AmCham EU
SPEAKING FOR AMERICAN BUSINESS IN EUROPE

Consultation response

Consultation on HLEG draft of AI ethics guidelines

AmCham EU speaks for American companies committed to Europe on trade, investment and competitiveness issues. It aims to ensure a growth-orientated business and investment climate in Europe. AmCham EU facilitates the resolution of transatlantic issues that impact business and plays a role in creating better understanding of EU and US positions on business matters. Aggregate US investment in Europe totalled more than €2 trillion in 2017, directly supports more than 4.7 million jobs in Europe, and generates billions of euros annually in income, trade and research and development.

American Chamber of Commerce to the European Union
Speaking for American business in Europe

Avenue des Arts/Kunstlaan 53, 1000 Brussels, Belgium • **T** +32 2 513 68 92
info@amchameu.eu • amchameu.eu • European Transparency Register: 5265780509-97

Contents

Introduction: Rationale and foresight of the Guidelines 3

Chapter I: Respecting fundamental rights, principles and values – ethical purpose 4

Chapter II: Realising trustworthy AI 7

Chapter III: Assessing trustworthy AI 9

General comments 10

Introduction: Rationale and foresight of the Guidelines

The American Chamber of Commerce to the European Union (AmCham EU) believes that key success factors for ethics in artificial intelligence (AI) guidelines are:

- A human-centric framework that is proportional, risk-based and flexible because AI is not a monolithic technology but its ethical risk changes drastically according to its use and context.
- A 'holistic' framework including high-level principles, best practices, voluntary and industry-driven standards and existing regulation.
- A common understanding of what the problems are and further research to enable these problems to be more effectively addressed (eg, technology can also be part of the solution).
- Encouraging companies to self-regulate: companies should establish guiding ethical principles for themselves that will apply throughout all their operations.
- Encouraging companies to adopt concrete governance practices: AI ethics should be built into business performance, not bolted-on as an afterthought. AI ethics should be part of the AI lifecycle, from the data models and product deployment, to the update of workflows, tools, and business processes. For example, companies could set up internal structures such as 'AI ethics boards' to discuss these issues.
- Encouraging companies to understand the key issues and tools to mitigate risk: as businesses and industry continue to pilot, adopt and rely on AI technologies to reshape the future of decision-making, AI that can be trusted to be transparent, fair, explainable and secure is imperative. Businesses need to continuously listen to concerns that might exist and adapt their ethical guidelines in developing tools to mitigate risks.
- Continuous dialogue with stakeholders (industry, researchers, etc) on the development of appropriate mechanisms, in particular for any consideration of 'regulatory' mechanisms.
- Increase overall awareness and foster trust through the entire value chain, from developers to users, as well as consumers and society at large.
- Encouraging authorities to collaborate with industry and civil society in building data ecosystems which help to generate datasets in quantity and quality which ensure and empower a fair and ethical AI.
- Encouraging policy-makers to cooperate at international level on ethical guidelines, helping to ensure an inclusive and global approach.

We welcome that the Commission's High-Level Expert Group (HLEG) on AI addresses the important topics of ethics and policy through an inclusive and multi-disciplinary approach. Moreover, we strongly agree with the objective of the HLEG to issue guidelines on ethics in AI (hereafter 'the Guidelines') that are actionable and proportionate and encourage companies to adopt a responsible and ethical approach to AI. This will be the real added value of the Guidelines and, whilst the current draft is a very good basis, it needs to be further 'operationalised' through the development of use cases. Finally, we believe that Europe is uniquely placed for the development of AI, and therefore agree with the intention to use these Guidelines to foster a reflection and discussion on the use of the technology at a global level.

We support that 'Trustworthy AI' (p.1-3) has two components: it must have an ethical purpose and it must be technically robust. This will make sure that AI is trusted by its users and that it will result in improving Europe's competitiveness. Whilst regulation already exists that applies to AI – as rightly highlighted in the Guidelines – a common approach on ethical questions, principles and values brings a huge benefit in generating user trust and facilitating a broader uptake of AI. Furthermore, building trust is also a mean to demystify some of the scepticism

around the technology. Trust is essential to create the needed dialogue for educating the public on what AI is and how it can be used. This could be clearly stated in the introductory section (p.1, 'Trustworthy AI').

The Guidelines provide a thoughtful and comprehensive set of ethical considerations designed to help developers and implementers of AI achieve 'trustworthy AI'. In offering these considerations, the Guidelines acknowledge that 'different situations raise different challenges' (p. iii). We strongly endorse this point and believe that contextual considerations merit greater attention in the Guidelines. The degree of risk of individual or societal harm, and the potential severity of such harm, will vary enormously depending on the specific AI application at issue. In fact, many of the ethical issues identified in these Guidelines only arise for AI systems that have a consequential – or meaningful – impact on individuals. We therefore urge the HLEG to make clear at the outset of the Guidelines that their recommendations are not 'one-size-fits-all', and instead should be tailored to each specific implementation of AI depending on a careful and thorough risk assessment.

We appreciate the target of putting a method in place to enable all stakeholders to formally endorse and sign up to the Guidelines on a voluntary basis, as this will help to support transparency for users and build trust. Considering the importance of the contextual elements – outlined above – only guidelines that are holistic, proportional, risk-based and flexible based on high-level principles, best practices, voluntary and industry-driven standards and existing regulation, seem appropriate for commitment to the proposal of the AI HLEG. Currently, the technical (and non-technical) methods (as mentioned in Chapter 2) can be outpaced too quickly by technology, and the detailed checklists (as described in Chapter 3) are both too specific and not relevant for all use cases. Thus, we would propose to not include these into the sections subject to formal endorsements.

Finally, we share the view that the issue of AI ethics requires a regular discussion and iteration. Therefore, it would be helpful to clarify in the final version of the Guidelines the process that would be followed for this continuous dialogue, as well as regularly updating the document.

Chapter I: Respecting fundamental rights, principles and values – ethical purpose

We value highly the approach of the HLEG to derive the responsible and trustworthy development, deployment and use of AI from the fundamental rights, ethical principles and values that underpin the commitment of the European Union (EU). The purpose of AI should be to bring benefit to individuals, society and business, and we believe that a human-centric approach to its growth is the prerequisite for its lasting success.

Taking into account that AI should create added value to different stakeholders, economic interests have to be considered as legitimate interests for a company in order to promote economic growth. Adhering to principles such as traceability, transparency and self-determination might come to an economic cost. Therefore, as an overall comment we recommend complementing the principle of 'beneficence' with a notion of proportionality in Chapter 1.

Fundamental rights of human beings (p. 7):

We appreciate that the commitment to preserve human rights and fundamental values is integrated into the AI ethics.

Under the section 'human dignity', we would suggest to remove the opposition between individual and data subject. To be treated as data subject does not mean that human dignity is negatively impacted.

Under the section 'Respect for democracy, justice and the rule of law', 'human-centric appeal, review and/or scrutiny of decisions made by AI systems' is not a right in itself but rather an '**opportunity**' which exists as a consequence of the rights explained in this section.

Ethical principles in the context of AI and correlating values (p.8):

We generally support the five ethical principles and correlated values proposed by the HLEG (do good, do no harm, preserve human agency, be fair and operate transparently), to ensure that AI is developed in a human-centric manner. Whilst we agree with the principles in general, we must stress that in some cases the principles

may be in conflict with one another and developers might be faced with contradictions. Therefore, we suggest creating a form of hierarchy within them and establish a resolution mechanism in cases of contradiction.

The role of experts in this process is welcome. However, we would like to stress that there will need to be cross-disciplinary or legal experts depending on the issues.

In addition, we encourage the HLEG to more explicitly recognise the fact that there will necessarily need to be a balancing between benefits and harms when deploying AI, and that some trade-offs may be unavoidable. Advancing the interests of certain individuals may inevitably impose harms on others (eg, an AI tool that makes one company more efficient might 'harm' rivals by making them relatively less able to compete).

Moreover, we would like to add the following comments below:

- The principle of beneficence: 'Do Good'

We agree with the AI HLEG that AI should be applied only when an added value can be generated for people and emphasise that this added value can also be of economic nature, such as an increase of efficiency, accuracy, reliability or reproducibility. We also recommend that the Guidelines adopt a broad understanding of beneficence. In fact, AI can be a tool to improve wellbeing, preserve dignity and foster sustainability but it can also serve more neutral objectives whose direct individual or social benefits are less clear. We therefore acknowledge that AI solutions may satisfy beneficence as long as they serve a useful purpose (to someone) that outweighs the risk and severity of potential harm to others.

We fully agree that AI can help with societal issues, such as fairness and inclusion. We would welcome initiatives from the European Commission to foster the discussion and research on increasing the benefit of AI regarding ethical and socio-economic challenges.

- The principle of non maleficence: 'Do no Harm'

We agree that AI systems should protect the dignity, integrity, liberty, privacy, safety and security of human beings. AI applications are being developed by humans and it must be understood that high efforts and continuous improvements are necessary to reduce potential risks.

Specific to each use case, the necessary quality level and fault tolerance, including fall back solutions in case of error and the required effort for testing, monitoring and controlling must always be defined under consideration of the field of application, such as: (i) how autonomous the AI may act (ie, differentiate between situations in its sole purpose as an information source, an assistant function in which the final decision is with the human, or if it is completely autonomous); (ii) how autonomous it may learn (ie, if re-training on the market possible and to what extent); (iii) the opportunities and possible risks and which machine learning method is used.

- The principle of autonomy: 'Preserve Human Agency'

AI can help people making better and more informed decisions. The two aspects highlighted in this paragraph are essential. First is the matter of choice: where possible, an alternative to being subject to direct or indirect AI decision-making should be provided to the user. However, it should also be considered that there are technological limits and that where real alternatives are possible today, there will be even more in the future. The right to opt out and withdraw from all AI decision-making does not seem realistic with the increasing use of the technology and will be difficult to implement. This might cause harm to others or prevent an authority from performing its duties for the common good. Such a 'right' cannot be horizontal – it must vary according to the use case and should be based on the type of AI system (the sensitivity of the use case).

Second, is the matter of transparency. When an interaction with AI is taking place and where crucial decisions are made by algorithms, a risk-adequate and use case specific approach is crucial. However, in the current phase of the deployment of AI, we believe that it should be transparent where and in what form AI is being used.

- The principle of justice: 'Be Fair'

AI systems should be designed in a way that the predictions resulting from training data are fair and as unbiased as possible. Because AI systems are designed by human beings and are trained using data that reflects our imperfect world, it's important that developers are aware how bias can be introduced into AI systems and how it can affect AI-based recommendations. We should target the reduction of unfair decisions (due to bias) and increase transparency. At a minimum, AI-based solutions will increase consistency and deliver a standardised approach/decision.

As noted in our comments on the definition of 'bias', removing all forms of bias from any finite might not be possible, which the draft Guidelines themselves recognise (see p.16). Hence, we would encourage the HLEG to revise this principle to target '**unfair**' bias.

- The principle of explicability: 'Operate transparently'

We fully agree with the AI HLEG that transparency and explainability are the key success factors to increase the acceptance and trust in AI systems. We agree that transparency means that the function of AI is explained an understandable manner, however, we would also add a contextual consideration in that the level of transparency depends on the application. In terms of 'business model transparency', we believe that the first basic requirement there is the need to inform individuals on whether or not they are interacting with an AI system. Beyond this, it means explaining the result, the base for decision-making and the benefit of the system. Providing the user with transparency, though, should not be afforded at the expense of a company's business model as this is sensitive information, impractical and often unachievable as they evolve frequently.

We agree that explicability is a precondition of trust, however the draft guidelines seem to confuse general principles with AI-specific issues by linking explainability de facto with 'informed consent'. 'Informed consent' is a GDPR term with a specific meaning, and therefore its use with regards to explicability must be more precisely defined. As GDPR allows data processing based on legal bases other than consent, like legitimate interest, we suggest removing this concept altogether and replacing it with the focus of these guidelines – trust.

Critical concerns raised by AI (p.11)

- 5.1. Identification without consent:

This chapter focuses on the consent in terms of privacy law. In addition to consent, the GDPR offers further legal grounds for data processing such as the contract and the legitimate interest. Therefore, the draft guidelines should not restrict the options that provide control to individuals as foreseen in data protection law.

Regarding identification, we agree that there must be differentiation between the identification of an individual and the tracing and tracking of an individual, but one must also be mindful of what practices are harmful and not harmful, lawful and unlawful. Although we agree that the identification without consent could be a critical concern in some scenarios, it might not be a critical concern in others but actually beneficial. We therefore recommend that the final Guidelines approach this issue with a specific focus on use cases where identification without consent poses an elevated risk of harm to individuals or society. Moreover, the idea developed in the section of 'developing entirely new and practical means by which citizens can give verified consent to being automatically identified by AI or equivalent technologies', is a dangerous path towards a situation in which citizens' choices are overridden by others who think they made the wrong decision, simply because it is believed that they did not give it enough consideration.

We also recommend that the Guidelines expressly acknowledge that different applications of AI might warrant different **types** of consent. In higher risk scenarios explicit consent might be appropriate, while in lower-risk scenarios, consent may be expressed implicitly, eg, by clearly informing a consumer that stepping into a store a store will entail the use of AI tracking to enable 'frictionless' shopping experiences.

Finally, the Guidelines should note that many of these issues relating to identification – and so to processing of personal data – are already governed by the GDPR and other EU law.

- 5.2. Covert AI systems:

We agree with the statement that AI developers and deployers should ensure that humans are made aware of – or able to request and validate – the fact that they are interacting with an AI identity. In addition, we would note that the principle is potentially under and over-inclusive, depending on how one understands the notion of 'interacting' (p. 11).

- 5.5. Potential longer-term concerns

We suggest deleting this section. With the technology evolving, long-term impacts cannot be predicted. The probability of potential occurrences as mentioned by the HLEG ('examples thereof are the development of Artificial Consciousness, ie, AI systems that may have a subjective experience of Artificial Moral Agents or of Unsupervised Recursively Self-Improving Artificial General Intelligence (AGI)' p.13), are currently relatively low and well into the future. The purpose of the Guidelines is to be practical and immediately applicable by focusing

on realistic and existing challenges while remaining attentive to future development of critical topics. There is no way to identify all possible scenarios, and we believe that the principles around which the Guidelines are built are broad enough to inform decisions on scenarios we do not foresee today. Ultimately, the goal of the Guidelines, and of any ethical principles in this space more generally, should be technology neutrality.

Chapter II: Realising trustworthy AI

The Guidelines' conception of 'data governance' is too narrow and is not reflective of the fact that governance structures necessary to develop AI ethically include a broader range of engineering and design practices (eg, access controls, systems documentation, etc.). We therefore urge the HLEG to recognise that data governance is complex in practice and will need to be tailored to individual scenarios.

In general, this chapter contains too many requirements and will make it challenging for developers to make them operational. For the sake of clarity, we would recommend amending and merging some as follows:

- 'Data **quality and** governance' and 'respect for privacy';
- 'Design for all' and 'non-discrimination';
- 'Governance of AI Autonomy' and 'respect for human autonomy'; and
- 'Robustness' and 'safety'.

Requirements of Trustworthy AI (p.14)

1. Accountability

As rightly written by the AI HLEG, the topic of accountability is highly dependable on the use case, the field of application, the autonomy of the AI and many more factors. A general approach or 'one-size-fits-all' solution should not be targeted. We would add the sentence, 'accountability might include the ability to contest the output and provide feedback on why a certain result is right/wrong', which is essential to learning systems.

2. Data **Quality and** Governance

We would amend the title as follows: 'Data **Quality and** Governance'. Data quality and data integrity potentially have a big impact on the AI-systems. It is important to be aware of how limited data sets, bias in data or other factors impacting data quality can directly affect AI-based recommendations. All stakeholders should aim at the reduction of unfair decisions (eg, due to bias), and increase transparency to continuously improve our data sets and AI systems. Furthermore, it needs to be recognised that in certain cases bias is intended because of the objective of the AI system.

3. Design for all

We support the observation that 'systems should be designed in a way that allows all citizens to use the products or services [...]' (p. 15), as our members' aim is for widespread adoption of AI technology in a way that is beneficial to society. At the same time, this requirement should also recognise that some flexibility may be needed when determining how to design for all, depending on the product or service concerned.

4. Governance of AI autonomy (human oversight)

We fully agree with the analysis done by the AI HLEG that the concrete ways to implement human oversight (through safety, accuracy, adaptability, privacy and explicability) will differ depending on the application and specific AI systems. More concretely, we propose to always review what level of autonomy in decisions should be applied (ie, distinguish between AI used only as a source of information, AI as an assistant with final decision by user, or AI that acts fully automated without human involvement). Furthermore, we find it essential to also review the level of autonomy in learning (may the AI learn on the market (retraining possible), with limited parameters (no safety relevant parameters), or if no learning or evolvement on the market is possible). As a third dimension, we suggest reconsidering the level of risk (eg, which persons or laws could be harmed and how). On this basis, a use-case specific decision should be taken. Moreover, a fallback solution for fully automated AI-systems should be prepared when human involvement is necessary.

5. Non-discrimination

As mentioned in our comments on the scope, most data scientists would agree that virtually any datasets will reflect at least some types of bias. Therefore, the Guidelines should not be aimed at **eliminating** all biases in datasets used to train AI, but rather at better understanding how limitations in datasets might impact the outputs of algorithms and taking all necessary steps to mitigate the risks these limitations might generate.

6. Respect for (and enhancement of) human autonomy

We agree that the user's well-being should be central to AI deployment and AI-systems should, where possible, promote conscious decisions by the users regarding the delegation of responsibility to the system. Applied well, AI could enable people to indicate much more precisely their preferences than would be otherwise practically feasible.

7. Respect for privacy

We have a strong framework for data privacy in Europe with the GDPR and this is applicable to AI-systems. The Guidelines could mention the use of encryption, pseudonymisation and other privacy protective techniques that reduce privacy risk to individuals while still allowing for the development of AI solutions that can be beneficial to society. Moreover, they could stress that AI can be used to enhance privacy and its potential to do so should be further explored. However, no new framework or regulation is needed specifically to address AI.

8. Robustness

The algorithms must be secure, reliable and robust enough to deal with errors or inconsistencies during the execution, deployment and user phase of the AI system. Therefore, there must be an extensive design and development phase, during which developers take appropriate measures to ensure safe and robust operation in the public.

Reliability & reproducibility – We do not see these elements as core to the development of 'trustworthy AI'. The draft Guidelines should consider the limits of reproducibility, especially when placed on the market and retraining (by the user) is possible. These aspects should be handled by the overall system design and by extensive testing before and after being placed on the market.

Accuracy – Accuracy of AI systems is limited and is directly linked to the data set used for training. More guidance is needed on what level of accuracy is required for AI systems, especially for sensitive use cases.

Fall-back plan - The fall-back plan should depend on the use case and may not always be necessary.

9. Safety

It is worth adding that in many applications machine-learning increases the overall performance of a system, including in terms of safety compared to a strictly rule-based system.

The AI system should be safe and not harm the user or his/her rights, it also should be reliable and do what is expected. Minimising the risks of the whole system, testing and quality monitoring will be key elements besides setting the right quality criteria (such as false positive vs. false negative rate).

10. Transparency

It is important that there is a certain basic level of transparency or explainability to earn the user's trust. On the other hand, greater transparency will be necessary for developers and operators to ensure quality monitoring and continuous improvement, as well as for use cases with potentially higher risks. Transparency levels provided should be contextualised and risk-based. However, achieving transparency can be complex and highly dependent on a host of variables, precluding anything resembling a 'one-size-fits-all' approach. Certain AI technologies, including deep neural networks, are so complex that they go well beyond what's comprehensible to humans. In these contexts, the overall goal of transparency would be ill-served. Stakeholders should be

required to provide '**meaningful information**' about choices and decisions concerning data sources and development processes and for uses that can have significant impact.

Technical and non-technical methods to achieve trustworthy AI (p.18)

The technical and non-technical methods listed to achieve 'Trustworthy AI' are a good non-exhaustive list, but more emphasis should be placed on the role of standardisation and codes of conduct. As the AI HLEG mentioned, the listed methods are meant to present only a sample of possible methods and should be continuously reviewed. They should only be considered as best practice and activities for further collaboration and research and therefore should not be included into the formal endorsements.

Traceability & auditability – We believe that the development of human-machine interfaces that provide mechanisms for understanding the system's behaviour is essential. However, the nature of auditability will be heavily context-dependent. In complex scenarios, third party auditors and expert controls will be more effective for technical support. In still other scenarios, internal organisational auditing and controls may suffice. In light of this, the Guidelines should do more to acknowledge that effective auditing, depending on the context, can include any of those mechanics.

Standardisation – We would like to stress that the nature of AI makes it difficult to imagine a horizontal standard that would be meaningful across applications and sectors.

Accountability governance – The draft Guidelines rightly stress the importance of having a data governance programme with competence over AI. Whether this is specifically deemed an Ethical AI review board or whether it has a broader mandate also capturing AI, is perhaps less relevant and should depend on the scale and nature of AI work that a company is performing.

When developing these mechanisms, the global dimension should not be forgotten. As mentioned in the executive summary of the draft Guidelines, we strongly agree with the view that AI and its development exists within a global ecosystem and therefore Europe should work tirelessly to shape the global debate on AI governance to promote trustworthy AI for all citizens.

Chapter III: Assessing trustworthy AI

We generally support the 10 requirements for 'Trustworthy AI', but believe that the long list of questions requires more work to be used by developers. In its current form, this chapter is frequently inconsistent and repetitive and the questions it seeks to answer are often too high-level to have any practical use.

Furthermore, precise questions for AI auditing or assessing will vary from use case to use case, and a tailored response needs to be provided for each specific situation or question. The development of use cases will be essential to make the guidelines practical and actionable. We strongly encourage the HLEG to continue and focus the work on such use cases, in cooperation with industry and civil society.

The following items are crucial:

- Proportional, risk-based flexible and voluntary guidelines (not all questions are necessary to consider for a 'simple AI tool');
- A 'holistic' approach with high-level questions to easily identify critical topics/use-cases, etc.;
- Clear definitions (e,g, AI taxonomy, levels of transparency, bias);
- Measurability of the questions; and
- Clear differentiation of the questions and thus how it should be implemented, eg:
    - What is use case specific or has a broader context;
    - What is legal/ethical topics and what are technical solutions; and
    - Responsibility of business/development function and of governance function.

General comments

Glossary (p.iv)

In the definition of 'artificial intelligence (p.iv), we believe the final Guidelines should provide greater clarity on what is meant by 'deciding the best action(s) to take'. In particular, Article 22 of the GDPR articulates the concept of a 'decision based solely on automated processing'. Is the AI definition set forth in the Guidelines coextensive with the GDPR, or is it narrower (or broader)?

In addition, we note that the Guidelines' definition of AI is narrower than many common understandings of the term. Many solutions in use today that are described as having an AI component do not necessarily 'decide' on a course of action; instead, many of them make connections, reveal correlations, or provide other insights that humans then use to decide on a course of action. We therefore believe that the proposed definition of AI should be modified to reflect this point.

The Guidelines define 'bias' as 'prejudice for or against something or somebody, that may result in unfair decisions' (p. iv). In view of most data scientists, virtually any datasets will reflect at least some types of bias (eg, traffic data collected in large cities might not accurately reflect traffic patterns in smaller cities). The goal should not be to *eliminate* all biases in datasets used to train AI, as this is effectively impossible for most (and possibly all) finite datasets. Rather, the goals should be: (i) to take steps to mitigate the risk that an AI solution might generate *unfair* biases; and (ii) to help people understand the scope, characteristics and limitations of the dataset(s) on which an AI solution was trained, so that people can better understand how these limitations might impact the outputs generated by the AI in any given application.

Finally, the Guidelines frequently use the terms 'transparency', 'explicability' and 'explainability' interchangeably. In our view, 'transparency' is a broader concept than 'explicability', the latter being also linked to an important separate term, 'intelligibility', that is somewhat overlooked in the Guidelines. We therefore encourage the HLEG to include each of these four terms in the glossary to help clarify intended meanings for stakeholders.