STANDARD FORM OF EXPLANATORY MEMORANDUM FOR EUROPEAN UNION LEGISLATION AND DOCUMENTS
8396/19*
COM (2019) 168 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS
Building Trust in Human-Centric Artificial Intelligence

Submitted by Department for Digital, Culture, Media and Sport on 3 May 2019

SUBJECT MATTER

1. The Europe Union's Artificial Intelligence (AI) Strategy was launched by the Commission in April 2018. It has three aims: boost the EU's tech and industrial capacity including uptake across the economy; prepare for socio-economic changes; and ensure an appropriate ethical and legal framework.

2. This Communication highlights the publication and launch of a piloting phase for the *'Ethics Guidelines for Trustworthy AI'*. The guidelines were developed by the Commission's High Level Expert Group (HLEG) - a multi-stakeholder group representative of different Member States alongside academic and industry perspectives. This group is responsible for developing broader AI policy recommendations due in May 2019. The guidelines are a key output of the EU's AI strategy and associated delivery plan - *The Coordinated Plan on AI* - published in December 2018. Following a stakeholder consultation earlier this year (December 2018 - February 2019), the HLEG presented revised guidelines to the Commission in March.

3. The piloting phase will be launched by the summer of 2019 and last until the end of the year. All interested stakeholders will be able to participate by indicating their interest through the European AI Alliance - a multi-stakeholder forum for engaging in a broad and open discussion of all aspects of AI development and its impact on the economy and society. After this piloting phase they will be further revised at the beginning of 2020.

4. The guidelines highlight trust as a prerequisite to ensure a human-centric approach to the development of AI. The document captures the founding principles of the EU as guiding principles for this work (respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights), as well as the EU Charter of Fundamental Rights which captures the personal, civic, political, economic and social rights enjoyed by people within the EU.

5. AI brings new challenges as machines are increasingly able to make decisions without human intervention and therefore have the potential to threaten these values. Broadly speaking, AI systems should adhere to the law, ethical principles and ensure no unintended harm is caused. In doing so, diversity should be considered at every stage of development to capture how these decisions could affect different groups across society. These guidelines are being developed to ensure this and account for

existing regulatory frameworks. The desired outcome is to establish an ethical level playing field across all Member States from developers, suppliers and users alike.

6.   The guidelines build on the work of the European Group on Ethics in Science and New Technologies and the Fundamental Rights Agency. The group has developed seven key requirements based on the following three components that are needed to develop 'trustworthy AI': 1) comply with the law; 2) fulfill ethical principles; and 3) be robust. The following are generic and applicable across all settings but context should always be taken into account to assess the relative weight of each requirement.

7.   The seven key requirements are:
a.   Human agency and oversight: *the primacy of the human being is highlighted here and that AI should act as enabling, rather than diminishing human autonomy and rights. In practice this means including appropriate degrees of control measures, human oversight and governance mechanism during its development and deployment; this is especially relevant for public authorities having oversight powers in line with their mandate.*

b.   Technical robustness and safety: *Algorithms must be secure reliable and robust enough to deal with errors and cope with erroneous outcomes. This includes being resilient against overt and subtle attempts to manipulate data and must include a plan of action in case of such issues. Outcomes should also be reproducible and should employ a safety and security-by-design approach to ensure their safety is verifiable.*

c.   Privacy and data governance: *Privacy and data protection must be guaranteed at all stages of the AI system's life cycle in order that individuals can trust their data will not be used to harm or discriminate against them. The quality and integrity of the data must be ensured to avoid embedding socially constructed biases. Finally access to this data must be adequately governed and controlled.*

d.   Transparency: *What and how a decision is made by the AI system should be documented in order to trace the history of the decision and enable the user to have redress. Explanation of how the decision was made should be adapted to its audience as far as possible, but especially regarding the extent to which it is used to shape the business model of an organisation. The limitations of the system should be communicated to stakeholders and users should always know when they are communicating with an AI system rather than a person.*

e.   Diversity, non discrimination and fairness: *Data sets may suffer from historic bias, incompleteness and bad governance models resulting in indirect discrimination. Such concerns should be tackled from the beginning of system development. Ensuring diverse design teams and public participation during development should be employed to mitigate this.*

f. Societal and environmental well-being: *The impact of AI on society as a whole, including the environment should be taken into account. AI's ability to contribute to UN Sustainable Development Goals should be encouraged, whilst caution around its ability to undermine the democratic process should be given careful consideration.*

g. Accountability: *Mechanisms to ensure responsibility and accountability for AI systems and their outcomes is critical to increasing trust. Impact assessment can be used to establish potential negative impacts of AI and articulate where trade offs might need to be made.*

8. The Commission will use this document to promote the EU's approach internationally and build a consensus on a human-centric AI, a key topic in international fora since Japan's G7 Presidency in 2016. They will continue to strengthen cooperation with like-minded third countries (such as Japan and Canada), explore how non-EU countries and international organisations can contribute to the pilot phase; and continue to play an active role in international fora such as the G7 and G20 as well as through bilateral engagement with like minded countries too.

SCRUTINY HISTORY

9. An EM was submitted in January 2019 on a Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Coordination Plan on Artificial Intelligence.

MINISTERIAL RESPONSIBILITY

10. The Secretary of State for Digital, Culture, Media and Sport and Minister for Digital and the Creative Industries have responsibility for AI and data ethics policy.

INTEREST OF THE DEVOLVED ADMINISTRATIONS

11. This Communication involves non-binding guidelines for the ethical use of AI - an emerging policy area. As such, the devolved administrations have been consulted in preparation of this EM to ensure their interests are represented.

LEGAL AND PROCEDURAL ISSUES

12. There are no legal or procedural issues. This is not a proposal for legislation. The guidelines are non-binding and as such do not create any new legal obligations.

APPLICATION TO THE EUROPEAN ECONOMIC AREA

13. The Communication only explicitly states applicability to Member States.

SUBSIDIARITY

14.   There are no subsidiarity concerns.

POLICY IMPLICATIONS (including Exit implications where appropriate)

15.   On 23 June 2016, the European Union (EU) referendum took place and the people of the United Kingdom voted to leave the EU. Until exit day, the UK remains a full member of the EU and all the rights and obligations of EU membership remain in force. During this period, the Government will also continue to negotiate, implement, and apply EU legislation.

16.   The UK has a strong tradition in the careful navigation and management of the ethical and social complexities of new technologies and a proportionate pro-innovation approach to regulation. The UK already benefits from well established and robustly enforced personal data laws, as well as wider regulations that guide how data driven activities and sectors can operate.Effective governance will build confidence in, and demand for, new data-driven and AI-based innovations, amongst businesses and citizens. The Coordinated Action Plan's focus on improving development and deployment of AI technologies as well as increasing productivity and economic growth align closely with those of the Office for AI and the Government's Industrial Strategy.

17.   To ensure the ethical deployment of AI and data driven technology, the Government has established the world's first Centre for Data Ethics and Innovation (CDEI) - an independent advisory body that identifies the measures needed to strengthen and improve the way data and AI are used and regulated. This includes articulating best practice and advising on how we address potential gaps in regulation.

18.   The Government welcomes these non-binding guidelines as a helpful contribution to the global debate on the ethical deployment and governance of artificial intelligence. We are content that they incorporate our feedback but believe the pilot phase will be critical to assessing their utility for and, crucially, their impact on, industry, as well as balancing the ethical concerns of academics and governments/public authorities. They have the potential to foster dialogue at a global level in the *responsible* deployment of technology, particularly through the G7 and G20.

19.   We will continue to engage with the Commission during the piloting phase, up until exit day. Thereafter, we will explore opportunities to continue engagement. Through the piloting phase the Commission is considering mechanisms for including feedback from non-EU and third countries. The UK is seen as a leader in this space - including through our establishment of the CDEI - and therefore our contribution is particularly valued.

20.   We generally welcome the rights-based approach proposed in the guidelines as a more practical way to define the ethical challenges surrounding the deployment of AI. This approach complements the protection of human rights and freedoms operationalised in the GDPR, and the derivation and development of ethical principles from these rights. The guidance also complements the [UK's Data Ethics Framework](#), which sets out clear principles for how data should be used in the public sector.

21.   Across the following areas of work, we will consider how the guidelines fit with UK Government's policy positions and the CDEI's findings. We will continue to look to engage with the Commission as appropriate:

a.   **Centre for Data Ethics and Innovation projects:** At Budget 2018 the Government commissioned the CDEI to study the use of data in shaping people's online experiences, and the potential for bias in decisions made using algorithms. An interim update on progress is expected in summer 2019. Final recommendations will be made early next year;

b.   **AI4Gov:** The Office for Artificial Intelligence and the Government Digital Service began work in January this year to bring a strategic understanding of AI adoption across public, commercial and international contexts. The primary objective of this work is to identify AI use cases/opportunities to increase productivity and service quality across central government; and

c.   **Data trusts:** The Office for Artificial Intelligence partnered with the Open Data Institute (ODI) and Innovate UK to explore whether a data trust could increase access to data while retaining public trust. On 15 April 2019, the ODI published a number of reports setting out its findings and recommendations for next steps towards creating a functioning data trust. The Office for AI is considering the ODI's recommendations to determine how to take this work forward.

CONSULTATION

22.   We will continue to contribute to the HLEG's stakeholder engagement opportunities representing HMG's position on further iterations of these guidelines, as appropriate.

IMPACT ASSESSMENT

23.   An Impact Assessment has not been included.

FINANCIAL IMPLICATIONS
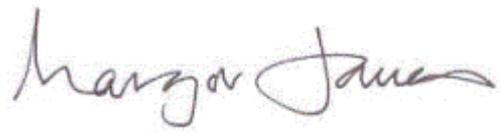
24.   There are no financial implications.

TIMETABLE

25.   After this piloting phase begins in summer 2019, the guidelines will be further revised at the beginning of 2020.

OTHER OBSERVATIONS

26.   There are no other observations.

MINISTERIAL NAME AND SIGNATURE

Margot James
Minister for Digital and the Creative Industries
Department for Digital, Culture, Media and Sport