that they constitute a high risk. By setting a comprehensive set of requirements and effective ex ante conformity assessment procedures, it makes violations for these applications much less likely before they are placed on the market. In addition, all providers of high-risk AI systems will have to establish and implement robust quality and risk management systems as well as post-market monitoring strategy that will provide efficient post-market supervision by providers and quick

remedial action for any emerging risks. An effective ex-post market surveillance control will be also carried out by national competent authorities having adequate financial and technical resources. Moreover, additional AI applications could be added as the need arises. Hence, option 3 is more effective than the baseline in achieving this objective.

Option 3+ would have the same legal effectiveness as option 3, but in addition allow companies that produce or use applications that have not been selected as high risk to nevertheless fulfil the obligations. Since risks to individuals in reality are not binary – either low or high – but follow a continuous graduation from zero to extremely high, providing such an incentive especially to applications which are at the edge of high risk but are not covered by the legal requirements could significantly further reduce the overall risk of violation. Thus, option 3+ would be more effective than the baseline in achieving this objective.

Option 4 would very effectively limit the risks by setting the same requirements as option 3, but for all AI applications. It would thus cover the high-risk applications of option 3, the applications at the edge of high risk that make the codes of conduct of option 3+ worthwhile, and all other applications as well, including many applications where there are no or only very low risks. Individuals would be comprehensively protected, and as a result, option 4 would be much more effective than the baseline effective in achieving this objective.

7.2.2. Second specific objective: Ensure legal certainty to facilitate investment and innovation in AI

Option 1 could not foster investment and innovation by providing legal certainty to AI developers and users. While the existence of the voluntary label and the compliance with the associated requirements could function as an indication that the company intends to follow recommended practices, from a legal point of view there would be only a small change compared to the baseline. Uncertainty regarding the application of EU fundamental rights and safety rules specifically to AI would remain, and the ensuing risk would continue to discourage investment. Thus, option 1 would not be more effective than the baseline in achieving objective 2.

Option 2 would improve investment and innovation conditions by providing legal certainty only for applications that have been regulated. Thus, option 2 would only be partially more effective than the baseline in achieving objective 2.

Option 3 would improve conditions for investment and innovation by providing legal certainty to AI developers and users. They would know exactly which AI applications across all Member States are considered to constitute a high risk, which requirements would apply to these applications and which procedures they have to undertake in order to prove their compliance with the legislation, in particular where ex-ante conformity assessments (third-party or through internal checks) are part of the enforcement system. Option 3 would thus be more effective than the baseline in achieving objective 2.

Given the rapid technological evolution, legal certainty would nevertheless not be absolute, since regulatory changes cannot be excluded over time, but only be minimised as far as possible. When proposing changes, European policy-makers would be supported in their analysis by a group of experts and by national administrations, which can draw on evidence from their respective monitoring systems.

Option 3+ would provide the same legal certainty as option 3. The additional code of conduct scheme would, as in option 1, function as an indication that the company is willing to take appropriate measures, but would not assure legal certainty to those participating. However, since unlike option 1 the applications covered by the codes of conduct would be medium to low risk applications, the need for legal certainty is arguably smaller than for those applications which are covered by the high-risk requirements. Option 3+ would thus be more effective than the baseline in achieving objective 2.

Option 4 would provide the same legal certainty as option 3, but for all AI applications. However, this increased legal certainty would come at the price of increased legal complexity for applications where there is no reason for such complications, since they do not constitute a high risk. It would thus simply be more effective than the baseline in achieving objective 2.

7.2.3. Third specific objective: Enhance governance and effective enforcement of the existing law on fundamental rights and safety requirements applicable to AI systems

Option 1 would moderately improve enforcement for those applications that have obtained the label. There would be specific monitoring by the issuer of the label, which could include audits; it is even possible that the label would require ex-ante verification. However, the limited coverage would preclude these improvements from being an overall enhancement of enforcement. Since the label would coexist with a series of national legislative frameworks, governance would be more complicated than in the baseline scenario. Hence, option 1 would not be more effective than the baseline in achieving objective 3.

Option 2 would presumably improve effective enforcement and governance for regulated applications, according to the specifications laid down in the relevant sectorial legislation. However, since these may very well differ from one area to the next, overall enforcement and governance of requirements related to AI applications may become more complicated, especially for applications that could fall into several regulated categories simultaneously. As a result, option 2 would only partially be more effective than the baseline in achieving objective 3.

Options 3, 3+ and 4 would all improve enforcement. For all three options, there would be the requirements to carry out ex-ante verification, either in the form of third party ex-ante conformity assessment (the integration of AI concerns into existing third party conformity assessments, and remote biometric identification in publicly accessible spaces) or in form of ex ante assessment through internal checks (mainly services). Compared to the baseline, this is a clear improvement in enforcement. Ex-post enforcement would also be considerable strengthened because of the documentation and testing requirements that will allow assessing the legal compliance of the use of an AI system. Moreover, in all of these options competent national authorities from different sectors would benefit from enhanced competences, funding and expertise and would be able to cooperate in joint investigations at national and cross border level. In addition, a European coordination mechanism is foreseen to ensure a coherent and efficient implementation throughout the single market. Of course, option 3 and the mandatory part of option 3+ would cover only high-risk of AI applications and would thus be more effective than the baseline in achieving objective 3, while Option 4 would cover all AI applications and would thus be much more effective than the baseline in achieving objective 3.

7.2.4. Fourth specific objective: Facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation

Option 1 would provide a certain improvement to the baseline, by establishing a common set of requirements across the single market and allowing companies to signal their adherence, thus allowing users to choose these applications. Consumers and businesses could therefore reasonably be confident that they purchase a lawful, trustworthy and safe product if it has obtained the label, no matter from which member state it originates. The single market would be facilitated only for those applications that have obtained the label. For all other applications, the baseline would continue to apply. There is also the real possibility that individual Member States esteem that the voluntary label does not sufficiently achieve objective 1 and therefore take legislative action, leading to fragmentation of the single market. Consequently, option 1 would only be partially more effective than the baseline in achieving objective 4.

Option 2 would provide a clear improvement to the baseline, which would however be limited to those products and services for which ad-hoc legislation (including amendments to existing legislations) is introduced. For those products, consumers and businesses could be certain that the

products and services they use are lawful, safe and trustworthy, and no Member States would be likely to legislate with respect to those products. This effect would come into being for each class of applications only after the ad-hoc legislation has been adopted. However, for products not covered by ad-hoc legislation, there would be no positive effect on consumer trust, and there is a real possibility of fragmentation of the single market along national borders, even assuming that the highest risk applications are those for which ad-hoc legislation would be agreed. Therefore, option 2 would only be partially more effective than the baseline in achieving objective 4.

Option 3 would provide a clear improvement to the baseline. On the one hand, for those cases that are covered, consumers and businesses can rely on the European framework to guarantee that the AI applications are lawful, trustworthy and safe coming from any Member States. On the other side, they can consider that those applications not covered by the legislation do not, in principle, constitute a high risk. Moreover, Member States are likely to refrain from legislation that would fragment the single market for low risk products[282], since they have agreed on a list of high-risk applications and since there is a mechanism to amend this list. As a result, option 3 would be more effective than the baseline in achieving objective 4.

Option 3+ would also provide a clear improvement to the baseline. It would have all the same effects of option 3, and in addition afford businesses the opportunity to signal their adherence to lawful, trustworthy and safe AI for those applications not considered high risk. .While it is uncertain how many non-high-risk applications would be developed in accordance with codes of conduct, the total increase in trust by business and consumers is at the minimum equivalent to option 3 and can be legitimately expected to be significantly higher. Option 3+ would thus be much more effective than the baseline in achieving objective 4.

Option 4 would create a comprehensive increase in trust by businesses and consumers, since they will know for all applications that providers had to fulfil the legal obligations. Moreover, since all risks will be covered, there is no risk of additional national legislation that could lead to fragmentation. However, one must also concede that the increase in costs for all AI applications (see discussion on proportionality below), including when there is no countervailing benefit because they do not extensively rely on user trust (e.g. industrial applications) can have the effect of fewer AI applications being offered, thus leading to a smaller market than otherwise. Option 4 thus only effectively achieves objective 4.

7.3. Efficiency

The costs of option 1 for AI providers and users would be similar for each AI application to the costs of option 3, if the requirements are identical, and if the enforcement mechanism is similar. On an aggregate level, the costs could be higher or lower than option 3, depending on how many companies introduce a code of conduct. However, the costs will be targeted in a less precise way because some costs AI applications that do not really need additional trust will incur them, and some applications that should undergo the requirements according to the risk-based approach will not do so. On the other hand, participation is voluntary and therefore left to the business decisions of companies. Hence, one can argue that it has no net cost – if the benefits did not outweigh the costs, companies would not participate. In that sense, option 1 would be cost effective. However, public administrations would still have to bear the costs to supervise the system, which could in principle cover all AI applications. Nevertheless, there would be no costs to policy-makers to determine high-risk applications, since any application can apply for the voluntary label.

Option 2 has overall low aggregate costs for AI providers and users, since it will only address specific problems and may often be implemented during regular revisions of existing regulations. However, the costs for each application can be significant, and the multiplicity of specific

[282] For high risk, they have already agreed and cannot adopt national measures that are contrary to the uniform rules agreed within the European horizontal instrument.

regulations may make compliance with them unnecessarily complicated. Nevertheless, it can be assumed that significant costs would only be occurred if the benefits were worth the effort. Public administrations would only incur costs in specific areas, where – in case of amending existing regulations - competent authorities would already be established. The costs of determining high risk applications would correspond to the choice of applications to be regulated. Overall, it can be assumed that option 2 is cost effective.

The costs of option 3 mainly consist in the burden on AI providers and users, which is in turn composed of the compliance costs and verification costs. While the costs for covered systems are moderate, the overall aggregate cost remains low due to the precise targeting of a small number of high-risk applications only. A limitation of third-party conformity assessments to AI systems that are safety components of products and remote biometric identification in publicly accessible spaces further limits the expenditure to the most relevant cases only. Moreover, the requirements are unified across applications, allowing for inexpensive and reusable compliance procedures. These costs are compensated by a strong positive impact on those applications where it is most needed. There would also be costs for public administrations that have to ensure enforcement, but they too would be limited, since the monitoring would only cover the applications classified as "high risk". For policy-makers there would be the additional costs of determining, based on solid evidence, what applications should be classified as high risk, which would however be small compared to overall compliance costs. The existence of an evidence base from the monitoring systems established by national competent authorities would help in minimising the risk that AI producers exploit their information advantages to misrepresent risks without prohibitive costs. Option 3 would thus be cost effective.

Regarding option 3+, for the mandatory part, the precise targeting ensures cost effectiveness. For the codes of conduct, the voluntary character ensures cost effectiveness. Overall, Option 3+ can be considered cost effective.

Option 4 has by far the highest aggregate costs for AI providers and users, since the costs per applications are the same, but the number of applications is far greater. These vastly increased costs are compensated only to little extent by an increased trust, since most of the additionally covered application do not rely on trust. Moreover, public administrations would have to monitor and enforce the system for all AI application, which would be significantly more resource-intensive than option 3. Thus, despite the fact that there would be no costs to policy-makers to determine high-risk applications, since all applications are covered, option 4 would not be cost effective.

7.4. Coherence

All options are fully coherent with the existing legislation on safety and fundamental rights. Options 1, 3, 3+ and 4 would promote or impose obligations to facilitate the implementation of existing legislation, and to address issues that existing legislation does not cover. Options 3, 3+ and 4 would make use of existing conformity assessment procedures wherever available. Option 2 would specifically cover applications where problems have arisen or are likely to arise that are not addressed by existing legislation.

All options are consistent with the separate initiative on liability, which, among others, aims to address the problems outlined in the Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics[283]. All options are equally coherent with the digital single market policy, by attempting to prevent the rising of barriers to cross-border commerce through the emergence of national and incompatible regulatory frameworks attempting to address the challenges raised by AI.

[283] European Commission, Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics,* COM/2020/64 final, 2020.

Options 3, 3+ and 4 are equally fully coherent with the overall strategy set out in Shaping Europe's digital future, which especially articulates a vision of "a European society powered by digital solutions that are strongly rooted in our common values", and with the European data strategy, which argues that the "vision stems from European values and fundamental rights and the conviction that the human being is and should remain at the centre." Building on these visions, both strategies attempt to accelerate the digital transformation of the European economy. Promoting legal certainty for the use of AI and ensuring it is trustworthy clearly contributes to this endeavour.

Option 1 has the same objective as the other initiatives but falls short in implementing these visions, since its non-binding character cannot guarantee the widespread respect of European values when it comes to AI applications. It is thus only partially coherent with European policy. Option 2 can only implement the respect of European values with regard to a subset of AI applications. It is thus equally only partially coherent with European policy

7.5 Proportionality

Options 1, 2, 3 and 3+ impose procedures that are proportional to the objectives pursued. Option 1 creates burdens only for companies who have voluntarily decided to so. Option 2 would only impose burdens when a concrete problem has arisen or can be foreseen, and only for the purpose of addressing this problem.

Option 3 only imposes burdens on a small number of specifically selected high-risk applications and only sets requirements that are the minimum necessary to mitigate the risks, safeguard the single market, provide legal certainty and improve governance. Only very limited transparency obligations are imposed where needed to inform affected parties that an AI system is used and provide them with the necessary information to enable them to exercise their right to an effective remedy. For high-risk systems, the requirements relating to data, documentation and traceability, provision of information and transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI and uncovered by other frameworks. A limitation of third-party conformity assessments to AI systems that are safety components of products and remote biometric identification in publicly accessible spaces also contributes to this precise targeting. Harmonized standards and supporting guidance and compliance tools will aim to help providers and users to comply with the requirements and minimize their costs. The costs incurred by operators are proportionate to the objectives achieved and the economic benefits that operators can expect from this initiative.

Option 3+ would have the same precise targeting plus allowing companies to follow voluntarily certain requirements for non-high-risk applications. Option 4, on the other hand, imposes burdens across all AI applications, whether justified by the risks each application poses or not. The aggregate economic cost for AI providers and AI users is therefore much higher, with no or only small additional benefits. It is thus disproportionate.

8. PREFERRED OPTION

As a result from the comparison of the options, **the preferred option is option 3+, a regulatory framework for high-risk AI applications with the possibility for all non-high-risk AI applications to follow a code of conduct.** This option would: 1) provide a legal definition of AI, 2) establish a definition of a high-risk AI system, and 3) set up the system of minimum requirements that high-risk AI systems must meet in order to be placed on or used on the EU market. The requirements would concern data, documentation and traceability, provision of information and transparency, human oversight and robustness and accuracy and would be mandatory for high-risk AI applications. Companies who introduce codes of conduct for other non-high-risk AI systems would do so voluntarily and these systems would be in principle shielded from unilateral Member States regulations.

Compliance would be verified through ex-ante conformity assessments and ex-post supervision and market surveillance. Ex-ante conformity assessments would be applicable to providers of all high-

risk AI systems. Every high-risk AI system will be certified for a specific intended purpose(s) so that its performance can be verified *in concreto*. If the purpose or the system's functionality are substantially changed by the user or a third party, they will have the same obligations as the provider in case the changed system qualifies as high-risk.

Regarding **high-risk AI systems which are safety components of products,**[284] the regulatory framework will integrate the enforcement of the new requirements into the existing sectoral safety legislation so as to minimise additional burdens. This integration will take place following an appropriate transitional period before the new AI requirements become binding for operators under the sectoral legislation. The mechanism of integration and extent of legal applicability of the horizontal instrument will depend on the nature and structure of the sectoral instruments in question.[285] In particular:

- Regarding **high-risk AI systems covered by NLF legislation,**[286] existing NLF conformity assessment systems would be applicable for checking the compliance of the AI system with the new requirements. The application of the horizontal framework would not affect the logic, methodology or general structure of conformity assessment under the relevant NLF product safety legislation (see Annex 5.3. - e.g. under the new Medical Device Regulation, the requirements of the horizontal AI framework would be applicable within the frame of the overall risk-benefit consideration which is at the heart of the assessment under that legislation). Obligations of economic operators and ex-post enforcement provisions (as described later in this text) of the horizontal framework will also apply to the extent they are not already covered under the sectoral product safety law.
- **Regarding high-risk AI systems covered by relevant Old Approach legislation**[287] **(e.g. aviation, cars),** applicability of the horizontal framework will be limited to the ex-ante essential requirements (e.g. human oversight, transparency) for high-risk AI systems, which will have to be taken into account when amending those acts or when adopting relevant implementing or delegated legislation under those acts.[288]

For other high-risk AI systems,[289] the conformity assessment could be done by the provider of the system based on ex ante assessment through internal checks. However, biometric remote

---

[284] See footnotes 229 and 300 for additional details.

[285] An overview of the impact and applicability of the horizontal framework to high-risk AI systems is provided in Annex 5.3.

[286] Based on up-to-date analysis, the concerned NLF legislations would be: Directive 2006/42/EC on machinery (which is currently subject to review), Directive 2009/48/EU on toys, Directive 2013/53/EU on recreational craft, Directive 2014/33/EU on lifts and safety components for lifts, Directive 2014/34/EU on equipment and protective systems intended for use in potentially explosive atmospheres, Directive 2014/53/EU on radio-equipment, Directive 2014/68/EU on pressure equipment, Regulation (EU) 2016/424 on cableway installations, Regulation (EU) 2016/425 on personal protective equipment, Regulation (EU) 2016/426 on gas appliances, Regulations (EU) 745/2017 on medical devices and Regulation (EU) 746/2017 on in-vitro diagnostic medical devices.

[287] Based on up-to-date analysis, the concerned old-approach legislation would be Regulation (EU) 2018/1139 on Civil Aviation, Regulation 858/2018 on the approval and market surveillance of motor vehicles, Regulation (EU) 2019/2144 on type-approval requirements for motor vehicles, Regulation (EU) 167/2013 on the approval and market surveillance of agricultural and forestry vehicles, Regulation (EU) 168/2013 on the approval and market surveillance of two- or three-wheel vehicles and quadricycles, Directive (EU) 2016/797 on interoperability of railway systems. Given the mandatory character of international standardization, Directive 2014/90/EU on marine equipment (which is a peculiar NLF-type legislation) will be treated in the same way as old-approach legislation.

[288] The direct or indirect applicability of requirements will fundamentally depend on the legal structure of the relevant old-approach legislation, and notably on the mandatory application of international standardisation. Where application of international standardisation is mandatory, requirements for the high-risk AI systems in the horizontal framework on AI will not directly apply but will have to be taken into account in the context of future Commission's activities in the concerned sectors.

[289] See footnote 231 and Annex 5.4.

identification in publicly accessible spaces would have to undergo an ex-ante third party conformity assessment, because of the particularly high-risks to breaches of fundamental rights.

In addition to ex-ante conformity assessments, there would also be an ex-post system for market surveillance[290] and supervision by national competent authorities designated by the Member States. In order to facilitate cross-border cooperation, a European coordination mechanism would be established which would function primarily via regular meetings between competent national authorities with some secretarial support at EU level. The EU body would be supported by an expert group to monitor technological developments and risks and provide evidence-based advice on the need for revision and updating of the high-risk use cases in public consultation of relevant stakeholders and concerned parties. This "Board on AI" will work in close cooperation with the European Data Protection Board, the EU networks on market surveillance authorities and any other relevant structures at EU level.

This option would best meet the objectives of the intervention. By requiring a restricted yet effective set of actions from AI developers and users, it would limit the risks of violation of fundamental rights and safety of EU citizens, but would do so in targeting the requirements only to applications where there is a high risk that such violations would happen. As a result, it would keep compliance costs to a minimum, thus avoiding an unnecessary slowing of uptake due to higher prices. In order to address possible disadvantages for SMEs, it would among others provide for regulatory sandboxes and access to testing facilities. Due to the establishment of the requirements and the corresponding enforcement mechanisms, citizens could develop trust in AI, companies would gain in legal certainty, and Member States would see no reason to take unilateral action that could fragment the single market. As a result of higher demand due to better trust, higher offers due to legal certainty, and the absence of obstacles to cross-border movement of AI systems, the single market for AI would be likely to flourish. The European Union would continue to develop a fast-growing AI ecosystem of innovative services and products embedding AI technology or stand-along AI applications, resulting in increased digital autonomy. As indicated in the introduction, the AI horizontal framework outlined in this preferred option will be accompanied by review of certain sectoral product safety legislation and new rules on AI liability.

With regard to review of safety legislation, as indicated in Section 1.3.2, review of some NLF sector-specific-legislation is ongoing in order to address challenges linked to new technologies. While relevant NLF product legislation would not cover aspects that are under the scope of the horizontal legislative instrument on AI for high-risk applications, the manufacturer would still have to demonstrate that the incorporation of a high-risk AI system covered by those NLF legislations into the product ensures the safety of the product as a whole in accordance with that NLF product legislation. In this respect, for example, the reviewed Machinery Directive 2006/42/EC could contain some requirements with regard to the safe integration of AI systems into the product (which are not under the scope of the horizontal framework). In order to increase legal clarity, any relevant NLF product legislation which is reviewed (e.g. Machinery Directive 2006/42/EC) would cross reference the AI horizontal framework, as appropriate. On the other hand, the General Product Safety Directive (GPSD) is also being reviewed to tackle emerging risks arising from new technologies. In line with its nature (see Section 1.3.2), the reviewed GPSD will be applicable, insofar as there are not more specific provisions in harmonised sector-specific safety legislation (including the future AI horizontal framework). Therefore, we can conclude that all revisions of safety legislation will complement and not overlap the future AI horizontal framework.

[290] For consistency purposes and in order to leverage on existing EU legislation and tools in the market surveillance domain, the provisions of the Market Surveillance Regulation 2019/1020 would apply, meaning the RAPEX system established by the General Product Safety Directive 2001/95/EC would be used for the exchange of relevant information with regard to measures taken by Member States again non-compliant AI systems.

Concerning liability, a longstanding EU approach with regard to product legislation is based on adequate combination of both safety and liability rules. This includes EU harmonised safety rules ensuring a high level of protection and the removal of barriers within the EU single market, and effective liability rules to provide for compensation where accidents nonetheless happen. For this reason, the Commission considers that only a combination of the AI horizontal framework with future liability rules can fully address the problems listed in this impact assessment specifically in terms of specific objectives 2 and 4 (legal certainty and single market for trustworthy AI). In fact, while the AI initiative shaped in this preferred option is an ex ante risk minimisation instrument to avoid and minimise the risk of harm caused by AI, the new rules on liability would be an ex post compensation instrument when such harm has occurred. Effective liability rules will also provide an additional incentive to comply with the due diligence obligations laid down in the AI horizontal initiative, thus reinforcing the effectiveness and intended benefits of the proposed initiative.

In terms of timing for the adoption,[291] the Commission has decided at political level that in order to provide clarity, consistency and certainty for businesses and citizens the forthcoming initiatives related to AI, as proposed in the White Paper on AI, will be adopted in stages. First, the Commission will propose the **AI horizontal legal framework (Q2 2021)** which will **set the definition for artificial intelligence**, a solid risk methodology to **define high-risk AI**, certain **requirements** for AI systems and certain obligations for the key operators across the value chain (providers and users). Second, the **liability framework** (expected Q4 2021/Q1 2022) will be proposed, possibly comprising a review of the Product Liability Directive and harmonising targeted elements of civil liability currently under national law. The future changes to the liability rules will take into account the elements of the horizontal framework with a view to designing the most effective and proportionate solutions with regard to liability for **damages/harm caused by AI systems** as well as ensuring effective **compensation of victims**. The AI horizontal framework and the liability framework will complement one another: while the requirements of the horizontal framework mainly aim to protect against risks to fundamental rights and safety from an ex-ante perspective, effective liability rules primarily take care of damage caused by AI from an ex-post angle, ensuring compensation should the risks materialise.[292] Moreover, compliance with the requirements of the AI horizontal framework will be taken into account for assessing liability of actors under future liability rules.[293]

Table 12: Forthcoming EU AI initiatives

AI INITIATIVE	MAIN ELEMENTS (SCOPE) WITH REGARD TO AI SYSTEMS
Horizontal legislation on AI (current proposal)	- Sets a **definition for "artificial intelligence"** - Sets risk assessment methodology and defines **high-risk AI systems** - Sets certain **minimum requirements for high risk AI systems** (e.g. minimum transparency of algorithm, documentation, data quality) - Sets **legal obligations with regard to the conduct of key economic operators** (providers and users) - Sets a **governance system** at national and EU level for the effective enforcement of these rules

[291] The new liability rules on AI are currently under reflection (see section 1.3. for more details on the issues at stake).

[292] The discussion on the requirements in the horizontal framework that relate to safety of a system and protection of fundamental rights 'ex ante' and 'ex post' placement on the market (that are both covered in the proposed horizontal framework) is discussed in another part of the text. For example, to ensure ex-post enforcement of requirements provided in the horizontal regulation, as discussed in the sections on enforcement, the proposal includes appropriate investigations by competent authorities with powers to request remedial action and impose sanctions.

[293] The relevant recital provision to this extent would be included in the proposed horizontal framework initiative.

New and adapted liability rules (under reflection - expected Q4 2021-Q1 2022) [294]	- Makes **targeted adaptations to liability rules**, to ensure that victims can claim compensation for damage caused by AI systems - May introduce **possible adaptations to the existing EU product liability rules** (based on strict liability), including notions of product, producer, defect as well as the defences and claim thresholds. - May propose **possible harmonisation of certain elements of national liability systems** (strict and fault-based) - May provide **possible specific considerations for certain sectors (e.g. healthcare)** - All possible **changes will take into account foundational concepts (e.g. the definition of AI) and legal obligations with regard to the conduct of key economic operators set by the AI horizontal framework.**
Sectoral safety legislation revisions	- The revisions **will complement, but not overlap with the horizontal AI framework** - May **set certain requirements to ensure that integration of the AI systems into the product is safe and the overall product performance is not compromised**

9. HOW WILL ACTUAL IMPACTS BE MONITORED AND EVALUATED?

Providing for a robust monitoring and evaluation mechanism is crucial to evaluate how far the regulatory framework succeeds in achieving its objectives. One could consider it a success if AI systems based on the proposed regulatory approach would be appreciated by consumers and businesses, with the European Union and Member States developing together a new culture of algorithmic transparency and accountability without stifling innovation. As a result, AI made in the EU incorporating a trust-based approach would become the world reference standard. AI made in Europe would be characterised by the absence of violations of fundamental rights and incidents physically harming humans due to AI.

The Commission will be in charge of monitoring the effects of the preferred policy option. For the purpose of monitoring, it will establish a system for registering stand-alone AI applications with implications mainly for fundamental rights in a public EU-wide database. This would also enable competent authorities, users and other interested people to verify if the high-risk AI system complies with the new requirements and exercise enhanced oversight over these AI applications posing increased rights to fundamental rights (Annex 5.4.). To feed this database, AI suppliers will be obliged to provide meaningful information about the system and the conformity assessment carried out.

Moreover, AI providers will be obliged to inform national competent authorities about serious incidents or AI performances which constitute a breach of fundamental rights obligations as soon as they become aware of them, as well as any recalls or withdrawals of AI systems from the market. National competent authorities will then investigate the incidents/breaches, collect all the necessary information and regularly transmit it with adequate metadata to the EU board on AI, broken down by fields of applications (e.g. recruitment, biometric recognition etc.) and calculated a) in absolute terms, b) as share of applications deployed and c) as share of citizens concerned.

. The Commission will complement this information on applied high-risk AI use cases by a comprehensive analysis of the overall market for artificial intelligence. To do so, it will measure AI uptake in regular surveys (a baseline survey has been carried out by the Commission in Spring 2020), and use data from national competent authorities, Eurostat, the Joint Research Center

[294] As indicated in Section 1.3.3., one of the elements under reflection is the possible Revision of the Product Liability Directive. The Product Liability Directive is a technology-neutral directive applicable to all products. If and when reviewed, it would also apply to high-risk AI systems covered under the AI horizontal framework.

(through AI Watch) and the OECD. It will pay particular attention to the international compatibility of the data collections, so that data becomes comparable between Member States and other advanced economies. The joint OECD/EU AI observatory is a first step on the way to achieve this.

The following list of indicators is provisional and non-exhaustive.

Table 13: Indicators for monitoring and evaluation

OBJECTIVE	INDICATOR	SOURCE
AI systems are safe and respect EU fundamental rights and values (negative indicators)	Number of serious incidents or AI performances which constitute a serious incident or a breach of fundamental rights obligations (semi-annual) by fields of applications and calculated a) in absolute terms, b) as share of applications deployed and c) as share of citizens concerned	National competent authorities; European Data Protection Board
Facilitate investment and innovation (positive indicators)	Total AI investment in the EU (annual) Total AI investment by Member State (annual) Share of companies using AI (annual) Share of SMEs using AI (annual) Projects approved through regulatory sandboxes and placed on the market (annual) Number of SMEs consulting on AI in Digital Innovations Hubs and Testing and Experimentation Facilities	Commission services and AI Watch; National competent authorities
Improve governance and enforcement mechanisms (negative indicators)	Number of recalls or withdrawals of AI systems from the market (semi-annual); by fields of applications and calculated a) in absolute terms, b) as share of applications deployed and c) as share of citizens concerned	National competent authorities
Facilitate single market	Level of trust in artificial intelligence (annual) (positive indicator) Number of national legislations that would fragment the single market (biannual) (negative indicator)	Commission services and AI Watch

Note: For a positive indicators, a higher value represents a better outcome. For a negative indicator a lower value represents a better outcome.

Taking into account these indicators and complementing with additional ad-hoc sources as well as qualitative evidence, the Commission will publish a report evaluating and reviewing the framework five years following the date on which it becomes applicable.

Glossary[295]

Acquis	The EU's 'acquis' is the body of common rights and obligations that are binding on all EU countries, as EU Members. Source: EUR-Lex glossary
AI	Artificial Intelligence
Artificial intelligence (AI) system	An AI system is a machine-based system that can, for a given set of human-defined objectives, generate output such as content, predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy. Source: based on OECD AI principles
Algorithm	Finite suite of formal rules (logical operations, instructions) allowing to obtain a result from input elements. This suite can be the object of an automated execution process and rely on models designed through machine learning. Source: Council of Europe AI Glossary
ALTAI	Assessment List for Trustworthy Artificial Intelligence, developed by the EU's High-Level Expert Group on Artificial Intelligence.
Autonomous systems	ICT-based systems which have a high degree of automation and can for instance perceive their environment, translate this perception into meaningful actions and then execute these actions without human supervision.
Algorithmic bias	AI bias or (or algorithmic) bias describes systematic and repeatable errors in a computer system that create unfair outcomes, such as favouring one arbitrary group of users over others. Source: ALTAI glossary
Black-box	In the context of AI and machine learning-based systems, the black box refers to cases where it is not possible to trace back the reason for certain decisions due to the complexity of machine learning techniques and their opacity in terms of unravelling the processes through which such decisions have been reached. Source: European Commission Expert group on Ethics of connected and automated vehicles study
Chatbot	Conversational agent that dialogues with its user (for example: empathic robots available to patients, or automated conversation services in customer relations). Source: Council of Europe Glossary
CJEU	Court of Justice of the European Union
CT	Computer Tomography
Data sovereignty	Concept that data is protected under law and the jurisdiction of the state of its origin, to guarantee data protection rights and obligations.
Data value chain	Underlying concept to describe the idea that data assets can be produced by private actors or by public authorities and exchanged on efficient markets like commodities and industrial parts (or made available for reuse as public goods) throughout the lifecycle of datasets (capture, curation, storage, search, sharing, transfer, analysis and visualization). These data are then aggregated as inputs for the production of value-added goods and services which may in turn be used as inputs in the production of other goods and services.
Deep Learning	A subset of machine learning that relies on neural networks with many layers of neurons. In so doing, deep learning employs statistics to spot underlying trends or data patterns and applies that knowledge to other layers of analysis. Source: The Brookings glossary of AI and emerging technologies

[295] If not indicated otherwise, Source of the definitions: DG CNECT Glossary.

Deepfakes:	Digital images and audio that are artificially altered or manipulated by AI and/or deep learning to make someone do or say something he or she did not actually do or say. Pictures or videos can be edited to put someone in a compromising position or to have someone make a controversial statement, even though the person did not actually do or say what is shown. Increasingly, it is becoming difficult to distinguish artificially manufactured material from actual videos and images. Source: The Brookings glossary of AI and emerging technologies
DIH	Digital Innovation Hub
Distributed computing	A model where hardware and software systems contain multiple processing and/or storage elements that are connected over a network and integrated in some fashion. The purpose is to connect users, applications and resources in a transparent, open and scalable way, and provide more computing and storage capacity to users. In general terms, distributed computing refers to computing systems to provide computational operations that contribute to solving an overall computational problem.
Embedded system	Computer system with a dedicated function within a larger system, often with real-time computing constraints comprising software and hardware. It is embedded as part of a complete device often including other physical parts (e.g. electrical, mechanical, optical). Embedded systems control many devices in common use today such as airplanes, cars, elevators, medical equipment and similar.
Facial Recognition	A technology for identifying specific people based on pictures or videos. It operates by analysing features such as the structure of the face, the distance between the eyes, and the angles between a person's eyes, nose, and mouth. It is controversial because of worries about privacy invasion, malicious applications, or abuse by government or corporate entities. In addition, there have been well-documented biases by race and gender with some facial recognition algorithms. Source: The Brookings glossary of AI and emerging technologies
GDPR	General Data Protection Regulation 2016/679
Harmonised Standard	A European standard elaborated on the basis of a request from the European Commission to a recognised European Standards Organisation to develop a standard that provides solutions for compliance with a legal provision. Compliance with harmonised standards provides a presumption of conformity with the corresponding requirements of harmonisation legislation. The use of standards remains voluntary. Within the context of some directives or regulations voluntary European standards supporting implementation of relevant legal requirements are not called 'harmonised standards'.
HLEG	High-Level Expert Group on Artificial Intelligence.
IoT **(Internet of Things)**	Dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual "things" have identities, physical attributes and virtual personalities and use intelligent interfaces and are seamlessly integrated into the information network.
ISO	International Organization for Standardization
Machine learning	Machine learning makes it possible to construct a mathematical model from data, including a large number of variables that are not known in advance. The parameters are configured as you go through a learning phase, which uses training data sets to find links and classifies them. The different machine learning methods are chosen by the designers according to the nature of the tasks to be performed (grouping, decision tree). These methods are usually classified into 3 categories: human-supervised learning, unsupervised learning, and unsupervised learning by reinforcement. These 3 categories group together different methods including neural networks, deep learning etc. Source: Council of Europe Glossary
Natural language processing	Information processing based upon natural-language understanding. Source: ISO

Neural Network	Algorithmic system, whose design was originally schematically inspired by the functioning of biological neurons and which, subsequently, came close to statistical methods. The so-called formal neuron is designed as an automaton with a transfer function that transforms its inputs into outputs according to precise logical, arithmetic and symbolic rules. Assembled in a network, these formal neurons are able to quickly operate classifications and gradually learn to improve them. This type of learning has been tested by tests on games (Go, video games). It is used for robotics, automated translation, etc. Source: Council of Europe Glossary
NLF **New legislative framework**	To improve the internal market for goods and strengthen the conditions for placing a wide range of products on the EU market, the new legislative framework was adopted in 2008. It is a package of measures that streamline the obligations of manufacturers, authorised representatives, importers and distributors, improve market surveillance and boost the quality of conformity assessments. It also regulates the use of CE marking and creates a toolbox of measures for use in product legislation. Source: European Commission, Internal Market, Industry, Entrepreneurship and SMEs
OECD	The Organisation for Economic Co-operation and Development.
PLD	Product Liability Directive, i.e. Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, OJ L 210, 7.8.1985, p. 29–33, ELI: http://data.europa.eu/eli/dir/1985/374/1999-06-04.
Self-learning AI system	Self-learning (or self-supervised learning) AI systems recognize patterns in the training data in an autonomous way, without the need for supervision. Source: ALTAI glossary
SME **Small- and Medium-sized Enterprise**	An enterprise that satisfies the criteria laid down in Commission Recommendation 2003/361/EC of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises (OJ L 124, 20.05.2003, p. 36): employs fewer than 250 persons, has an annual turnover not exceeding €50 million, and/or an annual balance sheet total not exceeding €43 million.
Supervised Learning	According to Science magazine, supervised learning is 'a type of machine learning in which the algorithm compares its outputs with the correct outputs during training. Supervised learning allows machine learning and AI to improve information processing and become more accurate'. Source: The Brookings glossary of AI and emerging technologies
Training data	Samples for training used to fit a machine learning model. Source: ISO
Trustworthy	Trustworthy AI has three components: 1) it should be lawful, ensuring compliance with all applicable laws and regulations; 2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values; and 3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the AI system's life cycle. Source: ALTAI glossary
Use case	Use case: A use case is a specific situation in which a product or service could potentially be used. For example, self-driving cars or care robots are use cases for AI. Source: ALTAI glossary