

---

## Consultation Response

### European Commission HLEG on AI Draft Ethics Guidelines for Trustworthy AI

1 February 2019

---

To be submitted electronically via [www.ec.europa.eu/futurium/en/ai-alliance-stakeholders-consultation/stakeholders-consultation-draft-ai-ethics-guidelines](http://www.ec.europa.eu/futurium/en/ai-alliance-stakeholders-consultation/stakeholders-consultation-draft-ai-ethics-guidelines)

The Association for Financial Markets in Europe (AFME) welcomes the opportunity to comment on **the EUROPEAN COMMISSION HIGH LEVEL EXPERT GROUP (HLEG) ON ARTIFICIAL INTELLIGENCE (AI) DRAFT ETHICS GUIDELINES FOR TRUSTWORTHY AI**. AFME represents a broad array of European and global participants in the wholesale financial markets. Its members comprise pan-EU and global banks as well as key regional banks, brokers, law firms, investors and other financial market participants. We advocate stable, competitive, sustainable European financial markets that support economic growth and benefit society.

AFME is the European member of the Global Financial Markets Association (GFMA) a global alliance with the Securities Industry and Financial Markets Association (SIFMA) in the US, and the Asia Securities Industry and Financial Markets Association (ASIFMA) in Asia.

AFME is registered on the EU Transparency Register, registration number 65110063986-76.

We summarise below our high-level response to the consultation, which is followed by answers to the individual questions raised.

#### General Comments

AFME commends the European Commission in appointing the High-Level Expert Group on Artificial Intelligence and establishing a forum – the European AI Alliance – to engage a broad and open discussion on the strategic importance of AI in Europe and globally. AFME welcomes the first step of the HLEG to draft AI guidelines on ethics as communicated in the March 2018 European Initiative on AI<sup>1</sup>. This a complex and challenging topic which requires significant discussion and input from a wide range of participants. AFME looks forward to engaging further with the HLEG's final Guidelines and its upcoming work on Policy and Investment Recommendations.

As with many industries, the application of AI has the potential to transform capital markets and is already impacting many aspects of how the industry operates, from trading and client interactions to risk management and operational processing. However, AI is a rapidly evolving technology that could have far reaching impacts on society. Care must be taken to ensure its use conforms to appropriate ethical standards applied within individual banks and does not unintentionally harm the market or clients. Equally, policy or regulatory frameworks must be supportive of the development of AI as to not stifle innovation and the potential benefits, while maintaining the appropriate balance against market and consumer protection.

Capital markets banks have existing codes of business conduct which include ethical principles or have separate, dedicated codes of ethics. These codes outline the responsibilities and obligations on a bank's individual employees' and on the overall bank, covering areas such as: complying with applicable laws and regulations; exercising fair judgement; and executing activities openly and fairly. They are designed to address significant risks that banks face, such as systemic, customer and reputational risks, and are reviewed

---

<sup>1</sup> <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>

regularly to ensure that they keep pace with developments in technology and markets and with shifts in ethical and cultural expectations.

The AFME responses to the sections posed by the Consultation are outlined below. Overall, we feel that the structure and content of the document may require further refinement in order to more clearly identify, and simplify, key concepts and recommendations. We believe that the Guidelines should more readily apply to the broadest application of relevant industries and AI use cases. We also believe that too quickly prescribing formal requirements and assessment criteria may fail to capture, or limit the maturity and continued adoption of, AI. For instance, it is not the case that AI applications that do not immediately meet the principles outlined in these Guidelines should be prohibited, but that further analysis may be necessary. These Guidelines should remain voluntary and follow the collection of wider stakeholder input at an industry level.

Given the tight timeframes for completion of the Guidelines, we encourage the HLEG to consider an additional consultation on Chapters II and III, to ensure that the diverse impacted sectors and interest groups have an opportunity to provide input. We suggest that this part of the paper be adopted in its final form at the same time as the Policy and Investment Recommendations. This consultation could be launched at the same time as the finalisation of Chapter I. If Chapter II and III are to be adopted at the same time as Chapter I we would like the paper to restate that these documents are intended as a living document.

Finally, we request confirmation that these Guidelines will be voluntary, and further clarity on the nature and timing of the attestation mechanism that will be used. We encourage the HLEG to ensure an opportunity for the public to review and provide input into the design of this mechanism to ensure that it works across a broad range of AI users.

We would be pleased to discuss the content of this response further.

## **Introduction: Rationale and Foresight of the Guidelines**

We would like to raise the following items for consideration:

**Glossary:** As with any other section of the Guidelines that is incomplete at this stage, we encourage the HLEG to ensure that changes to the glossary, including any revisions or additional terms are made available for public consultation before being finalised, to avoid the risk of the inclusion of definitions that may not be applicable to the very broad uses of AI.

**The role of AI ethics:** It would be helpful to emphasise in this section the subjectivity of 'ethics' as a concept, varying between individuals and cultures if the HLEG intends to successfully foster reflection and discussion at a global level. With reference to our 'purpose and scope' comments below, this should be taken into account in the form of flexibility for industry regulators and individual firms to apply the Guidelines to their individual situations, using them as part of their decision-making process for the use of AI.

**Benefits of AI:** The list of possible benefits of AI on page 1 refer to specific use cases, rather than the broader possibilities for the technology across all industries. We suggest that the Guidelines should acknowledge that AI has the potential to benefit all aspects of EU citizens and industry sectors, rather than starting from a narrow position, such as the examples listed on transportation, social welfare, climate change and natural resources.

**Purpose and scope:** It is unclear from this consultation exactly what status the final Guidelines will have. Under 'Scope of the Guidelines', it is noted that they should not be a substitute to any form of policy-making, regulation, or internal guidelines, and are not an official European Commission Document or legally binding. However, section B on page 3 states that '...it is important that AI developers, deployers and users also take actions and responsibility to actually implement these principles...'. Further clarity would be welcomed that

these Guidelines will be voluntary, as well as on the nature and timing of the attestation mechanism that will be used. As with the Glossary above, we are concerned that there will be no opportunity to input into the design of the mechanism, to ensure that it works across a broad range of AI users.

## **Chapter 1: Respecting Fundamental Rights, Principles and Values – Ethical Purpose**

Overall, we support many of the concepts outlined in this section. AI has the potential to bring many positive impacts for the financial services industry and Europe as a whole, and an ethical approach to AI should maximise the benefits for all.

With this in mind, we would like to raise the following items for consideration:

### **Section 1- The EU’s Rights-based Approach to AI Ethics:**

We are concerned that the statement that ‘adopting a rights-based approach will limit regulatory uncertainty’ could lead to some ambiguity. In our experience regulatory uncertainty is limited by considered and proportionate legislation, created in consultation with the relevant industry, and by ongoing dialogue with relevant regulators. From the perspective of AI in capital markets, use of AI is already covered by a number of existing regulations as part of a wider-framework of technology-agnostic, outcomes-based requirements.

### **Section 2 – From Fundamental Rights to Principles and Values:**

Defining “ethical purpose” through three discreet, yet interconnected, themes is complex and may not be easily understood by all persons that are required to interpret the Guidelines. We suggest that a simpler definition of ethical purpose should be developed that can be more easily consumed by all persons, including laypersons, that will need to refer to the Guidelines. In this respect, the Guidelines should mirror their own requirements for AI in being “comprehensible and intelligible by human beings at varying levels of comprehension and expertise”.

### **Section 3 – Fundamental Rights of Human Beings:**

Equality, non-discrimination and solidarity/the principle of justice: While fairness remains a key measure, it is important to note that fairness should not necessarily mean equality, i.e. that the AI application delivers the same output for all individuals or groups. This would impact on the effectiveness of AI models; whose results respond to mathematic processes on the input data.

### **Section 4 – Ethical Principles in the Context of AI and Correlating Values:**

The principle of non-maleficence (1): While AFME agrees that the aim of AI should be to ‘do no harm’, the definition of harm should be carefully considered. We suggest that it should instead be amended to ‘prevent harm’. For instance, if a firm uses an AI application to perform suitability checks on its clients, it should not be considered harmful to withhold services from clients that do not pass the assessment. Indeed, the ramifications of providing unsuitable services to individuals or groups can be significant and harmful to society more broadly. Furthermore, we consider that the principles of justice and explicability could be subsumed under this principle.

The principle of non-maleficence (2): We are concerned by the use of the term ‘negative profiling’. The activity of profiling is not in itself sinister. However, care should be taken that the processing of data on an individual, which may include profiling, does not have a negative impact on the individual. This obligation is in accordance with Article 22 of the General Data Protection Regulation (GDPR – Regulation 2016/679).

The principle of autonomy (1): There are situations in which it may be extremely important for an individual interacting with an AI application to be subordinate to that application (while human oversight of the application as a whole is maintained). For example, it may be necessary to prevent certain individuals from

over-riding AI applications concerned with safety systems or the detection of crime, such as anti-money laundering (AML).

The principle of autonomy (2): We agree that it is important, and indeed mandated under GDPR, that individuals should have the right not to be subject to solely automated decision making. However, this right does not preclude an automated recommendation being taken into account when the ultimate decision is made by a human.

The principle of autonomy (3): In addition, the extent to which an individual need to be made aware that they may be interacting with an AI application may depend on the function of that application and the materiality of the impact of that knowledge. For example, it may not be necessarily important to an individual to know that they are interacting with a 'chatbot' rather than a human when a firm is providing certain types of customer service.

The principle of explicability: While transparency may be crucially important for some applications of AI, the extent to which it is necessary to be able to explain the internal workings or decision logic of an AI application will vary depending on the function that application is performing. For example, an AI application that routes trade exceptions (e.g. a failed trade) to an operational process within a firm may not require a significant degree of transparency, provided that incorrect outcomes can be amended, and the application can learn from those amendments. There are important use cases where a lack of transparency in the decision making process of an AI provides a level of security, accuracy and fairness, for example applications that detect possible financial crime, cyber incidents or terrorist financing. It should also be borne in mind that AI is a technology which should augment, rather than replace, humans. Given that human decision making is not always transparent or fully explained, it would be better to frame this principle in terms of trust.

#### Section 5 – Critical Concerns Raised by AI:

We agree with the assessment that there may be situations in which it is important for AI systems to identify individuals, particularly the examples given of detection of fraud, money-laundering or terrorist financing.

As above, we note that consideration should be given to the extent to which an individual needs to be made aware that they may be interacting with an AI application.

Identification without consent: We suggest that the Guidelines' wording in relation to GDPR Article 6 should be slightly revised, as the Guidelines currently suggest that data processing is only valid to meet a legal obligation. However, GDPR Article 6 lists several bases for lawful data processing, of which compliance with a legal obligation is only one. We believe that the data processing requirements in GDPR are sufficient, and that it may be more appropriate to include a general statement stating this within the Guidelines. This would also future-proof the Guidelines in the event that amendments are made to the GDPR.

## **Chapter 2: Realising Trustworthy AI**

AFME commends the HLEG in seeking to design practical implementation guidelines for firms using AI. A principles-based approach to emerging technologies is appropriate to balance innovation risk and security. We are largely in agreement with the principles proposed and their intentions.

With this in mind, we would like to raise the following items for consideration.

#### Section 1 – Requirements of trustworthy AI:

Data governance: While AFME agrees that datasets may contain biases, it should not be assumed that this is 'inevitable' and/or that complete removal of all bias is a prerequisite for the use of such data within an AI

application. Instead, such risks should be mitigated, including via ongoing assessment of the application's outputs.

**Governance of AI autonomy:** While human oversight will always remain important, it should not be assumed that "...the greater degree of autonomy that is given to an AI system, the more extensive testing and stricter governance is required...". AFME believes this should be decided by an appropriate assessment framework, dependent on the system and/or industry. For example, greater governance and human oversight may be more appropriate for systems that interact directly with humans, rather than by the AI system's overall level of autonomy. AFME agrees with the statement that 'different levels or instances of governance' (including human oversight) will be necessary.

**Non-discrimination:** We support the statement that while it may be *possible* to remove bias from data, bias is inherent. It is important to acknowledge that, while mitigation for bias is a key standard for AI development, it is impractical to require the removal of all bias. A good test might be that use of the AI application leads to less bias than an alternative system or human process would.

**Robustness:** While contingency plans are an important part of any technology governance strategy, it should be considered that two of the key benefits of AI are the speed and scale at which data can be processed. It may therefore be that, in the event of a systems outage, humans would be unable to partially or fully backfill an AI system. The Guidelines should consider that robustness may be achieved by other means, and not just through human backfill.

**Transparency:** As above, we note that the ability to explain the internal workings or decision logic of an AI application will vary depending on the function that application is performing.

## Section 2 – Technical and non-technical methods to realise trustworthy AI:

AFME agrees with the assessment that both technical and non-technical methods must be used, and that good governance of AI should involve a continuous process of assessment and adjustment.

**Traceability and auditability:** We are concerned by the statement that "laypersons should be able to understand the causality of the algorithmic decision-making process and how it is implemented by organisations that deploy the AI system". AI has the potential to deliver huge benefits in a wide range of applications, but in some cases may be a complex technology. While proximate explanations (for individual decisions) are sometimes possible, global explanations of the algorithm, especially if in non-symbolic language for laypersons, are often not. As noted above, the explainability of AI will vary depending on the use to which it is being put. Firms should instead focus on developing sufficient understanding of the technology at management level and within control functions, to ensure appropriate oversight and governance.

**Regulation:** It is crucially important that regulatory bodies develop the skills and resources to respond to and support the development of AI within their industries. This will also allow development of AI as a regulatory tool, for example for assessing large quantities of data or predicting the build-up of risk.

**Standardisation:** AFME agrees that greater standardisation of terms and frameworks related to AI would be of great benefit. Given the cross-border nature of many industries and firms, it would be most useful if such standardisation occurred at a global level, considering initiatives taking place in other jurisdictions.

**Codes of conduct:** As noted above, there is a lack of clarity as to the exact status of this document and how the adherence process is intended to work in practice. Further consultation on this would be welcome.

**Education and awareness:** This is already a key priority for the capital markets industry. As the possible applications and benefits of AI expand, capital markets banks are increasingly investing in AI education and training for staff across their businesses.

### Chapter 3: Assessing Trustworthy AI

AFME agrees that assessment of Trustworthy AI will be important and welcomes the initial template from the HLEG as a draft for further consideration. Assessments of this nature will ensure that the benefits of the technology can be maximised while minimising the risks, and that the ethical considerations expressed in the Guidelines are addressed.

AFME believes that as with all new and developing technologies, it is important that the risks are considered and actively managed. A robust control framework, similar to those that are already in place for other technologies, should be a priority for any capital markets institution investing in the many forms of AI.

However, AFME believes that at this stage defining detailed assessment criteria in the form of a prescriptive checklist requires further stakeholder engagement. AFME believes that while high-level principles are useful (such as the MAS FEAT Principles<sup>2</sup>), detailed assessment criteria will need to be defined at both an industry and individual firm level, as it relates to the type and use of AI systems. Attempting to create overarching assessment criteria at this stage may inhibit firms adopting AI in the early stage of maturity or leave some important areas not fully considered sufficiently.

**Accountability:** AFME agrees that accountability for AI is integral for its ongoing use. Each firm's framework of governance and risk management should ensure accountability for the establishment of, and decisions involved in, each use of AI and for setting principles for implementation of policies, procedures and the allocation of responsibilities. For example, the assessment list has items related to governance (as it relates to human oversight, responsibility and accountability) in multiple requirements - Accountability, Data Governance, Governing AI Autonomy. AFME suggests that all governance related considerations should be consolidated under one requirement for consistency.

**Respect for privacy:** We are concerned that the questions listed under this section may not be specific enough for an AI contact. Consideration of privacy concerns should go beyond compliance with GDPR or issues of consent.

**Respect for human autonomy:** AFME suggest that this section more closely relates to the requirements for Transparency and for simplicity should be considered under that header. In addition, bullet four refers to users of AI having the facility to 'interrogate' algorithmic decisions in order to fully understand their purpose and data used. AFME suggests that it would be more appropriate for the AI system owner to be responsible for providing, on request, clear explanations and information related to an AI decision that relates to a user.

**Robustness:** AFME suggests that, as with all current technologies, the forms of attack that may impact an AI system will be broad, and may be both internal (for example, an insider threat within an organisation where the AI system resides), or external (for example, a cyber-hack). While many attack scenarios can be mitigated, we believe it is important to emphasise that it is a continuous process for firms to remain resilient within a dynamic threat landscape.

#### AFME contacts

Fiona Willis, [fwillis@gfma.org](mailto:fwillis@gfma.org) +44 (0)20 3828 92739

David Ostojitsch, [david.ostojitsch@afme.eu](mailto:david.ostojitsch@afme.eu) +44 (0)20 3828 92761

---

<sup>2</sup> <http://www.mas.gov.sg/News-and-Publications/Media-Releases/2018/MAS-introduces-new-FEAT-Principles-to-promote-responsible-use-of-AI-and-data-analytics.aspx>