Council of the European Union

Brussels, 23 April 2021
(OR. en)

8115/21
ADD 3

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 2/2
Subject:	COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT ANNEXES 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 2/2.

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

EUROPEAN
COMMISSION

Brussels, 21.4.2021
SWD(2021) 84 final

PART 2/2

COMMISSION STAFF WORKING DOCUMENT

IMPACT ASSESSMENT

ANNEXES

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}

ANNEXES: TABLE OF CONTENTS

1. ANNEX 1: PROCEDURAL INFORMATION

1.1. Lead DG, Decide Planning/CWP references

Lead DG: Directorate-General for Communications Networks Content and Technology (CNECT).

Decide: PLAN/2020/7453.

CWP: Adjusted Commission Work Programme 2020 COM(2020) 440 final: Follow-up to the White Paper on Artificial Intelligence, including on safety, liability, fundamental rights and data (legislative, incl. impact assessment, Article 114 TFEU, Q1 2021).

1.2. Organisation and timing

The initiative constitutes a core part of the single market given that artificial intelligence (AI) has already found its way into a vast majority of services and products and will only continue to do so in the future. It is based on Article 114 TFEU since it 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 AI systems embedded in products and services or provided as stand-alone AI applications.

The impact assessment process started with opening of a public consultation on the AI White Paper[1] on 19 February 2020, open until 14 June 2020. The inception impact assessments was published for stakeholder comments on 23 July 2020, open for comments until 10 September 2020. For details on the consultation process, see Annex 2.

The inter-service group (ISG) met on 10 November 2020 before submission of the Staff Working Document to the Regulatory Scrutiny Board (18 November 2020). The ISG consists of representatives of the Secretariat-General, and the Directorates-General CNECT, JUST, GROW, LS, HOME, SANTE, FISMA, AGRI, JRC, DEFIS, TRADE, ENV, ENER, EMPL, EAC, MOVE, RTD, TAXUD, MARE, EEAS, ECFIN and CLIMA.

A meeting with the Regulatory Scrutiny Board was held on 16 December 2020. The Regulatory Scrutiny Board issued a negative opinion on 18 December 2020. The inter-service group met again on 18 January before re-submission of the Staff Working Document to the Regulatory Scrutiny Board (22 February 2021).

Based on the Board's recommendations of 18 December, the Impact Assessment has been revised in accordance with the following points.

1.3. Opinion of the RSB and responses

The Impact Assessment report was reviewed by the Regulatory Scrutiny Board. Based on the Board's recommendations, the Impact Assessment has been revised to take into account the following comments:

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

Comments of the RSB	*How and where comments have been addressed*
(B) Summary of findings	
(1) The report is not sufficiently clear on how this initiative will interact with other AI initiatives, in particular with the liability initiative.	The report has been substantially reworked, especially in the introduction, sections 1.3, 4.2 and 8, to better explain how this initiative interacts with other AI initiatives such as the safety revisions and the AI liability initiative, emphasizing the complementarity between the three initiatives and their different scopes**.** Regarding links with the liability initiative, the AI horizontal initiative is an ex ante risk minimisation instrument including a system of continuous oversight to avoid and minimise the risk of harm caused by AI, whilst the initiative on liability rules would be an ex post compensation instrument when such harm has occurred (Sections 1.3.3 and 8). Concerning the product safety revisions, these aim primarily at ensuring that the integration of AI systems into the overall product will not render a product unsafe and compliance with the sectoral rules is not affected. On the other hand, the AI legislation will set a single definition of AI, a risk assessment methodology and impose minimum requirements specific to the high-risk AI system to address both safety and fundamental rights risks (Section 1.3.2. and 8). Section 8 and Annex 5.3 explain in detail how the AI horizontal initiative will work in practice for sectoral safety legislation (old and new approach). Annex 5.3 also lists all pieces of sectoral product legislation that will be affected by the horizontal AI initiative.
(2) The report does not discuss the precise content of the options. The options are not sufficiently linked to the identified problems. The report does not present a complete set of options and does not explain why it discards some.	The report has been substantially re-worked to explain in detail the content of all policy options and their linkages to the problem identified in the impact assessment. The report now sets out in detail the five requirements for AI systems and how they are linked to the problems and the drivers (opacity, autonomy, data dependency etc.) (Policy Option 1). The prohibited practices are now clearly explained and justified with links to the problems and relevant justifications and recommendations for their prohibition (Policy Option 2). Option 2 also lists all the sectoral initiatives that would have to be undertaken and their content, including an ad hoc specific initiative that would further restrict the use of remote biometric identification systems at public spaces. The risk assessment methodology has been explained with the precise criteria defined (Policy Option 3). All high-risk AI use cases (not covered by product sectoral legislation) are listed and justified by applying the methodology in a new Annex 5.4 supported with evidence. Annex 5.3. explains, on the other hand, the methodology for high-risk AI covered by sectoral product safety legislation and lists the relevant acts that would be affected by the new horizontal initative. The compliance procedures and obligations for providers have been further explained for Policy Options 1, 2 and 3, linking them to the problems the AI regulatory initiative aims to solve. The same has been done for obligations of users for Options 2 and 3. Measures to support innovation are further explained in Option 3 (e.g. sandboxes, DIHs) and how they will operate and help to address the problems. Option 3+ has been reworked and now explains in detail the possibility for codes of conduct as a voluntary mechanism for non-high risk AI applications. Details on the enforcement and governance system at national and EU

	level have been given for all policy options. For each of these different issues, alternative policy choices/sub-otions have been considered and explanations given why they have been discarded. A new table 7 summarises the selected and discarded policy sub-options.
(3) The report does not show clearly how big the relative costs are for those AI categories that will be regulated by this initiative. Even with the foreseen mitigating measures, it is not sufficiently clear if these (fixed) costs could create prohibitive barriers for SMEs to be active in this market.	Section 6.1.3 has been reworked in order to put costs in relation to regulated AI applications. The report now provides a perspective on the level of costs by estimating costs of other regulatory requirements and explains why there is hardly any risk of depriving EU of certain technological innovations. It also distinguishes one-off and running costs and analyses which activities companies would have to undertake even without regulatory intervention. The role of regulatory sandboxes for SMEs has been clarified, with guidance from competent authorities to facilitate compliance and reduce costs (Section 5.4.).
(C) What to improve	
(1) The content of the report needs to be completed and reworked. The narrative should be improved and streamlined, by focusing on the most relevant key information and analysis.	The content of the report has been streamlined and focuses now more on the most relevant key information, such as how this initiative will interact with other AI initiatives, how the options are designed and what is their precise content, how the high-risk cases are selected. The context and the objectives of the proposal (especially Section 4.2.4.) have been detailed. The policy options have been further completed and better linked to the identified problems (Section 5). The report now presents a complete set of options and explains why it discards some.
(2) The report should clearly explain the interaction between this horizontal regulatory initiative, the liability initiative and the revision of sectoral legislation. It should present which part of the problems will be addressed by other initiatives, and why. In particular, it should clarify and justify the policy choices on the relative roles of the regulatory and liability initiatives.	The interaction between the AI horizontal initiative, the liability initiative and sectoral product safety revisions has been further explained and analysed (Introduction and Section 1.3.). Section 4.2. explains which parts of the problems will be addressed by the horizontal AI initiative and which parts by the liability and the sectoral product safety revisions. Policy choices on the relative roles of the regulatory and liability initiatives are clarified in Section 8. Annex 5.3. explains in detail how the AI horizontal initiative will work in practice for sectoral safety legislation (old and new approach) and lists all acts that will be affected by the horizontal AI initiative.
(3) In the presentation of the options, the report focusses mainly on the legal form, but it does not sufficiently elaborate on the content. The report should present a more complete set of options, including options that were considered but discarded. Regarding the preferred option, the report should give a firm justification on what basis it selects the four prohibited practices. There should be a clear definition and substantiation of the definition and list of high-risk systems. The same applies to the list of obligations. The report should indicate how high risks can be reliably identified, given the problem drivers of complexity, continuous adaptation and unpredictability. It should consider possible alternative options for the prohibited practices, high-risk systems, and obligations. These are choices that policy makers need to be informed about as a basis	The report now describes in detail the content of all policy options and clearly links them to the problem identified in the impact assessment. For each of the key dimensions linked to the content and the enforcement and governance system, it presents alternative policy choices and explains why it discards some. Two new tables are added: Table 6 Summary of the content of all Policy Options and Table 7 Summary of selected sub-option and discarded alternative sub-options. To improve readability, summary tables of the content of each policy option have also been added. Alternatives for the proposed mandatory AI requirements are discarded in Policy Option 1 (e.g. social and environmental well-being, accessibility, proposed by EP), but could be addressed via voluntary codes of conduct (Option 3+). Option 3 explains now in detail the risk assessment methodology for classification of a system as high-risk distinguishing between AI systems as safety components of products and other high-risk AI systems (stand-alone). For the second category, the methodology with the concrete criteria for assessment has been explained in detail and applied in Annex 5.4. More details are given how the high-risk cases are selected and on what evidence basis, starting from a larger pool of 132 ISO use cases and other possible applications (Annex 5.4.). In option 3, the report explains also

for their decisions.	that the methodology focusing on the severity and likelihood of harms that is appropriate to address the problem drivers of complexity, continuous adaptation and unpredictability. Alternative ways of how the risk assessment could be done are also discarded – e.g. burden placed on the provider for the risk assessment (Policy Option 3). Alternative prohibited practices are also considered, such as the complete prohibition of remote biometric identification systems and other cases requested by civil society organisation (Policy Option 2). Alternative ways of the proposed compliance procedure and obligations for providers and users are also analysed and discarded (Policy Option 3).
(4) The report should be clearer on the scale of the (fixed) costs for regulated applications. It should better analyse the effects of high costs on market development and composition. The report should expand on the costs for public authorities, tasked to establish evolving lists of risk rated AI products. It should explain how a changing list of high-risk products is compatible with the objective of legal certainty. The analysis should consider whether the level of costs affects the optimal balance with the liability framework. It should reflect on whether costs could be prohibitive for SMEs to enter certain markets. Regarding competiveness, the report should assess the risk that certain high-risk AI applications will be developed outside of Europe. The report should take into account experiences and lessons learnt from third countries (US, China, South Korea), for instance with regard to legal certainty, trust, higher uptake, data availability and liability aspects.	Section 6.1.3. has been reworked in order to put costs in relation to regulated AI applications. The report now also provides a perspective on the level of costs by estimating costs of other regulatory requirements. Section 6.1.4. has been strengthened to assess the impact on SMEs, and support measures for SMEs have been spelt out. Section 6.1.5. now discards specifically the possibility that certain high-risk AI applications will only be available outside of Europe as a result of the regulatory proposal. A new annex with an overview of development in third countries has been added (Annex 5.1.). Section 5.4.2.c) explains how a changing list of high-risk AI systems is compatible with the objective of legal certainty. The powers of the Commission would be preliminarily circumscribed by the legislator within certain limits. Any change to the list of high-risk AI use cases would also be based on the solid methodology defined in the legislation, supporting evidence and expert advice. 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 presenting the proposed content of the various policy options (Section 5.), the report also takes into account experiences and lessons learnt from third countries.
(5) The report should explain the concept of reliable testing of innovative solutions and outline the limits of experimenting in the case of AI. It should clarify how regulatory sandboxes can alleviate burden on SMEs, given the autonomous dynamics of AI.	Section 5.4. has been detailed and now outlines better the limits of experimenting with AI technologies (Policy Option 3). The role of regulatory sandboxes in the mitigation of burden on SMEs has been better clarified, since options 3 and 3+ foresee implementation of regulatory sandboxes allowing for the testing of innovative solutions under the oversight of the public authorities in order to alleviate the burden on SMEs (Section 6.1.4.). It has been clarified that no exemption will be granted, and that benefits to SMEs will come from lower costs for specialist legal and procedural advice and from faster market entry.
(6) The report should better use the results of the stakeholder consultation. It should better reflect the views of different stakeholder groups, including SMEs and relevant minority views, and discuss them in a more balanced way throughout the report.	The report has been reworked and completed with additional breakdowns of stakeholder views based on the public consultation on the White Paper on AI, for instance on the various problems identified in the impact assessment, on the need for regulation, on sandboxes, on costs and administrative burdens, on the limitation of requirements to high-risk applications, on the definition of AI, on the use of remote biometric identification systems in public spaces.
(7) The report should make clear what success would look like. The report should elaborate on monitoring arrangements and specify indicators for monitoring and evaluation.	The report has been further elaborated and detailed on monitoring and evaluation (Section 9). Success has been defined two-fold: 1) Absence of violation of safety and fundamental rights of individuals; 2) Rapid uptake of AI based on widespread trust. Thus, AI made in EU would become a world reference. Additional details on the reporting systems and the indicators have been provided: AI providers would be obliged to report safety incidents and

	breaches of fundamental rights obligations when brought to their attention; competent authorities would monitor and investigate incidents; the Commission would maintain a publicly accessible database of high-risk AI systems with mainly fundamental rights implications; and it will also monitor uptake of AI and market developments. Indicators for monitoring and evaluation are specified.

Second submission to the Regulatory Scrutiny Board

(1) The report should explain the methodology and sources for its cost calculations in the relevant annex. It should include a detailed discussion of where and why the presented costs deviate from the supporting study. The report should better discuss the combined effect of the foreseen support measures for SMEs (lower fees for conformity assessments, advice, priority access to regulatory sandboxes) and the (fixed) costs, including for new market entrants.	Annex 4 has been expanded to provide more details on the methodology extracted from the support study. An explanation on where and why assumptions and figures differ from the support study was provided. A new section has been added as the end of 6.1.4 setting out how the support measures provide benefits to SMEs and how far this counteracts the costs generated by the regulation.

1.4. Evidence, sources and quality

To ensure a high level of coherence and comparability of analysis for all potential policy approaches, an external study was procured to feed into the impact assessment. It reviewed available evidence of fundamental rights or safety-related risks created by AI applications, as well as assessed the costs of compliance with the potential requirements outlined in the AI White Paper. The study also reviewed evidence of potential compliance costs based on the review of literature or other countries and analysed results of the public consultation launched by the White Paper. The estimation of the costs of compliance can be found in Annex 4 of this impact assessment.

In order to gather more evidence following the consultation on the AI White Paper, the Commission organised in July, September and November 2020 five (closed) expert webinars on (1) Requirements for high-risk AI, (2) Standardisation, (3) Conformity assessment and (4) Biometric identification systems (5).

On 9 October 2020, the Commission organised the Second European AI Alliance Assembly, with the participation of over 1 900 viewers. Featuring Commissioner Thierry Breton, representatives of the German Presidency of the European Council, Members of the European Parliament as well as other high-level participants, the event focused on the European initiative to build an Ecosystem of Excellence and of Trust in Artificial Intelligence.[2] The sessions included plenaries as well as parallel workshops and breakout sessions. Viewers were able to interact and ask questions to the panellists.

Furthermore, the Commission held a broad stakeholder consultation on the White Paper. There were numerous meetings with companies, business associations, civil society, academia, Member States and third countries' representatives. In addition, Commission representatives participated in more than fifty (online) conferences and roundtables,

[2] European Commission, *Second European AI Alliance Assembly*, 2020.

organised by Member States, civil society, business associations, EU representations and delegations and others.

In addition, the IA takes into account the analysis and the work that contributed to the Ethical Guidelines adopted by the high-Level Expert Group on AI (HLEG AI) and the results of the testing of the Assessment List of the HLEG AI. The guidelines are based on the analysis of more than 500 submissions from stakeholders. The Assessment List of the HLEG AI, adopted in the second half of 2019, where more than 350 organisation participated.

Finally, to further support evidence based analysis, the Commission has conducted extensive literature review, covering academic books, journals and well as a wide spectrum of policy studies and reports, including by non-governmental organisations. They have been quoted in the main body of the Impact Assessment.

2. ANNEX 2: STAKEHOLDER CONSULTATION

In line with the Better Regulation Guidelines,[3] the stakeholders were widely consulted as part of the impact assessment process.

2.1. The public consultation on the White Paper on Artificial Intelligence

The main instrument was the public consultation on the White Paper on Artificial Intelligence that ran from 19 February to 14 June 2020. The questionnaire of the consultation was divided in three sections:

- **Section 1** referred to the specific actions, proposed in the White Paper's Chapter 4 for the building of an ecosystem of excellence that can support the development and uptake of AI across the EU economy and public administration;
- **Section 2** referred to a series of options for a regulatory framework for AI, set up in the White Paper's Chapter 5;
- **Section 3** referred to the Report on the safety and liability aspects of AI.[4]

The summary below only address the questions relating to **Sections 2** and **3** of the public consultation, where the regulatory framework is discussed.

The consultation targeted interested stakeholders from the public and private sectors, including governments, local authorities, commercial and non-commercial organisations, experts, academics and citizens. Contributions arrived from all over the world, including the EU's 27 Member States and countries such as India, China, Japan, Syria, Iraq, Brazil, Mexico, Canada, the US and the UK.

The public consultation included a set of **closed questions** that allowed respondents to select one or more options from a list of answers. In addition to the given options, respondents could provide **free text answers** to each question of the questionnaire or insert **position papers** with more detailed feedback.

In total, 1 215 contributions were received, of which 352 were from companies or business organisations/associations, 406 from citizens (92% EU citizens), 152 on behalf of academic/research institutions, and 73 from public authorities. Civil society's voices were represented by 160 respondents (among which 9 consumers' organisations, 129 non-governmental organisations and 22 trade unions), 72 respondents contributed as 'others'.

Out of 352 business and industry representatives 222 were individual companies/businesses, while 130 came from business associations, 41.5% of which were micro, small and medium-sized enterprises. The rest were business associations. Overall, 84% of business and industry replies came from the EU-27. Depending on the question, between 81 and 598 of the respondents used the free text option to insert comments.

Over 450 position papers were submitted through the EU Survey website, either in addition to questionnaire answers (over 400) or as stand-alone contributions (over 50). This brings the overall number of contributions to the consultation to over 1 250. Among the position papers, 72 came from non-governmental organisations (NGOs), 60 from business associations, 53 from large companies, 49 from academia, 24 from EU citizens,

---

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

[4] European Commission, , *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics* COM/2020/64 final, 2020.

21 from small and medium enterprises (SMEs), 19 from public authorities, 8 from trade unions, 6 from non-EU citizens, 2 from consumer organisations, with 94 not specified.

Main concerns

In the **online survey**, the overwhelming majority of participants (95%) responded to the section on the regulatory options for AI. Out of the concerns suggested in the White Paper, 90% and 87% of respondents found the possibility of AI breaching fundamental rights and the use of AI that may lead to discriminatory outcomes, respectively, as the most important ones. The possibility that AI endangers safety or takes actions that cannot be explained were also considered as (very) important by respectively 82% and 78% of respondents. Concerns over AI's possible lack of accuracy (70%) and lack of compensations following harm caused by AI (68%) follow.

The most reoccurring out of **390 free text answers** received for this question, highlighted the benefits of AI, to express the need of a balanced regulatory approach and the avoidance of 'overregulation' (48 comments). However, other comments add to the concerns related to AI. According to those, future regulation should pay attention to issues such as the transparency of decisions made by AI (32), the attribution of accountability for those decisions (13) as well as ensuring the capacity of human beings to making their own choices without being influenced by algorithms (human agency / human in the loop) (19). A number of non-governmental organisations underlined the need for a democratic oversight (11) while aspects such as equality (11), data quality (7), labour rights (5), safety (4) and others[5] were mentioned as well.

The importance of fundamental rights and other ethical issues was also underlined by many **position papers**. 42 position papers, 6 of which are arguing in favour of human rights impact assessments, mentioned the issue as one of their top three topics. Fundamental rights issues were mostly emphasized by NGOs (16), 5 of which were in-favour of introducing a human rights / fundamental rights impact assessment for AI. In addition, many respondents brought up other ethical issues such as discrimination and bias (21), the importance of societal impacts (18), data protection (15), civil society involvement (9) and human oversight (7).

What kind of legislation

In the relevant **online survey** question[6], 42% of respondents found the introduction of a new regulatory framework on AI as the best way to address the concerns listed in the previous paragraph. Among the main arguments used by participants in **226 free text answers** was that current legislation might have gaps when it comes to addressing issues related to AI and therefore a specific AI legislation is needed (47 comments). According to other comments such legislation should come along with appropriate research and gap analysis processes (39).

Other free text answers, however, highlighted that this process should take place with caution in order to avoid overregulation and the creation of regulatory burdens (24). 33% of participants to the online questionnaire thought that **the gaps identified, could be**

[5] Other arguments mentioned (minimum frequency): accuracy (10) , collective harms caused by AI (5), involvement of civil society (5), manipulation (5), power asymmetries (e.g. between governments and citizens; employers and employees; costumers and large companies) (4), safety (4), legal reediness & review, environmental impact of AI (4), unemployment/employment related discrimination (3), privacy & data protection (3), compensation (2), cybersecurity (2), intentional harmful abuse from AI (2), More R&D in AI can help address concerns (2), external threats to humanity (2), intellectual property rights (2) and media pluralism (2).

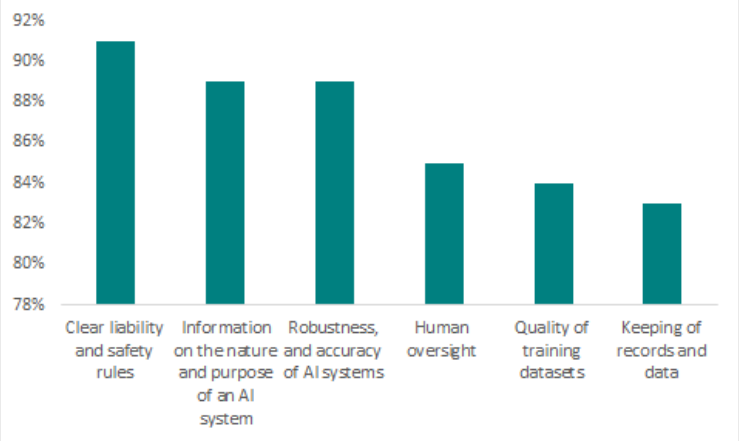
[6] 'Do you think that the concerns expressed above can be addressed by applicable EU legislation? If not, do you think that there should be specific new rules for AI systems?'

addressed through the adaptation of current legislation, in a way that new provisions do not overlap with existing ones. Standardisation (17 comments) or the provision of guidelines (14 comments) were some alternative solutions mentioned in the free text answers[7] while others mentioned that there should be a regular review of existing legislation, accounting for technological change (2 comments). On the same topic, only 3% of participants in the **online survey** thought that **current legislation is sufficient**, while the rest declared to have other opinions (18%) or no opinion at all (4%).

Mandatory requirements

The vast majority of **online respondents** seemed to overwhelmingly agree with compulsory requirements introduced by the White Paper in the case of high-risk applications. Clear liability and safety rules were supported by 91% of respondents and were followed by information on the nature and purpose of an AI system (89%), robustness, and accuracy of AI systems (89%). Human oversight (85%), quality of training datasets (84%) and the keeping of records and data (83%) followed.

Figure 1: Agreement to introduce compulsory requirements in the case of high-risk applications (in %)



Source: online survey, multiple-choice questions

In the 221 free text answers received on this topic, 35 referred to other documents and standards (e.g. German Data Ethics Commission – mentioned in 6 comments) while 33 of them called for criteria that are more detailed and definitions that would allow the limitation of requirements to high-risk applications only. However, other comments did not support a simple distinction between 'high' and 'low' risk AI. Some partly coordinated responses (16) were in favour of an impact assessment on fundamental/human rights impact assessment while others supported that all AI applications should be regulated as such (16) or based on use cases (13). Like for the question above, comments repeated that requirements should be proportionate (8) and avoid overregulation or any kind of unnecessary burdens for companies (6).[8]

In the **position papers**, the requirements were often not the main topics. When they were one of the major issues, the majority in favour of legislation was somewhat smaller. While many position papers did not mention regulatory requirements in their top three

[7] To that aim, some comments suggested changes in the GDPR and others supported that legislation should be technology neutral.

[8] Further comments to this question referred to human oversight (3), the difficulty of assessment and categorisation of AI (2), the need to align the definition of high-risk with international standards (2) and continuously review them for change (2), the use of the precautionary principle in general (2) and that of GDPR for risks related to privacy (1).

topics (54%), 23% generally agreed with the White Paper's approach to regulatory requirements for high-risk AI, while 12% generally disagreed. Some stakeholders also expressed other opinions (12%).

Among the 12% of stakeholders who expressed another opinion (47 in total), some argued that no new AI requirements were needed (7), while others asked for additional requirements (e.g. on intellectual property or AI design) to be considered (7). Other comments highlighted that the requirements must not stifle innovation (6), or that they needed to be more clearly defined (3).

'Human oversight' was the most mentioned requirement (109 mentions), followed by 'training data' (97), 'data and record keeping' (94), 'information provision' (78) and 'robustness and accuracy' (66).

Many business associations (73%) and large companies (59%) took a stance on regulatory requirements, while the other stakeholder types, including SMEs, did not take a stance on the issue as often. In addition, business stakeholders tended to broadly agree with the Commission on the issue as presented on the AI White Paper (31%). Those who expressed other opinions mainly highlighted that new rules/requirements were not needed (3.7%), or that requirements should be proportionate (2.2%).

Only 39% of academic stakeholders mentioned regulatory requirements (19). When they did, they tended to be in favour of them (22%) or they expressed other opinions (10%). The positioning of NGOs was similar: while only 38% mentioned the regulatory requirements, those who did were also mostly in favour of them (21%).

High-risk applications

Concerning the scope of this new possible legislation, participants where asked on whether it should be limited to high-risk applications only. While 42.5% of **online questionnaire** respondents agreed that the introduction of new compulsory requirements should only be limited to high-risk AI applications, another 30.6% doubted such limitation. The remaining 20.5% had other opinions and 6.3% had no opinion at all. It is interesting to note that respondents from industry and business were more likely to agree with limiting new compulsory requirements to high-risk applications with a percentage of 54.6%.

However, several online respondents did not appear to have a clear opinion regarding what high-risk means: although 59% of respondents supported the definition of high-risk provided by the White Paper[9], only 449 out of 1215 (37% of consultation participants) responded to this question.

[9] 'An AI application should be considered high-risk where it meets the following two cumulative criteria: First, the AI application is employed in a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur. This first criterion ensures that the regulatory intervention is targeted on the areas where, generally speaking, risks are deemed most likely to occur. The sectors covered should be specifically and exhaustively listed in the new regulatory framework. For instance, healthcare; transport; energy and parts of the public sector. (…)
Second, the AI application in the sector in question is, in addition, used in such a manner that significant risks are likely to arise. This second criterion reflects the acknowledgment that not every use of AI in the selected sectors necessarily involves significant risks. For example, whilst healthcare generally may well be a relevant sector, a flaw in the appointment scheduling system in a hospital will normally not pose risks of such significance as to justify legislative intervention. The assessment of the level of risk of a given use could be based on the impact on the affected parties. For instance, uses of AI applications that produce legal or similarly significant effects for the rights of an individual or a company; that pose risk of injury, death or significant material or immaterial damage; that produce effects that cannot reasonably be avoided by individuals or legal entities.' (European Commission,

In the **59 free text answers** received**,** 10 found the definition provided in the White Paper unclear and asked for more details/criteria to be provided. Other comments found problematic the clause according to which 'there may also be exceptional instances where, due to the risks at stake, the use of AI applications for certain purposes is to be considered as high-risk' (7) while some suggested additional criteria for the definition of 'high-risk' (4). Coordinated responses (4) support existing (sectorial) definitions of 'high and low risk' while others (4) suggested the identification of high-risk application instead of sectors. For others, the classification of entire sectors as 'high-risk' could bring disadvantages and hamper innovation (3). Other comments focus on the importance of 'legal certainty' (5) which could be reduced by overly frequent reviews of high-risk sectors (3)[10].

Consultation participants were also asked to indicate AI applications or uses which according to them can be considered as high-risk. The table below lists the top answers received:

Table 1: Other AI Applications and uses that can be considered as "high-risk" according to free text answers

TOP AI APPLICATIONS AND USES CONSIDERED AS "HIGH-RISK"	MIN. NO. OF MENTIONS
Applications related to autonomous weapons / defense sector	41
Remote biometric identification (e.g. facial recognition)	34
Applications in critical infrastructure (e.g. electricity, water supply, nuclear)	28
Reference to other documents/standards	25
Applications related to health	22
Applications in HR and employment	21
Applications analysing/manipulating human behaviour	18
Applications in predictive policing	18
Applications enabling mass surveillance	15
Applications used in political communication / disinformation	12
Applications related to security, law enforcement	12

The definition of 'high-risk' seemed to be the most important point of for stakeholders submitting **position papers** as well. A large number of papers commented that this definition was unclear or needed improvement (74 out of 408 stakeholders). Many believed that the simple distinction between high and low risk was too simplified and some proposed to introduce more levels of risk. Some believed that the definition was too broad, while others believed that it was too narrow.

In this context, some stakeholders proposed alternative approaches to defining 'high-risk' with more risk levels: some position papers (at least 6) suggested following a gradual approach with five risk levels, as proposed by the German Data Ethics Commission to create a differentiated scheme of risks. Other stakeholders (at least 5) suggested the

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

[10] Other comments: In favour of 'human rights impact assessments' (2). The context of use of an AI is important for assessing its risk (2). The binary separation in high/low risk is too simplified (2). The criteria for 'high risk' do not go far enough (2). Against listing the transport sector as 'high risk' (2). The risk framework should be proportionate (2). Agree with limit to high-risk applications, but should also apply to non-AI systems (2). Reference to other documents/standards (2).

adoption of risk matrixes, which combine the intensity of potential harm with the level of human implication/control in the AI decision. The probability of harm was another criterion for risk, repeatedly mentioned by stakeholders.

Similarly, many position papers addressed the two-step approach proposed in the White Paper to determining 'high-risk' AI. At least 19 position papers considered the approach inadequate, at least 5 argued against the sectoral approach and many others put forth a diverse set of suggestions and criticism.

One notable suggestion for the risk assessment approach was to take into account all subjects affected by the AI application: multiple stakeholders argued that not only individual risks, but also collective risks should be considered, as there were also risks affecting society as a whole (e.g. with regards to democracy, environment, human rights). The impression that the definition of 'high-risk' needs to be clarified was shared by all stakeholder types.

The two-step risk assessment approach received most comments from business stakeholders. At least 5 business associations and large companies argued against the sectoral approach to determining high-risk and were supportive of a contextual assessment. On the contrary, two out of the three SMEs that mentioned the risk assessment approach expressly, supported the sectoral approach.

Remote biometric identification in public spaces

Online questionnaire respondents were concerned about the public use of such systems with 28% of them supporting 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. 17.1% declared to have no opinion.

In the 257 free text answers received, participants mainly referred to the concerns that the use of biometric identification brings. According to 38 comments, remote biometric identification in public spaces endangers fundamental rights in general while other comments supported such concerns by referring to other documents, standards (30) or even the GDPR (20). 15 comments referred to the risk of mass surveillance and imbalances of power that the use of such technology may bring, 13 others referred to privacy while 10 more mentioned that biometric identification endangers the freedom of assembly/expression. However, there were also 13 comments referring to possible benefits coming from the use of this technology while 13 more mentioned that the use of remote biometric identification in public spaces should be allowed for specific purposes only, e.g. security, criminal or justice matters. Other 8 comments stressed that there are uses of facial recognition that do not pose 'high risks' or endanger fundamental rights and the introduction of guidelines could be beneficial for the correct use of the technology (8). The management of such systems by qualified staff could, according to 7 more comments, guaranty human oversight in its use.[11]

[11] Additional comments: Excessive regulation hinder innovation / imposes costs (7). Allow remote biometric identification in public spaces only if proportionate (6).More research/information is necessary (6). Allow remote biometric identification in public spaces only for specific purposes: security / criminal justice matters, only in specific cases (6). The existing framework is sufficient (6). Stakeholders should be consulted (6). Allow remote biometric identification in public spaces only under

Among the **position papers**, a part of the stakeholders specifically mentioned remote biometric identification in public spaces (96) as one of their top three topics. Of these, a few argued for a ban of remote biometric identification in public spaces (19), and 7 respondents for a moratorium. A few more were in favour of conditioning its use to tight regulation and adequate safeguards in public spaces (19). Almost half of the stakeholders who positioned themselves in favour of a ban of biometric identification in public spaces were NGOs. This contrasts with the 34 business stakeholders who mentioned biometric identification, among which only one was in favour of a ban. A moratorium for remote biometric identification in public spaces was also mentioned by academic stakeholders: four research institutions were in favour of a moratorium of biometric identification until clear and safe guidelines were issued by the EU.

Enforcement and voluntary labelling

To make sure that AI is trustworthy, secure and in respect of European values, the White Paper suggested conformity assessment mechanisms for high-risk applications. The public consultation proposed several options to ensure that AI is trustworthy, secure and in respect of European values. 62% of **online survey respondents** supported a combination of ex-post and ex-ante market surveillance systems. 3% of respondents supported only ex-post market surveillance. 28% supported external conformity assessment of high-risk applications. 21% of respondents supported ex-ante self-assessment.

To the options above, respondents added further ones through **118 free text answers**. Among those, 19 suggested an (ex-ante) assessment of fundamental rights while 14 comments were in favour of self-assessment and 11 more suggested that independent external bodies/experts should ensure assessment. There were also 8 comments supporting that existing assessment processes are sufficient while 8 others where against ex-ante assessment as that might be a burden for innovation.

Voluntary labelling systems could be used for AI applications that are not considered of high-risk. 50.5% of online respondents found it useful or very useful, while another 34% did not agree with the usefulness of such system. 15.5% of respondents declared that they did not have an opinion on the matter.

Still, in **301 free text answers,** 24 comments appeared to be generally in favour of voluntary labelling, 6 more supported self-assessment and 46 more made reference to other documents and existing international standards that could be used as an example for such practices (e.g. the AI HLEG's assessment list, the energy efficiency label[12] or the EEE EPPC and IEEE-SA). According to 6 comments, labelling systems need to be clear and simple while 18 comments stressed that clearer definitions and details are needed. Other 16 comments called for the involvement of stakeholders in development of labelling systems or (7 more comments) the appointment of an independent body should be responsible for the voluntary labelling. The importance of enforcement and control of voluntary labelling was stressed by 12 more comments and a harmonised EU-wide approach was suggested by 5 others. Moreover, 8 comments mentioned that systems need to be flexible to adapt to technological changes.

---

other specific conditions (5). Facial recognition may be needed for autonomous vehicles (coordinated response, car makers) (5). Legislation needs to be clear and simple (4). The definition of 'public space' is unclear (4). Strict rules for the storage of biometric data are important (3). Remote biometric identification in public spaces is useful for social distancing during the COVID-19 epidemic (3). Regulation should only be considered in case of consumer harm (2). Human oversight is overestimated (2). A moratorium would leave the field to other, less free countries and reduce accuracy of systems (2). Are vehicles a 'public space'? (2). EU-level harmonisation is important (2).

[12] European Commission, *About the energy label and eco-design*, 2020.

However, 27 replies seemed to be sceptical towards voluntary labelling systems in general and 25 more towards self-labelling/self-regulation in particular. Some of these comments mentioned that such systems may be used according to the interest of companies, according to 16 more, it is likely that such systems favour bigger players who can afford it while 23 more stressed it imposes costs that can hamper innovation for smaller ones. Moreover, 12 comments mentioned the issue of labelling for 'low risk' categories, which can create a false sense of risks, others 7 comments mentioned that the distinction among low and high risk is too simplified while 5 more said that they can create a false sense of security.[13]

52 **position papers** addressed the proposed voluntary labelling scheme as one of their top three topics. 21 of them were sceptical of labelling, either because they believed that it would impose regulatory burdens (especially for SMEs) or because they were sceptical of its effectiveness. Some stakeholders argued that such a scheme was likely to confuse consumers instead of building trust. On the other hand, 8 position papers were explicitly in favour, and many other stakeholders provided a diverse set of comments.

The voluntary labelling scheme received most comments through position papers submitted by business stakeholders: most of business associations (11) and SMEs (3) were sceptical of the idea, due to the costs it could impose on them or a suspected lack of effectiveness. The position of large companies mentioning voluntary labelling was quite the opposite: most tended to be in favour of it (4).

Safety and liability implications of AI, IoT and robotics

The overall objective of the safety and liability legal frameworks is to ensure that all products and services, including those integrating emerging digital technologies, operate safely, reliably and consistently, and that damage that has already occurred is remedied efficiently.

60.7% of **online respondents** supported a revision of the existing Product Liability directive to cover particular risks engendered by certain AI applications. 63 % of respondents supported that national liability rules should also be adapted for all AI applications (47 %) or specific AI applications (16 %) to better ensure a proper compensation in case of damage, and a fair allocation of liability. Amongst those businesses that took a position on this question (i.e. excluding ‘no opinion’ responses), there is equally clear support for such adaptations, especially amongst SMEs (81 %).

Among the particular AI related risks to be covered, **online respondents** prioritised cyber risks with 78% and personal security risks with 77%. Mental health risks followed with 48% of respondents flagging them, and then risks related to the loss of connectivity, flagged by 40% of respondents. Moreover, 70% of participants supported that the safety legislative framework should consider a risk assessment procedure for products subject to important changes during their lifetime.

In **163 free text answers**, 23 respondents added to the risks those of discrimination/manipulation, which according to 9 others can be caused by profiling practices or automated decision making (5 comments), while 14 more (mainly NGOs) focused on the particular discrimination risk linked to online advertisement. This can also be related to another set of comments (14 in total) according to which, such risks may cause differentiated pricing, financial detriments, filter bubbles or interference in political

[13] Additional comments: All AI should be regulated (5). In favour of a mandatory labelling system (4). In B2B trust is created through contractual agreements (3). Standards need to be actively promoted to become effective (2). Not products/services should be labelled, but an organisation's quality of AI governance (2).

processes (other 2 comments mentioning the risks of disinformation can be relevant here as well). Risks to personal data (11 comments), or those deriving from cyber-attacks (7 comments), risks for people with disabilities (10 comments) as well as general health risks (8 comments) were among other risks mentioned.[14] For the specific risks deriving from cyber security and connectivity loss in the automotive sector, a coordinated response of four carmakers, noted that other regulations tackle them already.

In the **173 free text answers** regarding the risk assessment procedures of the safety and liability framework, as pointed by 11 comments 'AI systems change over time'. Therefore, 16 comments mention that risk assessments need to be repeated in case of changes (after placement on the market). To the same regard, 13 comments pointed that clearer definitions of e.g. 'important changes' should be given during that process and 11 others that risk assessment should only be required in case of a significant change to a product (partly coordinated response).

According to 12 comments, assessment procedures could build up on the existing GDPR Impact Assessment or even involve GDPR and data protection officers (coordinated response of 10 stakeholders).[15]

52 **position papers** addressed issues of liability as one of their top three topics, most of them providing a diverse set of comments. 8 believed that existing rules were probably sufficient and 6 were sceptical of a strict liability scheme. Those who were sceptical often argued that a strict liability scheme was likely to stifle investment and innovation, and that soft measures like codes of conduct or guidance documents were more advisable. At the same time, other contributions to the public consultation from the entire range of stakeholders expressed support for a risk-based approach also with respect to liability for AI, and suggested that not only the producer, but also other parties should be liable. Representatives of consumer interests stressed the need for a reversal of the burden of proof.

When it comes to liability, some business associations and large companies thought that existing rules were probably already sufficient (7) or they were sceptical of strict liability rules and possible regulatory burdens (5). Almost none of the other stakeholder types shared this position. A few businesses submitted position papers in favour of harmonising liability rules for AI.

Other issues raised in the position papers

The position papers submitted also raised some issues that were not part of the questionnaire.

How to define artificial intelligence? (position papers only)

As the White Paper does not contain its own explicit definition of AI, this analysis of the position papers took the definition of the HLEG on AI as a reference point. The HLEG

[14] Additional comments: Risks caused by autonomous driving / autonomous systems (5). Risks linked to loss of control / choice (7). Weapons / lethal autonomous weapon systems (4). Risks for fundamental rights (3). Risks for nuclear safety (2). Significant material harm (2). Risks to intellectual property (IP) (2). Risks to employment (1).

[15] Additional comments: Recommendations on when the risk assessment should be required (8). There is no need for new AI-specific risk assessment rules (7). Existing bodies should be involved and properly equipped (4). Independent external oversight is necessary (not specified by whom) (4). Overly strict legislation can be a barrier for innovation and impose costs (4). New risk assessment procedures are not necessary (4). Trade unions should be involved (3). Long-term social impacts should be considered (3).Human oversight / final human decisions are important (3). Fundamental rights are important in the assessment (2). Legal certainty is important (2). Risk assessments are already obligatory in sectors like health care (2).

definition of AI includes systems that use symbolic rules or machine learning, but it does not explicitly include simpler Automatic Decision Making (ADM) systems.

Position papers were analysed to determine whether and why stakeholders shared or did not share this definition or have other interesting comments on the definition of AI.

The majority of position papers made no mention of the definition of AI (up to 70%, or 286 out of 408) among their top three topics. A majority of 15.7% had a different definition than the one suggested by the HLEG (64). 9.3% found the definition was too broad (37), out of which 2.7% said that AI should only include machine learning (11). Stakeholders highlighted that a too broad definition risks leading to overregulation and legal uncertainty, and was not specific enough to AI. Another 6.6% believed that the definition was too narrow (27), with 3.7% saying that it should also include automated decision-making systems (15). Stakeholders highlighted that the definition needed to be future proof: if it was too narrow, it risks disregarding future aspects of next-generation AI.

2.7% of stakeholders agreed with the AI HLEG definition of AI (11) but 5.4% of position papers stated that the AI HLEG's definition is unclear and needs to be refined (22). To improve the definition, stakeholders propose, for example: to clarify to what extent the definition covers traditional software; to distinguish between different types of AI; or to look at existing AI definitions made by public and private organisations. Finally, 2.2% of stakeholders provided their own definition of AI (9).

The majority of business stakeholders believed that the AI HLEG's definition was too broad. This trend was strongest for business associations. On the contrary, the majority of academic and NGO stakeholders believed that the HLEG's definition is too narrow.

At least 24 business stakeholders believed that the definition was too broad, while only 5 believed that it was too narrow and only 4 agreed with it. Business stakeholders were also relatively numerous in saying that the definition is unclear or needs to be refined (at least 11). The majority of academic and NGO stakeholders believed that the AI HLEG's definition was too narrow (6 and 8) and only 1 academic and 4 NGO stakeholders believed that the definition was too broad.

Costs - What costs could AI regulation create? (Position papers only.)

Costs imposed by new regulations are always a contentious topic. Some see costs imposed by regulation as an unnecessary burden to competitiveness and innovation; others see costs as a necessary by-product of making organisations comply with political, economic or ethical objectives.

In order to better understand stakeholder's perspective on the costs of AI regulation, position papers were analysed for mentions of two main types of costs: (1) compliance costs, generally defined as any operational or capital expense faced by a company to comply with a regulatory requirement; and (2) administrative burdens, a subset of compliance costs, covering 'red tape' such as obligations to provide or store information.

84% of stakeholders do not explicitly mention costs that could be imposed by a regulation on AI as one of the top three topics (344). 11% of stakeholders (46) mention compliance costs in general and 7% of stakeholders (29) (also) mention administrative burdens in particular. It must be noted that some stakeholders mentioned both types of costs.

Some stakeholders warned against the costs incurred by a mandatory conformity assessment, especially for SMEs or companies operating on international markets. Some highlighted that certain sectors were already subject to strict ex-ante conformity controls

(e.g. automotive sector) and warned against the danger of legislative duplication. Several stakeholders also saw a strict liability regime as a potential regulatory burden and some noted that a stricter regime could lead to higher insurance premiums.

Some respondents also put forth other arguments related to costs, such as the potential cost saving effects of AI, the concept of 'regulatory sandboxes' as a means to reduce regulatory costs, or the environmental costs created by AI due to high energy consumption.

17% of all types of business stakeholders mentioned compliance costs and 13% (also) mentioned administrative burdens, while up to 74% of business stakeholders did not explicitly mention costs among their top three topics. Among business stakeholders, business associations are the ones that mentioned costs the most. Out of all mentions of costs from all stakeholders (75 in total), 56% came from business stakeholders (42).

Academic stakeholders also mentioned costs more often than other types of stakeholders, but also not very often overall. 13% of academic stakeholders mentioned compliance costs and 9% (also) mentioned administrative burdens, while 82% did not explicitly mention costs in their top three topics. Other stakeholders mentioned costs more rarely.

Governance - Which institutions could oversee AI governance? (Position papers only)

The institutional structure of AI governance is a key challenge for the European regulatory response to AI. Should AI governance, for example, be centralised in a new EU agency, or should it be decentralised in existing national authorities, or something in between? In order to better understand this issue, the position papers were analysed regarding their position on the European institutional governance of AI.

Most stakeholders (up to 77% or 314) did not address the institutional governance of AI.

Among the 23% of position papers who did address this issue in their top three topics, 10% of stakeholders were in favour a new EU-level institution, with 6% of stakeholders being in favour of some form of a new EU AI agency (24) and 4% in favour of a less formalised EU committee/board (15). At the same time, at least 3% of stakeholders were against establishing a new institution (14): they argued that creating an additional layer of AI-specific regulators could be counterproductive, and they advocated for a thorough review of existing regulation frameworks, e.g. lessons learned from data protection authorities dealing with GDPR, before creating a new AI-specific institution/body.

1% of stakeholders were in favour of governance through national institutions (6) and another 1% of stakeholders were in favour of governance through existing competent authorities (5) (without specifying whether these would be on the EU or national level). In addition, stakeholders also mentioned other ideas, such as the importance of cooperation between national and/or EU bodies (7); multi-stakeholder governance involving civil society and private actors (6); or sectorial governance (4).

While only 32% of academic stakeholders mention the issue in their position papers among the top three topics, they tended to be in favour of an EU AI agency (10%), but many provided a diverse set of other arguments. 24% of large companies and business associations provided a position on the issue while SMEs practically did not mention it. All business stakeholders tended to be more sceptical of formal institutionalisation: 8% of business associations and 4% of large companies are against a new institution, 5% of associations and 2% of large companies are in favour of a less formalised committee/board, and the others share other more specific positions.

Most trade unions and EU or non-EU citizens did not have a position on the issue, but if they did, the majority was in favour of an EU AI agency (25% of trade unions and 17% of EU and non-EU citizens). However, it must be noted that these percentages are very volatile due to the low number of respondents with a position on the issue.

2.2. Analysis of the results of the feedback from the inception impact assessment

The Inception Impact Assessment elicited 132 contributions from 130 different stakeholders – two organizations commented twice – from 24 countries all over the world. 89 respondents out of 130 had already answered the White Paper consultation.

Table 2: Participating Stakeholders (by type)

STAKEHOLDER TYPE	NUMBER
Business Association	55
Company/Business Organization	28
NGO	15
EU citizen	7
Academic/Research Institution	7
Other	6
Consumer Organization	5
Trade Union	4
Public Authority	3

Table 3: Participating Stakeholders (by country)

COUNTRY	NUMBER	COUNTRY	NUMBER
Belgium	49	Finland	2
Germany	17	Hungary	2
US	11	Poland	2
Netherlands	8	Portugal	2
UK	8	Sweden	2
France	6	Bulgaria	1
Ireland	3	Czech Republic	1
Italy	3	Estonia	1
Spain	3	Japan	1
Austria	2	Lithuania	1
Denmark	2		

Summary of feedback

Stakeholder mostly requested a narrow, clear and precise definition for AI. Stakeholders also highlighted that besides the clarification of the term of AI, it is important to define 'risk', 'high-risk', 'low-risk', 'remote biometric identification' and 'harm'.

Some of the stakeholders caution the European Commission not to significantly expand the scope of future AI regulation to ADM, because if AI were to be defined as ADM, it would create regulatory obligations that hamper development.

Several stakeholders warn the European Commission to avoid duplication, conflicting obligations and overregulation. Before introducing new legislation, it would be crucial to clarify legislative gaps, to adjust the existing framework, focus on effective enforcement and adopt additional regulation only where necessary. It is essential to review EU legislation in other areas that are potentially applicable to AI and make them fit for AI. Before choosing any of the listed options, existing regulation needs to be carefully analysed and potential gaps precisely formulated.

There were many comments underlining the importance of a technology neutral and proportionate regulatory framework.

Regulatory sandboxes could be very useful and are welcomed by stakeholders, especially from the Business Association sector.

Most of the respondents are explicitly in favour of the risk-based approach. Using a risk-based framework is a better option than blanket regulation of all AI applications. The types of risks and threats should be based on a sector-by-sector and case-by-case

approach. Risks also should be calculated taking into account the impact on rights and safety.

Only a few respondents agreed that there is no need for new regulation for AI technologies: option 0 "baseline". Less than 5% of the stakeholders supported option 0.

There was a clear agreement among those stakeholders who reflected on option 1 that either per se or in combination with other options, 'soft law' would be the best start. Around one third of the stakeholders commented option 1 and more than 80% of them were in favour of it. Most of the supportive comments arrived from the business association (more than 75%) and company/business sector.

Option 2 'voluntary labelling system' per se was not supported, since it seems to be premature, inefficient and ineffective. More than one third of the stakeholders had a view on the voluntary labelling system of which nearly 75% disagreed with option 2. It is argued mostly by the business association, company/business and NGO sectors that voluntary labelling could create heavy administrative burden and would only be useful if it is flexible, robust and clearly articulated. If the Commission would decide to introduce voluntary certification, it should be carefully addressed as it can result in a meaningless label and even increase non-compliant behaviour when there are no proper verification mechanisms.

The three sub-options 3a (legislation for specific AI applications), 3b (horizontal framework for high-risk AI applications) and 3c (horizontal framework for all AI applications) were commented on by more than 50% of the respondents. There is a majority view – more than 90% of the stakeholders who reflected on this question – that if legislation is necessary, the EU legislative instrument should be limited to 'high-risk' AI applications based on the feedback mostly of business associations, companies and NGOs. Legislation limited only to specific applications could leave some risky application out of the regulatory framework.

The combination of different options was a very popular choice, nearly one third of the respondents supported option 4 'combination of any of the options above'. Most variations included option 1 'soft law'. The most favoured combination with nearly 40% was option 1 'soft law' with sub-option 3b 'high-risk applications', sometimes with sub-option 3a 'specific applications'. Mainly business associations and companies supported this combination. Especially NGOs, EU citizens and others preferred the combination of option 2 'voluntary labelling' and sub-option 3b. In small numbers, option 1, option 2 and sub-option 3b, the combinations of option 2, sub-option 3a and/or sub-option 3b were also preferred. Option 1 and sub-option 3b are often viewed favourably per se or in combination also. Sub-option 3c was not popular at all.

Among those who formulated their opinion on the enforcement models, more than 50%, especially from the business association sector were in favour of the combination of ex-ante risk self-assessment and ex-post enforcement for high-risk AI applications.

In case of an ex-ante enforcement model, there are many respondents who caution against third party ex-ante assessments and instead recommend self-assessment procedures based on clear due diligence guidance. New ex-ante conformity assessments could cause significant delays in releasing AI products has to be taken into account. Ex-ante enforcement mechanisms without any background are causing a lot of uncertainty. Ex-ante conformity assessments could be disproportionate for certain applications.

If ex-post enforcement would be chosen, it should be used with the exception of sectors where ex-ante regulation is a well-established practice. Ex-post enforcement should only be implemented in a manner that complements well against ex-ante approaches.

2.3. Stakeholder outreach

The following consultation activities (in addition to the open public consultation and the Inception Impact Assessment feedback) were organised:

2.3.1. Event on the White Paper with larger public

In addition to the public consultations, the Commission also consulted stakeholders directly. On 9 October 2020, it organised the Second European AI Alliance Assembly with more than 1 900 participants across different stakeholder groups, where the issues addressed in the impact assessment were intensely discussed. The topical workshops held during the event on the main aspects of the AI legislative approach included biometric identification, AI and liability, requirements for Trustworthy AI, AI Conformity assessment, standards and high-risk AI applications. The AI Alliance is a multi-stakeholder forum launched in June 2018 in the framework of the European Strategy on Artificial Intelligence. During the conference, participants could interact with the different panels, which were made up of diverse stakeholders, through Sli.do. Overall, there were 647 joined participants, with 505 active participants. 338 questions were asked, and attracted over 900 likes among themselves. Over the course of the day, over 1 000 poll votes were cast.

2.3.2. Technical consultations

The Commission organised five online workshops with experts from different stakeholder groups:

- online workshop on conformity assessment on 17 July 2020 with 26 participants from the applying industry, civil society and conformity assessment community;
- online workshop on biometrics on 3 September 2020 with 17 external participants from stakeholders such as the Fundamental Rights Agency, the World Economic Forum, the French Commission Nationale de l'Informatique et des Libertés and academia;
- online workshops on standardisation on 29 September 2020 with 27 external participants from UNESCO, OECD, Council of Europe, CEN-CENELEC, ETSI, ISO/IEC, IEEE, ITU;
- online workshop on potential requirements on 9 October 2020 with 15 external experts on AI, mainly from academia;
- online workshop on children's rights and AI on 12 November 2020 with external experts.
- AI expert group for home affairs on surveillance technologies and data management by law enforcement on 17 December 2020.

In addition, the contractor for the analytical study organised two online workshops with experts as follows:

- online validation workshop on costs assessment on 28 September 2020 with 40 external experts;
- online validation workshop conformity assessment on 7 October 2020 with 25 external experts.

The Commission services also participated in many seminars (more than 50) and held a numerous meetings with a large variety of stakeholders from all groups.

2.3.3. Outreach and awareness raising events in Member States and International outreach

Due to the coronavirus, the planned outreach activities in Member States had to move online and were fewer than initially planned. Nevertheless, Commission services discussed the approach in meetings with large numbers of stakeholders in several Member States, including France, Germany and Italy. They also exchanged views with international bodies, in particular the Council of Europe, the G8 and G20 as well as the OECD. The EU approach was discussed in bilateral meetings with a number of third countries, for example Japan and Canada.

2.3.4. European AI Alliance platform

The Commission also used the European AI Alliance, launched in June 2018, which is a multi-stakeholder online platform intended for broad engagement with academia, industry and civil society to discuss the European AI and gather input and feedback. The European AI Alliance has more than 3 700 members representing a wide range of fields and organisations (public authorities, international organisations, consumer organisations, industry actors, consultancies, professional associations, NGOs, academia, think tanks, trade unions, and financial institutions). All Member States are represented, as well as non-EU countries.

3. ANNEX 3: WHO IS AFFECTED AND HOW?

3.1. Practical implications of the initiative

3.1.1. Economic operators/business

This category comprises **developers of AI applications, providers that put AI applications on the European market and operators/users of AI applications** that constitute a particularly **high risk** for the safety or fundamental rights of citizens. The initiative applies to AI systems operated or used in Europe and the respective operators, independent of whether they are based in Europe or not. According to their respective role in the AI life-cycle they would all have to comply with clear and predictable obligations for taking measures with a view to preventing, mitigating and monitoring risks and ensuring safety and respect of fundamental rights throughout the whole AI lifecycle. Before placing their product on the market, providers in particular will have to ensure that the high-risk AI systems comply with essential requirements, addressing more specifically the underlying causes of risks to fundamental rights and safety (such as requirements relating to data, traceability and documentation, transparency of AI systems and information to be provided, robustness and accuracy and human oversight). They will also have to put in place appropriate quality management and risk management systems, including to identify and minimise risks and test the AI system ex ante for its compliance with the requirements and relevant Union legislation on fundamental rights (e.g. non-discrimination).

Once the system has been placed on the market, providers of high-risk AI systems would be obliged to continuously monitor, manage and mitigate any residual risks, including reporting to the competent authorities incidents and breaches of fundamental rights obligations under existing Union and Member States law.

Where feasible, the requirements and obligations will be operationalised by means of **harmonized standards** that may cover the process and the requirements (general or specific to the use case of the AI system). This will help providers of high-risk AI systems to reach and demonstrate compliance with the requirements and improve consistency.

In addition, for a subset of the high-risk applications (safety components of products and remote biometric identification in publicly accessible spaces), companies would have to submit their applications to third-party ex-ante conformity assessment bodies before being able to place them on the market. When harmonized standards exist and the providers apply those standards, they would not be required to undergo an ex-ante third party conformity assessment; this option would be applicable to safety components depending on the relevant sectoral safety rules for conformity assessment. For all other high-risk applications, the assessment would be carried out via an ex ante conformity assessment though internal checks.

For non-high risk AI systems, the instrument will impose minimal requirements and obligations for increased transparency in two limited cases: obligation to disclose that the human is interacting with an AI system and label deep fakes when not used for legitimate purposes.

The initiative will give rise to new compliance costs. Apart from authorisation and on-going supervisory costs, developers, providers and operators will need to implement a range of operational changes. The individual costs arising from this will largely depend on the extent to which respective AI developers, providers and operators have already implemented measures on a voluntary basis. An EU regulatory framework, however, avoids the proliferation of nationally fragmented regimes. It will thus provide AI system

developers, operators and providers with the opportunity to offer services cross-border throughout the EU without incurring additional compliance costs. As the initiative pre-empts the creation of national regimes in many Member States, there can be a significant indirect cost saving in this regard for cross-border operations. Concerning AI system developers, the initiative aims to facilitate competition on a fair basis by creating a regulatory level playing field. It will also help to strengthen consumer and investor trust and should thereby generate additional revenue for AI systems developers, providers and operators.

3.1.2. Conformity assessment, standardisation and other public bodies

Standardisation bodies will be required to develop standards in the field.

Conformity assessment bodies would have to establish or adapt conformity assessment procedures for the products covered. In case of third-party conformity assessment they also would have to carry them out.

Member States would have to equip **competent national authorities** (e.g. market surveillance bodies etc.) adequately to supervise the enforcement of the requirements, including the supervision of the conformity assessment procedures and also the ex-post market monitoring and supervision. The ex-post system will monitor the market and investigate compliance with the obligations and requirements for all high-risk AI systems already placed on the market and used in order to effectively enforce the existing rules and sanction non-compliance.

Authorities will also have to participate in meetings as part of a coordination mechanism at EU level to provide uniform guidance about the interpretation of the new rules and consistency.

The Commission will also encourage **voluntary compliance with codes of conduct** developed by industry and other associations.

Supervisors will face a range of new tasks and supervisory obligations stemming from the framework. This has cost implications, both as concerns one-off investments and ongoing operational costs. Supervisors will need to invest in particular in new monitoring systems and ensure a firm enforcement of regulatory provisions. They will also need to train staff to ensure sufficient knowledge of these newly regulated markets and employ additional employees to stem the additional work. The costs for specific national authorities depends on (1) the number of AI applications monitored, and (2) the extent to which other monitoring systems are already in place.

3.1.3. Individuals/citizens

Citizens will benefit from an increased level of safety and fundamental rights protection and higher market integrity. The mandatory information and transparency requirements for high-risk AI systems and enforcement rules will enable citizens to make more informed decisions in a safer market environment. They will be better protected from possible activities that might be contrary to the EU fundamental rights or safety standards. In summary, citizens will carry lower risks, given the European regulatory approach. It can however not be excluded, that some of the compliance costs will be passed on to the citizens.

3.1.4. Researchers

There will be a boost for research, since some of the requirements (such as those related to robustness) will require continuous research and testing of products.

3.2. Summary of costs and benefits

Table 4: Overview of Benefits (total for all provisions) – Preferred Option

DESCRIPTION	AMOUNT	COMMENTS
Direct benefits		
Fewer risks to safety and fundamental rights	Not quantifiable	Citizens
Higher trust and legal certainty in AI	Not directly quantifiable	Businesses
Indirect benefits		
Higher uptake	Not directly quantifiable	Businesses
More beneficial applications	Not quantifiable	Citizens

Not quantifiable: impossible to calculate (e.g. economic value of avoiding fundamental rights infringements)

Not directly quantifiable: could in theory be calculated if many more data were available (or making large numbers of assumptions)

Table 5: Overview of costs – Preferred option

		CITIZENS/ CONSUMERS		BUSINESSES		ADMINISTRATIONS	
		One-off	**Recurrent**	**One-off**	**Recurrent**	**One-off**	**Recurrent**
Comply with substantial require-ments	**Direct costs**			€ 6000 – 7000 per application	€ 5000 – 8 000 per application		
	Indirect costs						
Verify compliance	**Direct costs**			€ 3000 – 7500 per application			
	Indirect costs			Audit QMS €1000 – 2000 per day, depending on complexity	Renew audit, €300 per hour, depending on complexity		
Establish competent authorities	**Direct costs**						1-25 FTE per MS; 5 FTE at EU
	Indirect costs						

4. ANNEX 4: ANALYTICAL METHODS

Summary of the elements of the compliance costs and administrative burden

This annex summarises the key elements of the compliance costs and administrative burdens for enterprises, based on 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.*[16]

The cost assessment achieved by the consultant relies on the Standard Cost Model, a widely known methodology to assess administrative burdens. It has been adopted by several countries around the world, including almost all EU Member States and the European Commission in its Better Regulation Toolbox.

A specific version of the model is used in this case: it features standardised tables with time estimates per administrative activity and level of complexity. The cost estimation is built on time expenditure for activities induced by the new requirements under the proposed regulation. The assessment is based on cost estimates of an average AI *unit* of an average firm, estimated to cost around USD 200,000 or EUR 170,000[17].

The costs assessed here refer to two kinds of direct compliance costs:

- *Substantive compliance costs*, which encompass those investments and expenses faced by businesses and citizens to comply with substantive obligations or requirements contained in a legal rule. These costs are calculated as a sum of capital costs, financial costs and operating costs.
- *Administrative burdens* are those costs borne by businesses, citizens, civil society organisations and public authorities as a result of administrative activities performed to comply with the information obligations (IOs) included in legal rules.

The approach broadly corresponds to the methodology adopted by the German government and developed with the Federal Statistical Office (Destatis). The table below shows a correspondence table used for the cost assessment in this document that allocate specific times to specific activities, differentiating each activity in terms of complexity levels.

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

[17] For AI costs, see https://www.webfx.com/internet-marketing/ai-pricing.html, https://azati.ai/how-much-does-it-cost-to-utilize-machine-learning-artificial-intelligence/ and https://www.quytech.com/blog/ai-app-development-cost/

Reference table for the assessment of compliance costs

	Time			Cost (Euros)		
	Easy	Moderate	Complex	Easy	Moderate	Complex
Administrative activities						
Familiarising oneself with the Information obligation	3	3	60	1,60	1,60	32,00
Procuring data	2	10	120	1,07	5,33	64,00
Filling in forms, labelling, classifying	3	5	30	1,60	2,67	16,00
Performing calculations	3	20	185	1,60	10,67	98,67
Checking data and inputs	1	8	60	0,53	4,27	32,00
Correcting errors	2	10	60	1,07	5,33	32,00
Processing data	3	20	240	1,60	10,67	128,00
Transmitting and publishing data	1	2	5	0,53	1,07	2,67
Internal meetings	6	60	600	3,20	32,00	320,00
External meetings	10	60	480	5,33	32,00	256,00
Payment	1	3	23	0,53	1,60	12,27
Photocopying, filing, distribution	1	2	10	0,53	1,07	5,33
Cooperating in an audit by public authorities	5	60	540	2,67	32,00	288,00
Corrections which have to be made as a result of the audit	4	30	480	2,13	16,00	256,00
Procuring additional information in case of	3	15	120	1,60	8,00	64,00
Training courses	2	30	480	1,07	16,00	256,00
Substantive costs						
Procuring goods and services						
Procuring services and/or hiring additional staff						
Supplying own services						
Adjustment of internal processes						
Supervisory measures						
Storage, inventory management, production						

Source: Consultant's elaboration based on Normenkotrollrat (2018)

The translation of activities into cost estimates was obtained by using a **reference hourly wage rate of EUR 32**, which is the average value indicated by Eurostat for the Information and Communication sector (Sector J in the NACE rev 2 classification)[18].

Two workshops were organised to discuss the cost estimates, one with businesses and one with accreditation bodies and standardisation organisations were invited to another workshop to discuss the team's estimates on conformity costs.

Compliance costs regarding data

This requirement, as defined in the White Paper (pp.18-19), includes the following main activities:

- Providing reasonable assurances that the use of the products or services enabled by the AI system is safe (e.g. ensuring that AI systems are trained on datasets that are sufficiently broad and representative of the European context to cover all relevant scenarios needed to avoid dangerous situations).
- Take reasonable measures to ensure that the use of the AI system does not lead to outcomes entailing prohibited discrimination, e.g. obligation to use sufficiently representative datasets, especially to ensure that all relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination are appropriately reflected.
- Ensuring that privacy and personal data are adequately protected during the use of AI-enabled products and services. For issues falling within their respective scope, the GDPR and the Law Enforcement Directive regulate these matters.

[18] Stakeholders' feedback suggests that EUR 32 is too low, but they are operating in more advanced economies. Given the economic differences across the EU, the EU average is a reasonable reference point here.

Thus, the types of activities that would be triggered by this requirement include, among others:

- familiarising with the information obligation (one-off);
- assessment of data availability (this may require an internal meeting);
- risk assessment (this may require an internal meeting);
- testing for various possible risks, including safety-related and fundamental rights-related risks, to then adopt and document proportionate mitigating measures;
- anonymisation of datasets, or reliance on synthetic datasets; or implementation of data minimisation obligations;
- collecting sufficiently broad datasets to avoid discrimination.

For an average process, and a normally efficient firm, a reasonable cost estate for this activity is €2763.[19]

Administrative burden regarding documents and traceability

This requirement aims to enable the verification and enforcement of compliance with existing rules. The information to be kept relates to the programming of the algorithm, the data used to train high-risk AI systems, and, in certain cases, keeping the data themselves. The White Paper (p. 19) prescribes the following actions:

- Keeping accurate records of the dataset used to train and test the AI system, including a description of the main characteristics and how the dataset was selected;
- Keeping the datasets themselves;
- Keeping documentation on programming and training methodologies, processes and techniques used to build, test and validate the AI system;
- Keeping documentation on the functioning of the validated AI system, describing its capabilities and limitations, expected accuracy/error margin, the potential 'side effects' and risks to safety and fundamental rights, the required human oversight procedures and any user information and installation instructions;
- Make the records, documentation and, where relevant, datasets available on request, in particular for testing or inspection by competent authorities.
- Ensure that confidential information is protected (e.g. trade secrets).

As a result, this obligation requires a well-trained data officer with the necessary legal knowledge to manage data and records and ensure compliance. The cost could be shared among different products and the data officer could have other functions, too. For an average process, and an efficient firm, a reasonable cost estimate per AI product for this activity is €4 390.[20]

Administrative burden regarding provision of information

[19] See support study chapter 4, section 4.2.1.
[20] See support study chapter 4, section 4.2.2.

Beyond the record-keeping requirements, adequate information is required on the use of high-risk AI systems. According to the White Paper (p. 20), the following requirements could be considered:

- Ensuring clear information is provided on the AI system's capabilities and limitations, in particular the purpose for which it is intended, the conditions under which it can be expected to function as intended, and the expected level of accuracy in achieving the specified purpose. This information is especially important for deployers of the systems, but it may also be relevant to competent authorities and affected parties.
- Making it clear to citizens when they are interacting with an AI system and not a human being.

Hence, the types of activities that would be triggered by this requirement include:

- **Provide information on the AI system's characteristics, such as**
  - Identity and contact details of the provider;
  - Purpose and key assumptions/inputs to the system;
  - What the model is designed to optimise for, and the weight according to the different parameters;
  - System capabilities and limitations;
  - Context and the conditions under which the AI system can be expected to function as intended and the expected level of accuracy/margin of error, fairness, robustness and safety in achieving the intended purpose(s);
  - Potential 'side effects' and safety/fundamental rights risks;
  - Specific conditions and instructions on how to operate the AI system, including information about the required level of human oversight.
- Provide information on whether an AI system is used for interaction with humans (unless immediately apparent).
- Provide information on whether the system is used as part of a decision-making process that significantly affects the person.
- Design AI systems in a transparent and explainable way.
- Respond to information queries to ensure sufficient post-purchase customer care. This activity was stressed by stakeholders with experience in GDPR compliance.

Given the overlaps with activities foreseen under other requirements, only the familiarisation with the specific information obligations and their compliance has been computed, rather than the cost of the underlying activities. However, it is worth noting that this requirement may also entail changes in the design of the system to enable explainability and transparency.

For an average process, and a normally efficient firm, a reasonable cost estimate for this activity is €3 627.[21]

Compliance costs regarding human oversight

[21] See support study chapter 4, section 4.2.3.

The White Paper acknowledges that the type and degree of human oversight may vary from one AI system to another (European Commission, 2020a, p.21). It will depend, in particular, on the intended use of the AI system and the effects of that use on affected citizens and legal entities. For instance:

- Output of the AI system does not become effective unless it has been previously reviewed and validated by a human (e.g. the rejection of an application for social security benefits may be taken by a human only).
- Output of the AI system becomes immediately effective, but human intervention is ensured afterwards (e.g. the rejection of an application for a credit card may be processed by an AI system, but human review must be possible afterwards).
- Monitoring of the AI system while in operation and the ability to intervene in real time and deactivate (e.g. a stop button or procedure is available in a driverless car when a human determines that car operation is not safe).
- In the design phase, by imposing operational constraints on the AI system (e.g. a driverless car shall stop operating in certain conditions of low visibility when sensors may become less reliable, or shall maintain a certain distance from the vehicle ahead in any given condition).

Therefore, the possible activities involved in compliance with this requirement are the following, based on the questions of the Assessment List for Trustworthy Artificial Intelligence[22] developed by the High-Level Expert Group on Artificial Intelligence:

- monitoring the operation of the AI system, including detection of anomalies, dysfunctions, and unexpected behaviour;
- ensuring timely human intervention, such as a "stop" button or procedure to safely interrupt the running of the AI system;
- conducting revisions in the design and functioning of the currently deployed AI system as well as implementing measures to prevent and mitigate automation bias on the side of the users;
- overseeing overall activities of the AI system (including its broader economic, societal, legal and ethical impact);
- implementing additional hardware/software/systems assisting staff in the above-mentioned tasks to ensure meaningful human oversight over the entire AI system life cycle;
- implementing additional hardware/software/systems to meaningfully explain to users that a decision, content, advice or outcome is the result of an algorithmic decision, and to avoid that end-users over-rely on the AI system.

This leads to a total estimate of €7 764.[23]

Compliance costs regarding robustness and accuracy

According to the White Paper on Artificial Intelligence (European Commission, 2020a, p. 20), 'AI systems must be technically robust and accurate if they are to be trustworthy. These systems, therefore, need to be developed in a responsible manner and with ex ante due and proper consideration of the risks they may generate. Their development and

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

[23] See support study chapter 4, section 4.2.4.

functioning must be such to ensure that AI systems behave reliably as intended. All reasonable measures should be taken to minimise the risk of harm.' Accordingly, the following elements could be considered:

- Requirements ensuring that the AI systems are robust and accurate, or at least correctly reflect their level of accuracy, during all lifecycle phases;
- Requirements ensuring that outcomes can be reproduced;
- Requirements ensuring that AI systems can adequately deal with errors or inconsistencies during all lifecycle phases;
- Requirements ensuring that AI systems are resilient against overt attacks and against more subtle attempts to manipulate data or algorithms, and that mitigating measures are taken in such cases.

Compliance with this requirement entails technical and organizational measures tailored to the intended use of the AI system, to be assessed since the design phase of an AI system and throughout the moment in which the system is released on the market. It includes measures to prevent and mitigate automation bias, particularly for AI systems used to provide assistance to humans; and measures to detect and safely interrupt anomalies, dysfunctions, unexpected behaviour.

For every single AI product the following activities are envisaged:

1 | On accuracy:

- familiarising oneself with accuracy requirements;
- calculating an established accuracy metric for the task at hand;
- writing an explanation of the accuracy metric, understandable for lay people;
- procure external test datasets and calculating additional required metrics.

2 | On robustness:

- familiarising oneself with robustness requirement;
- brainstorming on possible internal limitations and external threats of the AI model;
- describing limitations of the AI system based on knowledge of the training data and algorithm;
- conducting internal tests against adversarial examples (entails possible retraining, changes to the algorithm, 'robust learning');
- conducting internal tests against model flaws (entails possible retraining, changes to the algorithm);
- conducting tests with external experts (e.g. workshops, audits);
- conducting robustness, safety tests in real-world conditions (controlled studies, etc.).

Moreover, additional labour is very likely be necessary to perform these tasks so that the development complies with requirements and to keep records of testing results for future conformity assessment.

For an average process, and a normally efficient firm, a reasonable cost estate for this activity is €10 733.33.[24]

The business-as-usual factor

All of the above costs estimates relate to the total cost of the activities. However, economic operators would already take a certain number of measures even without explicit public intervention. To calculate this so-called business-as-usual factor, it is assumed that in the best prepared sector at most 50% of compliance costs would be reduced through existing practices. All sectors of the economy are then benchmarked with regard to their digital intensity against the best performing sector (e.g. in a sector with half the digital intensity only half as much can be accounted for business-as-usual). Next, for each sector future growth in digital intensity is forecast by extrapolating from recent years and a weighted average is calculated. As a result, the above costs are discounted by a factor of 36.67%[25].

Instances where the data used in the impact assessment diverges from the data in the study

All the cost estimates are based on the support study. However, a few adjustments were made.

Firstly, all the figures have been rounded and where possible expressed as ranges of values. That is because the precise figures given above are the result of the mathematical modelling used in the study. However, given the assumption necessary for the calculation, the result really are only rough estimates, and indicating amounts to a single euro would signal a precision which is not backed up by the methodology. So, for example, the study's business-as-usual factor 36.37% is used in the impact assessment as a "roughly one third" reduction.

Secondly, the compliance costs regarding robustness and accuracy have not been taken into account. Indeed, an economic operator trying to sell AI systems would anyway have to ensure that their product actually works, i.e. robustness and accuracy. This cost would therefore only arise for companies not following standard business procedures. While it is important that these requirements are included in the regulatory framework so that substandard operators need to improve their procedures, it would be misleading to include these costs for an average company. Including these costs in the overall estimate would only makes sense if one takes into account that a large share of AI providers supplies products that are either not accurate or not robust. There is no evidence to suggest that this is the case.

Note also that the compliance costs regarding human oversight have not been added with the other compliance costs into one single amount but kept separate, since it is overwhelmingly a recurring cost for AI users rather than a one-off cost for AI suppliers like the other compliance costs.

Finally, companies supplying high-risk AI systems in general already have a quality management system in place. For products, that is fundamentally because of already existing Union harmonisation legislation on product safety, which includes quality system-based conformity assessment procedures and, in some cases, also ad-hoc obligations for economic operators related to the establishment of a quality management system. Companies supplying high-risk stand-alone AI systems, such as remote biometric identification systems in publicly accessible places, which are controversially discussed topics, will equally often either already have a quality management system in

[24] See support study chapter 4, section 4.2.5.
[25] See support study chapter 4, section 4.4.2.1

place or introduce one if they want to market such a system subject to reinforced public scrutiny.,. Analogue to the reasoning above, while it is important that these requirements are included in the regulatory framework so that substandard operators need to improve their procedures, it would be misleading to include these costs for an average company.

5. ANNEX 5

5.1. ETHICAL AND ACCOUNTABILITY FRAMEWORKS ON AI INTRODUCED IN THIRD COUNTRIES

The present initiative on AI appears to be a frontrunner when it comes to proposing a comprehensive regulatory framework for AI. Governments in third countries are looking at the EU as a standard-setter (e.g. India; Japan); less eager to take action to impose regulatory constraints on AI (e.g. China); or more inclined towards sectoral approaches, rather that all-encompassing frameworks (the US). To date, no country has enacted a comprehensive regulatory framework on AI. However, a number of initiatives around the globe were taken into account in the analysis:

- ✓ The Australian government is developing a voluntary AI Ethics framework, which includes a very broad definition of AI and eight voluntary AI Ethics principles. Guidance is developed to help businesses apply the principles in their organisations.
- ✓ In Canada, a Directive on Automated Decision-Making came into effect on April 1, 2020 and it applies to the use by public authorities of automated decision systems that "provide external services and recommendations about a particular client, or whether an application should be approved or denied." It includes an Algorithmic Impact Assessment and obligations to inform affected people when such systems are used.
- ✓ In March 2019, the Japanese Cabinet Office released a document titled "Social Principles of Human-Centric AI". This document defines three basic principles: (i) Dignity, Diversity and Inclusion and Sustainability. In July 2020, the Ministry of Economy, Trade and Industry, published a White Paper with respect to big data, the Internet of Things, AI and other digital technologies. The argument is that in order for regulations to keep up with the changes in technology and foster innovation, a new regulatory paradigm is needed.
- ✓ In early 2020, the Personal Data Protection Commission of Singapore revised after consultation a Model AI Governance Framework, which offers detailed and readily-implementable guidance to private sector organisations to address key ethical and governance issues when deploying AI solutions.
- ✓ In 2019, in the UK, the Office for AI published a "Guide on using artificial intelligence in the public sector" advising how the public sector can best implement AI ethically, fairly and safely. The Information Commissioner's Office (ICO) has also published a Guidance on AI Auditing Framework,[26] providing 'best practices' during the development and deployment of AI systems for ensuring compliance with data protection laws.
- ✓ 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 AI applications in the public sector. Federal agencies must consider 10 principles including promoting public trust in AI, considering issues of fairness, non-discrimination, safety, and security, and assessing risks, costs, and benefits. The most recent U.S. President's Executive

[26] https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-artificial-intelligence-and-data-protection/

Order from 3 December 2020 on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, 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.

- ✓ Senate and House bills for the Algorithmic Accountability Act were proposed in the US-Congress in April 2019: they have required "impact assessments" on "high-risk" automated decision systems." Similar bills were recently introduced by New Jersey, Washington State and New York City.
- ✓ In February 2020, the New York City Council also proposed a bill for the use of automated employment decision tools, which requires an independent bias audit of these systems and informing job candidates that such systems have been used and that they are regulated by this act.
- ✓ Still in the Unites States, a Commercial Facial Recognition Privacy Act was proposed in March 2019. If enacted, the bill would generally prohibit organisations from using "facial recognition technology to collect facial recognition data" of end-users without providing notice and obtaining their consent.
- ✓ The Government of New Zealand, together with the World Economic Forum, in 2020 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.

5.2. FIVE SPECIFIC CHARACTERISTICS OF AI

(1) Complexity: [multiplicity of elements that constitute an AI system and complexity of a value chain]

AI systems often have many different components and process very large amounts of data. For example, advanced AI models frequently have more than a billion parameters. These amount of parameters are not in practice understandable for humans, including for their designers and developers.

A system can be complex but still comprehensible from an ex-post perspective. For example, in the case of a rule-based system with a high number of rules, a human might not be able to say in advance what output the system would produce in a given context, but once there is an output, it can be explained based on the rules.

(2) Transparency/ Opacity: [the process by which an AI system reaches a result]

The opacity refers to the lack of transparency on the process by which AI system reaches a result. An AI system can be transparent (or conversely opaque) in three different ways: with respect to how exactly the AI system functions as a whole (*functional transparency*); how the algorithm was realized in code (*structural transparency*) and how the program actually run in a particular case, including the hardware and input data (*run transparency*).

Algorithms often no longer take the form of more or less easily readable code, but instead resemble a 'black-box'. This means that while it maybe be possible to test the algorithm as to its effects, but not to understand how those effects have been achieved.

Some AI systems lack transparency because the rules followed, which lead from input to output, are not fully prescribed by a human. Rather, is some cases, the algorithm is set to learn from data in order to arrive at a pre-defined output in the most efficient way, which might not be representable by rules which a human could understand. As a result, AI systems are often **opaque** in a way other digital systems are not ('the so called black box effect'). Independently from technical characteristics, a **lack of transparency** can also stem from systems relying on rules and functionalities that are not publicly accessible and of which a meaningful and accurate description is not publicly accessible.

The complexity and lack of transparency (opacity of AI) makes it difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights.

(3) Continuous adaptation: [the process by which an AI system can improve its own performance by 'learning' from experience] **and Unpredictability:** [the outcome of an AI system cannot be fully determined]

Some AI systems are not completed once put into circulation, but by their nature depend upon subsequent input, in particular on updates or upgrades. Often they need to interact with other systems or data sources in order to function properly. They therefore need to remain open by design, i.e. permit external input either via some hardware plug or through some wireless connection, and come as hybrid combinations of hardware, software, continuous software updates, and various continuous services.

"Many systems are designed to not only respond to pre-defined stimuli, but to identify and classify new ones and link them to a self-chosen corresponding reaction that has not been pre-programmed as such".[27] Some AI systems can be used to automatically adapt or 'learn' while in use. In these cases, the rules being followed by the system will adapt based on the input which the system receives. This continuous adaptation will mean that the same input may produce different outputs at different times, thus rendering the system unpredictable to a certain extent.

Continuous adaptation can give rise to new risks that were not present when the system was placed on the market. These risks are not adequately **addressed** ***in the existing legislation which predominantly focuses on safety risks present at the time of placing on the market.***

(4) Autonomous behaviour: [functional ability of a system to perform a task with minimum or no direct human control or supervision]

AI systems can increasingly perform tasks with less, or entirely without, direct human intervention.[28] A certain and increasing degree of autonomy (level of autonomy is a continuum) is one of the key aspects of certain AI systems.[29] This continuum ranges from systems where actions of a system are under full supervisions and control of a human to the more sophisticated AI systems that "combine environmental feedback with the system's own analysis regarding its current situation" and thus have minimum or no human supervision in real time. This increasing degree of autonomous behaviour of some AI systems for a particular task

[27] Report from the Expert Group on Liability and New Technologies – New Technologies Formation, European Commission, 2019, p.33.

[28] This is independent and separate from the ability of certain systems to alter the rules which they follow while in use, i.e. 'continuous adaptation' characteristic discussed above.

[29] See for example, SAE International standard J3016 "Levels of Driving Automation" that defines the six levels of driving automation for road vehicles, from no automation to full automation.

combined with their increasing 'ability' to 'interact' with the external environment may present a particular challenge for transparency and human oversight.[30]

Autonomy is not by itself a technological feature but rather the result of a design decision allowing a more or less constrained interaction between the system and the environment in pursuit of a task. The high-level objectives are defined by humans, however the underlying outputs and mechanisms to reach these objectives are not always concretely specified. The partially autonomous behaviour that developers foresee for certain AI systems is usually strongly tied to a specific context and function. Within the given context, these AI systems are designed to help reach conclusions or take decisions within pre-set boundaries without the involvement of a human operator.

Autonomy can affect the safety of the product, because certain AI systems increasingly can perform tasks with less, or entirely without, direct human intervention and in complex environments this may lead to situations where AI system may actions which have not been fully foreseen by their human designers with limited possibilities to override the AI system decision.

(5) Data

Many AI systems "increasingly depend on external information that is not pre-installed, but generated either by built-in sensors or communicated from the outside, either by regular data sources or by ad hoc suppliers. Data necessary for their proper functioning may, however, be flawed or missing altogether, be it due to communication errors or problems of the external data source, due to flaws of the internal sensors or the built-in algorithms designed to analyse, verify and process such data".[31]

The accuracy of AI systems might be unevenly distributed in relation to different kinds of input data, depending on the data with which the system was trained. Furthermore, algorithms that are based on statistical methods produce probabilistic outputs, always containing a certain degree of error, no matter if they are designed to adapt while in use or not. Certain AI systems, due to the way and context they are exploited, present a risk of algorithmic bias as a consequence of several factors such as the considered dataset or machine learning algorithm. For instance, a machine learning algorithm may be "trained" on data to build a model that, once deployed, will process new data in a certain way (e.g. performing classification or pattern recognition). As an example, an application that is developed to recognize patterns would in the case of one common method (supervised learning) "learn" from the training data which characteristics (often called "features") are relevant indicators for certain patterns so that the application can be used to recognize such patterns. The trained application can then be used to analyse future input data.

Both the training data and the input data (used to obtain an output) risk being discriminatory if they are unsuitable or inaccurate. For example, in recruitment contexts it is plausible that a developer only has data about accepted candidates, but no data about the would-be performance of candidates that were not hired. Besides the data, potential discrimination can also originate in the design of algorithms that

[30] For more detailed discussion of concept of autonomy see e.g. The International Committee of the Red Cross, Autonomy, artificial intelligence and robotics: Technical aspects of human control, 2019. This report cautiously explain "the perception of both autonomy and AI is constantly shifting, as advances in technology mean that some systems once considered "autonomous" and "intelligent" are now classed merely as "automated". Importantly, there is no clear technical distinction between automated and autonomous systems, nor is there universal agreement on the meaning of these terms."

[31] See above, p.33.

are used to process the data. Relevant factors include the problem formulation, the underlying conception of a good result, potential biases in the conception of the software code, such as in the choice of input data and variables or in the benchmark for the evaluation of the outcome, which is often used to further optimise an application. There is a particular risk of biased outcomes in the case of machine learning applications. By automating at least parts of the process by which the rules are generated according to which an algorithm will produce results, it becomes possible that discriminatory rules are automatically generated. This is even likely where the data used to train a machine learning application reflects societal biases, if there is no adequate procedure to counteract these biases.

The dependence of AI systems on data and their 'ability' to infer correlations from data input can in certain situations can affect the values on which the EU is founded, create real health risk, disproportionately adverse or discriminatory results, reinforce systemic biases and possibly even create new ones.

5.3. INTERACTION BETWEEN THE INITIATIVE ON AI AND EXISTING SECTORAL PRODUCT SAFETY LEGISLATION

Section 1.

Existing product safety legislation does not contain specific requirement for safety and trustworthiness of AI systems. The proposed horizontal framework on AI will establish such new requirements for high-risk AI systems for certain sectoral product safety legislation (new and old approach). The acts concerned under the NLF framework and the old approach are enumerated respectively in sections A and B below.

The table below summarises how these new requirements for high-risk AI systems will be implemented and interact with existing sectoral product safety legislation.

Table 6: Overview of impact and applicability of the existing safety legislation and the AI horizontal framework to high-risk AI systems

	High-risk AI / existing safety legislation	Interaction	Overall Impact
1	AI systems covered by certain sectoral **safety legislation, following New Legislative Framework (NLF)**	The AI system will be high-risk if it is a safety component of a product or a device that is subject to a third party conformity assessment under the NLF legislation. Requirements and obligations for high-risk AI systems set by the AI horizontal framework will become directly applicable and will automatically complement the existing NLF legislation.	• The new ex ante requirements for high-risk AI systems set in the AI horizontal framework will complement the existing sectoral safety requirements under NLF sectoral legislation. • The conformity assessment procedures already existing under NLF would also apply for the checks of the new AI specific requirements. • New obligations for providers and users will apply to the extent these are not already existing under the NLF sectoral act. • The ex-post enforcement of the new rules for AI systems will be carried out by the same NLF market surveillance authorities responsible for the product.
2	AI systems covered by certain sectoral **safety legislation, following Old Approach** (e.g. aviation, cars)	AI systems that are safety components of products under relevant old approach legislation will always be considered high-risk. The new requirements for high-risk AI systems set by the AI horizontal framework will have to be taken into account when adopting relevant	• The new ex-ante requirements for high-risk AI systems set in the AI horizontal framework will complement the existing sectoral requirements under old approach (when relevant implementing or delegated legislation under those acts will be adopted). • The conformity assessment or authorisation procedures existing under the sectoral old approach legislation would also apply for the

	implementing or delegated legislation under those acts.	checks of the new AI requirements. • The AI horizontal framework will not create any new obligations for providers and users. • The ex-post enforcement rules of the AI horizontal framework will not apply.

A. **Interaction between the proposal for AI horizontal framework and NLF safety legislation (row 1 in table above)**

The proposed horizontal framework on AI will establish new requirements for high-risk AI systems that will **complement** the existing product safety NLF legislation.[32] An AI system will be high-risk if it is a safety component of a product or a device which undergoes a third party conformity assessment under the relevant sectoral NLF legislation.[33] A safety component of a product or device is understood as a component which provides the safety functions with regard to that specific product or device.

Based on up-to-date analysis **the concerned NLF legislation** that will fall under the scope of the new AI horizontal initiative include:

- 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;
- Regulation (EU) 746/2017 on in-vitro diagnostic medical devices.

The objective is to ensure that **the new AI horizontal framework** (which is in itself an NLF-type framework for the new safety requirements it creates) can be **fully and smoothly integrated into the existing procedures and enforcement and governance systems established under the NLF legislation**.

[32] NLF product legislation also covers some non-embedded AI systems which are considered products by themselves (e.g. devices by themselves under the Medical Device Regulations or AI safety components placed independently on the market which are "machinery by themselves under the Machinery Directive). If those non-embedded AI systems are subject to third-party conformity assessment under the relevant sectoral framework, they will be high-risk for the purpose of the AI horizontal framework.

[33] This approach is justified because the conformity assessment of any sectoral legislation already presupposes a risk assessment on the safety risks posed by the products covered by that instrument. It makes therefore sense to rely on the risk classification of a product under the relevant NLF legislation to define when an AI-driven safety component (of that product) should be considered high-risk.

The new requirements for AI systems set by the AI horizontal framework would become directly applicable and be checked in the context of the **conformity assessment system already existing under the relevant NLF instrument**.

The Notified Bodies assessing the compliance of the provider with the new AI requirements **would be the ones already designated under the relevant NLF legislation.** However, the competence of the Notified Bodies in the field of AI should be assessed as part of the designation process under the relevant NLF instrument.

Obligations for certain operators in the value chain – namely manufacturers, importer, distributor, authorised representative - are generally already established in the existing NLF legislation. Obligations for economic operators (notably for providers and users) of the new AI horizontal framework apply to the extent these are not already existing under the NLF sectoral act.

With regard to **market surveillance**, Regulation (EU) 2019/1020 on market surveillance will apply to the AI horizontal framework. The ex-post enforcement of the new rules for high-risk AI systems will be carried out by the same NLF market surveillance authorities responsible for the product under the existing NLF legislation.

Ongoing or future **reviews of NLF product legislation** will not address aspects which are covered by the AI horizontal instrument. In order to increase legal clarity, any relevant NLF product legislation being reviewed (e.g. Machinery Directive 2006/42/EC subject to an ongoing review) would cross reference the AI horizontal framework, as appropriate. However, any reviewed NLF product legislation may aim to ensure that the incorporation of the AI system into the product does not compromise the safety of the product as a whole. In this respect, for example, the reviewed Machinery Directive 2006/42/EC could contain requirements for the safe integration of AI systems into the product (not covered by the AI horizontal framework).

B. Interaction between the proposal for AI horizontal framework and old-approach safety legislation (row 2 in table above)

Compared to NLF legislation, the applicability of the AI horizontal framework will be different for the old approach product safety legislation. This is because the old approach legislation follows a system of enforcement, generally based on detailed legal safety requirements (with possible integration of international standards into law) and a stronger role of public bodies in the approval system – an approach very different from the NLF logic followed by the AI horizontal initiative.

The horizontal framework on AI will establish new requirements for high-risk AI systems (e.g. transparency, documentation, data quality) that will be integrated into the existing old approach safety legislation. AI systems that are safety components of products under the old approach legislation will always be considered as high-risk AI systems.[34]

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;

[34] This is because products regulated under the old approach legislation always undergo third party conformity assessments or authorisation procedures in the legislations that will be covered by the new AI initiative.

- 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.
- Directive 2014/90/EU on marine equipment (which is a peculiar NLF-type legislation, but given the mandatory character of international standardization in that field, will be treated in the same way as old-approach legislation).

The new requirements for high-risk AI systems set by the AI horizontal framework will have to be taken into account in the future when amending the sectoral legislation or when adopting relevant implementing or delegated acts under that sectoral safety legislation.

Existing conformity assessment/authorization procedures, obligations of economic operators, governance and ex-post enforcement under the old approach legislation will not be affected by the AI horizontal framework.

The application/relevance of the AI horizontal initiative on AI to the old approach safety legislation will be thus limited only to the new safety requirements for high-risk AI systems, when relevant implementing or delegated acts under that sectoral safety legislation will be adopted.

5.4. LIST OF HIGH RISK AI SYSTEMS (NOT COVERED BY SECTORIAL PRODUCT LEGISLATION)

For AI systems that are mainly with fundamental rights implications and not covered by sectoral product safety legislation,[35] the Commission has done the initial assessment for identifying the relevant high-risk AI systems by screening a large pool of AI use cases, covering:

- High-risk AI use cases included in **the EP report[36];**
- **A list of 132 AI use cases identified by a recent ISO report**[37] and other methodologies[38];
- **Results from the study** accompanying the report, analysis by **AI Watch** and **extensive complementary research** of other sources such as analysis of case-law, academic literature and reports from international and other organisations (problem definition 2 in the impact assessment presents in short some of the most prominent use cases with significant fundamental rights implications);

---

[35] For AI systems which are safety components of products covered by sectoral product safety legislation see Annex 5.3.

[36] 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, 2020/2012(INL).

[37] For example, classification of products as high-risk means that the AI safety component should also be treated similarly; See also Article 29 Data Protection working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.

[38] Final Draft of ISO/IEC TR 24030 - AI Use Cases. The Opinion of the German Data Ethic Commission proposing a pyramid of 5 levels of criticality of the AI systems. The Council of Europe Recommendation CM/Rec(2020)1 refers to "high risk" when the algorithmic systems is used in processes or decisions that can produce serious consequences for individuals or in situations where the lack of alternatives prompts a particularly high probability of infringement of human rights, including by introducing or amplifying distributive injustice.

- Results from the piloting **of the draft HLEG ethic guidelines** in which more than 350 stakeholders participated, including 50 in-depth case studies;
- Results from the **public consultation on the White Paper** that identify specific use cases as high-risk (or request their prohibition) and additional targeted consultations with stakeholders[39];

The risk assessment methodology described in the impact assessment has been applied to this large pool of use cases and the assessment of the Commission has concluded that the **initial list of high-risk AI systems presented below should be annexed** to Commission's proposal of the AI horizontal instrument. Other reviewed AI use cases not included in this list have been discarded either because they do not cause harms to the health and safety and/or the fundamental rights and freedom of persons, or the probability and/or the severity of these harms has not been estimated as 'high' by applying the indicative criteria for risk assessment.[40]

Table 7: List of high-risk AI use cases (stand-alone) identified following application of the risk assessment methodology

HIGH-RISK USES	POTENTIAL HARMS	ESPECIALLY RELEVANT INDICATIVE CRITERIA*	EVIDENCE & OTHER SOURCES
AI systems intended to be used for the remote biometric identification of persons in publicly accessible spaces	Intense interference with a broad range of fundamental rights (e.g. private life and data protection, human dignity, freedoms expression, freedom of assembly and association) Systemic adverse impact on society	Already used by an increasing number of public and private actors in the EU Potentially very severe extent of multitude of harms High potential to scale and adversely impact a plurality of people Vulnerability of affected people (e.g. people cannot object freely, imbalance if	AlgorithmWatch and Bertelsmann Stiftung, *Automating Society Report 2020*, 2020 (pp. 38-39, p. 104); European Data Protection Board, *Facial recognition in school renders Sweden's first GDPR fine*, 2019; European Data Protection Board, *EDPS Opinion on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence*

[39] The Commission has also carried out targeted consultations on specific topics that have informed its assessment: 1) Principles and Requirements for trustworthy AI; 2) Biometrics, 3) Children's rights, 4) Standardisation, 5) Conformity assessments, 6) Costs of implementation. These workshops were focusing on collecting data, evidence and complementing the public consultations on the White Paper and the Inception Impact Assessment.

[40] See Option 3 in section 5.3 of the Impact assessment. A specific assessment of the probability and severity of the harms will be done to determine if the AI system generates a high-risk to the health and safety and the fundamental rights and freedom of persons based on a set of criteria that will be defined in the legal proposal. The criteria for assessment 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.

	at large (i.e., on democratic processes, freedom and chilling effect on civic discourse)	used by public authorities) Indication of harm (legal challenges and decisions by courts and DPAs)	*and trust*, 2020 (pp. 20-21); Agency for Fundamental Rights, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, 2019; Court of Appeal, United Kingdom, Decision R (Bridges) v. CC South Wales, EWCA Civ 1058 of 11 August 2020; Buolamwini, I./ Gebru, T., *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 2018; National Institute of Standards and Technology, U.S. Department of Commerce, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, 2019.
AI systems intended to be used to dispatch or establish priority in the dispatching of emergency first response services, including firefighters and medical aid	Injury or death of person(s), damage of property (i.e. by de-prioritising individuals in need of emergency first response services) Potential interference with fundamental rights (e.g. human dignity, right to life, physical and mental integrity, non-discrimination)	Already used by some public authorities (firefighters, medical aid) Potentially very severe extent of harm High potential to scale and adversely impact a plurality of people (due to public monopoly) Vulnerability and high dependency on such services in emergency situations Irreversibility of harm very likely (due to physical character of the harm) Not regulated by safety legislation	European Agency for Fundamental Rights, *Getting The Future Right – Artificial Intelligence and Fundamental Rights*, 2020 (pp. 34-36); ISO A.97 System for real-time earthquake simulation with data assimilation, ISO/IEC TR 24030 - AI Use Cases 2020 (p. 101).
AI systems intended to be used as safety components in the management and operation of essential public infrastructure networks, such as roads or the supply of water, gas and electricity	Injury or death of person(s) Potential adverse impact on the environment Disruptions of ordinary conduct of critical economic and social activities	Potentially very severe extent of harm to people, environment and ordinary conduct of life High potential to scale and adversely impact people and also the environment (potentially large scale due to criticality of essential public infrastructure networks) Dependency on outcome (high degree of dependency due to potential to impact	German Data Ethics Commission, *Opinion of the Data Ethics Commission*, 2020. ISO A.109 AI dispatcher (operator) of large-scale distributed energy system infrastructure, ISO/IEC TR 24030 - AI Use Cases 2020 (p. 42); ISO A.29 Enhancing traffic management efficiency and infraction detection accuracy with AI technologies, ISO/IEC TR 24030 - AI Use Cases 2020 (pp. 103-104); ISO A.49 AI solution for traffic signal optimization based on multi-source data fusion, ISO/IEC TR

		sensitive access to basic utilities) Irreversibility of harm very likely due to the safety implications Not regulated by safety legislation	24030 - AI Use Cases 2020 (p. 104); ISO A.122 Open spatial dataset for developing AI algorithms based on remote sensing (satellite, drone, aerial imagery) data, ISO/IEC TR 24030 - AI Use Cases 2020 (pp. 114-115).
AI systems intended to be used for determining access or assigning individuals to educational and vocational training institutions, as well as for assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions	Intense interference with a broad range of fundamental rights (e.g. non-discrimination, right to education, private life and data protection, effective remedy, rights of children) Adverse impact on financial, educational or professional opportunities; adverse impact on access to public services;	Already used by some educational institutions Potentially very severe extent of harm High potential to scale and adversely impact a plurality of people (public education) Dependency on outcome (access to education critical for professional and economic opportunities) Insufficient remedies and protection under existing law Indication of harm (opacity, existing legal challenges/case-law)	Tuomi, I., *The use of Artificial Intelligence (AI) in education*, *European Parliament,* 2020 (pp. 9-10); UNESCO, *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*, 2019 (pp. 32-34); AlgorithmWatch and Bertelsmann Stiftung, *Automating Society Report 2020*, 2020 (p. 280); Burke, L., *The Death and Life of an Admissions Algorithm*, *Inside Higher Ed*, 2020; Department of Education Ireland, *Press release on errors detected in Leaving Certificate 2020 Calculated Grades Process*, 30 September 2020 ISO A.73 AI ideally matches children to daycare centers, ISO/IER TR 24030 – AI Use Cases 2020 ISO A.83 IFLYTEK intelligent marking system, ISO/IER TR 24030 – AI Use Cases 2020 (p.39).
AI systems intended to be used for recruitment – for instance in advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests – as well as for making decisions on promotion and termination of work-related contractual relationships, for task allocation, monitoring or evaluating work performance and behaviour	Intense interference with a broad range of fundamental rights (e.g. workers' rights, non-discrimination, private life and personal data, effective remedy) Adverse impact on financial, educational or professional opportunities	Growing use in the EU Potentially very severe effect of adverse decisions in employment context on individuals' professional and financial opportunities and their fundamental rights High degree of vulnerability of workers vis-à-vis (potential) employers Insufficient remedies and protection under existing law Indication of harm (high probability of historical biases in recruitment used as training data, opacity, case-law for unlawful use);	Datta, A. et al., *Automated Experiments on Ad Privacy Settings, Proceedings on Privacy Enhancing Technologies*; 2015 (pp. 92-112); Electronic Privacy Information Center, *In re HireVue*, 2019; Geiger, G., *Court Rules Deliveroo Used 'Discriminatory' Algorithm*, *Vice*, 2020; [Italy, Tribunale di Bologna, Decision of 31 December 2020, to be published.]; Sánchez-Monedero, J. 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; Upturn, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, 2018;

			ISO A.23 VTrain recommendation engine, ISO/IEC TR 24030 - AI Use Cases 2020 (p. 38)
AI systems intended to be used to evaluate the creditworthiness of persons or establish their credit score, with the exception of AI systems developed by small scale users for their own use	Adverse impact on economic, educational or professional opportunities; Adverse impact on access to essential public services Intense interference with a broad range of fundamental rights (e.g. non-discrimination, private life and personal data, effective remedy)	Growing use by credit bureaux and in the financial sector Lack of transparency of AI based decisions making it impossible for individuals to know what type of behaviour will be relevant to assign them to their statistical group Risk of high number of cases of indirect discrimination which are not likely to be captured by existing anti-discrimination legislation Potentially severe harm (due to reduced access to economic opportunities when the services are provided by large scale operators, e.g. credit enabling investments and use of the score to determine access to other essential services e.g. housing, mobile services etc.) Insufficient remedies and protection under existing law (robust financial service legislation, but assessment also done by unregulated entities; no binding specific requirements for AI) Indication of harm (high probability of historical biases in past credit data used as training data, opacity, case law)	AlgorithmWatch, 2020, *SCHUFA, a black box: OpenSCHUFA results published*, 2018; European Agency for Fundamental Rights, *Getting The Future Right – Artificial Intelligence and Fundamental Rights*, 2020 (pp. 71-72); Finland, National Non-Discrimination and Equality Tribunal, Decision 216/2017 of 21 March 2017; European Banking Authority, *Report on Big Data and Advanced Analytics*, 2020 (pp. 20-21); ISO A.27 Credit scoring using KYC data, ISO/IEC TR 24030 - AI Use Cases 2020 (pp. 43-44); ISO A.119 Loan in 7 minutes, ISO/IEC TR 24030 - AI Use Cases 2020 (p. 46).
AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility for social security benefits and services, as well as to grant, revoke, or reclaim social security benefits and services	Intense interference with a broad range of fundamental rights (e.g. right to social security and assistance, non-discrimination, private life and personal data protection, good administration, effective remedy)	Growing use in the EU Potentially very severe extent of harm (due to potentially crucial importance essential of social security benefits and services for individuals well-being) High potential to scale and adversely impact a plurality of persons or groups (due to the public character of the social security benefits and	Allhutter, D. et al., *AMS Algorithm on trial, Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften*, 2020; Netherlands, Court of The Hague, Decision C-09-550982 / HA ZA 18-388 of 5 February 2020 on Syri; Kayser-Bril, N., *In a quest to optimize welfare management, Denmark built a surveillance behemoth*, AlgorithmWatch, 2020; Niklas, J., *Poland: Government to scrap controversial unemployment scoring system*, *AlgorithmWatch*,

	Adverse impact on financial, educational or professional opportunities or on a person's course of life; adverse impact on access to public services;	services) High degree of dependency on the outcome (due to lack of alternative for recipients) and high degree of vulnerability of recipients vis-à-vis public authorities Indication of harm (opacity, high probability of past biased training data, challenges/case-law)	2019; Wills, T., *Sweden: Rogue algorithm stops welfare payments for up to 70,000 unemployed*, *AlgorithmWatch,* 2019; European Agency for Fundamental Rights, *Getting The Future Right – Artificial Intelligence and Fundamental Rights*, 2020 (pp. 30-34).
Predictive policing and certain other AI systems in law enforcement, asylum, migration, border control with significant impacts on fundamental rights	Intense interference with a broad range of fundamental rights (e.g. effective remedy and fair trial, non-discrimination, right to defence, presumption of innocence, right to liberty and security, private life and personal data, freedom of expression and assembly, human dignity, rights of vulnerable groups) Systemic risks to rule of law, freedom and democracy	Growing use in the EU Potentially very severe extent of harm (due to severe consequences of decisions and actions in this context) Potential to scale at large and adversely impact a plurality of people (due to large number of individuals affected) High degree of dependency (due inability to opt out) and high degree of vulnerability vis-à-vis law enforcement) Limited degree of reversibility of harm Insufficient remedies and protection under existing law Indication of harm (high probability of historical biases in criminal data used as training data, opacity)	AlgorithmWatch, *Automating Society*, 2019 (pp. 37-38, 100); Council of Europe, *Algorithms and human rights*, 2017, (pp. 10-11, 27-28); European Agency for Fundamental Rights, *Getting The Future Right – Artificial Intelligence and Fundamental Rights*, 2020 (pp. 68-74); González Fuster, G., *Artificial Intelligence and Law Enforcement – Impact on Fundamental Rights*, *European Parliament*, 2020; Gstrein, O. J. et al., *Ethical, Legal and Social Challenges of Predictive Policing*, *Católica Law Review,* 3:3, 2019 (pp. 80-81); Oosterloo, S. & van Schie, G., *The Politics and Biases of the "Crime Anticipation System" of the Dutch Police*, *Information, Algorithms, and Systems*, 2103 (pp. 30-41); 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. European Crime Prevention Network, Predictive policing, Recommendations paper, 2014. Wright, R., *Home Office told thousands of foreign students to leave UK in error*, *Financial Times,* 2018. Warrel, H., *Home Office drops 'biased' visa algorithm*, *Financial Times,* 2020; Molnar P. and Gill L., *Bots at the Gate: A human rights analysis of automated decision-making in Canada's immigration and refugee system*, *University of Toronto*,

			2018. ISO A.14 Behavioural and sentiment analytics, ISO/IEC TR 24030 - AI Use Cases 2020 (pp. 96-97). Roxanne research project that uses AI to enhance crime investigation capabilities
AI systems used to assist judicial decisions, unless for ancillary tasks	Intense interference with a broad range of fundamental rights (e.g. effective remedy and fair trial, non-discrimination, right to defence, presumption of innocence, right to liberty and security, human dignity as well as all rights granted by Union law that require effective judicial protection) Systemic risk to rule of law and freedom	Increased possibilities for use by judicial authorities in the EU Potentially very severe impact and harm for all rights dependent on effective judicial protection High potential to scale and adversely impact a plurality of persons or groups (due to large number of affected individuals) High degree of dependency (due to inability to opt out) and high degree of vulnerability vis-à-vis judicial authorities) Indication of harm (high probability of historical biases in past data used as training data, opacity)	Council of Europe, *Algorithms and human rights*, 2017, (pp. 11-12); European Commission for the Efficiency of Justice, *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, 2018; *U.S. Wisconsin Supreme Court Denied Writ of Certiorari of 26 June 2017, Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016)*; Decision of the French Conseil Constitutionnel of 12 June 2018, Décision n° 2018-765 DC.Propublica, *Machine Bias: There's software used across the country to predict future criminals, and it's biased against blacks*, 2016

5.5.: ANALYSES OF IMPACTS ON FUNDAMENTAL RIGHTS SPECIFICALLY IMPACTED BY THE INTERVENTION

Impact on the right to human dignity

All options will require that humans should be notified of the fact that they are interacting with a machine, unless this is obvious from the circumstances or they have already been informed.

Options 2 to 4 will also prohibit certain harmful AI-driven manipulative practices interfering with personal autonomy when causing physical or psychological harms to people.

Impacts on the rights to privacy and data protection

All options will further enhance and complement the right to privacy and the right to data protection. Under options 3 to 4, providers and users of AI systems will be obliged to take mitigating measures throughout the whole AI lifecycle, irrespective of whether the AI system processes personal data or not.

All options will also require the creation of data governance standards in the training and development stage. This is expected to stimulate the use of privacy-preserving techniques for the development of data-driven machine learning models (e.g. federated learning, 'small data' learning etc.). New requirements relating to transparency, accuracy, robustness and human oversight would further complement the implementation of the data protection acquis by providing rules that address developers and providers who might not be directly bound by the data protection legislation. These options will harmonize and enhance technical and organisational standards on how high-level principles should be implemented (e.g. security, accuracy, transparency etc.), including in relation to high-risk AI applications.

Under options 2 to 4, AI used for some particularly harmful practices would also be prohibited such as general purpose scoring of citizens and use of AI-enabled technology that might manipulate users through specific techniques that are likely to cause physical or psychological harms.

Options 2 to 4 will also prohibit certain uses of remote biometric identification systems in publicly accessible spaces and subject the permitted uses to higher scrutiny and additional safeguards on top of those currently existing under the data protection legislation.

Impacts on the rights to equality and non-discrimination

All option will aim to address various sources of risks to the right to non-discrimination and require that sources of biases embedded in the design, training and operation of AI systems should be properly addressed and mitigated. All options except option 1 will also envisage limited testing obligation for users, taking into account the residual risk.

High quality data and high quality algorithms are essential for discrimination prevention. All options would impose requirements for documentation requirements in relation to the data and applications used and, where applicable, use of high quality data sets that should be relevant, accurate and representative for the context of application and the intended use. Obligations will also be imposed for testing and auditing for biases and adoption of appropriate bias detection and correction measures for high-risk AI system. Transparency obligations across the full AI value chain about the data used to train an algorithm (where

applicable), its performance indicators and limitations will also help users to minimize the risk of unintentional bias and discrimination.

All options will also include additional requirements for accuracy and human oversight, including measures to minimize 'automation bias' that will help to reduce prohibited discriminatory impacts across protected groups.

Under options 2 to 4, providers and users of AI systems will be allowed to process sensitive data for the sole purpose of bias detection and mitigation and subject to appropriate safeguards. This will strike a fair balance and reconcile the right to privacy with the right to non-discrimination in compliance with the data protection legislation and the EU Charter of Fundamental Rights.

When properly designed AI systems could positively contribute to reducing bias and existing structural discrimination especially in some sectors (e.g. recruitment, police, law enforcement). For example, predictive policing might, in some contexts, lead to more equitable and non-discriminatory policing by reducing reliance on subjective human judgements.

Impact on the right to freedom of expression

Options 2 to 4 are expected to indirectly promote the right to freedom of expression insofar that increased accountability on the use of data shared by individuals could contribute to preventing the risk of a chilling effect on the right to freedom of expression.

An obligation to label deep fakes generated by means of AI could have an impact on the right to freedom of expression. That is why this obligation should not apply when the deep fakes are disseminated for legitimate purposes when authorised by law or to exercise freedom of expression or arts subject to appropriate safeguards for the rights of third parties and the public interests.

Impacts on the right to an effective remedy and fair trial and the right to good administration

The overall increased transparency and traceability of the system in the scope of all options will also enable affected parties to exercise their right to defence and right to an effective remedy in cases where their rights under Union or national law have been breached.

In addition, options 3 to 4 would require that certain AI systems used for judicial decision-making, in the law enforcement sector and in the area of asylum and migration should comply with standards relating to increased transparency, traceability and human oversight which will help to protect the right to fair trial, the right to defence and the presumption of innocence (Articles 47 and 48 of the Charter) as well as the general principles of the right to good administration. In turn, increased uptake of trustworthy AI in these sectors will contribute to improving access to legal information, possibly reducing the duration of judicial proceedings and to enhancing access to justice in general.

Finally, concerning restrictions potentially imposed by authorities, the established remedy options would always be available to providers and users of AI systems who are negatively affected by the decisions of public authorities.

Impacts on rights of special groups

All options are expected to positively affect the rights of a number of special groups. First, **workers' rights** will be enhanced since recruitment tools and tools used for career management or monitoring will likely be subjected to the mandatory requirements for

accuracy, non-discrimination, human oversight, transparency etc. In addition to that, workers (in a broad sense) are often the back-end operators of AI systems, so the new requirements for training and the requirements for safety and security will also support their rights to fair and just working conditions (Article 31 of the Charter).

The **rights of the child** (Art. 24 of the Charter) are expected to be positively affected when high-risk AI systems are affecting them (e.g. for decision-making purposes in different sectors such as social welfare, law enforcement, education etc.). Providers of the high-risk AI system should also consider children's safety by design and take effective measures to minimize potential risks. Under option 3, this will concern only products and services considered to be 'high-risk', while under option 4 any product embedding AI, such as AI-driven toys, will have to comply with these requirements.

Option 2, 3, 3+ and 4 would also prohibit the design and use of AI systems with a view to distorting children's behaviour in a manner that is likely to cause them physical or psychological harm which would also help to increase overall safety and integrity of children who are vulnerable due to their immature age and credulity.

Overall, increased use of AI applications can be very beneficial for the enhanced protection of children's rights, for example, by detecting illegal content online and child sexual abuse, providing that it does not lead to a systematic filtering of communications, identifying missing children, providing adaptive learning systems tailored to each student's needs and progress to name only a few examples.

Impact on the freedom to conduct a business and the freedom of science

All options will impose some restrictions on the freedom to conduct business (Article 16 of the Charter) and the freedom of art and science (Article 13 of the Charter) in order to ensure responsible innovation and use of AI. While under option 1, these restrictions will be negligent since compliance with the measures will be voluntary, options 2 to 4 envisage binding obligations that will make the restrictions more pronounced.

Under options 2, 3 and 3+, these restrictions are proportionate and limited to the minimum necessary to prevent and mitigate serious safety risks and likely infringements of fundamental rights. However, option 4 would impose requirements irrespective of the level of risk, which might lead to disproportionate restrictions to the freedom to conduct a business and the freedom of science. These restrictions are not genuinely needed to meet the policy objective and they would prevent the scientific community, businesses, consumers and the society at large from reaping the benefits of the technology when it poses low risks and does not require such an intense regulatory intervention.

Impact on intellectual property rights (Article 17(2) of the Charter)

Often economic operators seek out copyright, patent and trade secret protection to safeguard their knowledge on AI and prevent disclosure of information about the logic involved in the decision-making process, the data used for training the model etc.

The increased transparency obligations under options 2 to 4 will not disproportionately affect the right to an intellectual property since they will be limited only to the minimum necessary information for users, including the information to be included in the public EU database.

When public authorities and notified bodies are given access to source code and other confidential information, they are placed under binding confidentiality obligations.