High-Level Expert Group on Artificial Intelligence publishes Ethics Guidelines for Trustworthy AI

Fahy, R.F.

Publication date
2019

Document Version
Final published version

Published in
IRIS

Link to publication

Citation for published version (APA):
Fahy, R. F. (2019). High-Level Expert Group on Artificial Intelligence publishes Ethics Guidelines for Trustworthy AI. *IRIS*, *2019*(7), 3. [3].

OBSERVATOIRE EUROPÉEN DE L'AUDIOVISUEL
EUROPEAN AUDIOVISUAL OBSERVATORY
EUROPÄISCHE AUDIOVISUELLE INFORMATIONSSTELLE

IRIS

Legal Observations
of the European Audiovisual Observatory

IRIS 2019-7/3

European Commission: High-Level Expert Group on Artificial Intelligence publishes Ethics Guidelines for Trustworthy AI

On 8 April 2019, the High-Level Expert Group on Artificial Intelligence (AI), which is an independent expert group set up by the European Commission, published its Ethics Guidelines for Trustworthy AI. The Guidelines are timely, given that both the Council of Europe (see IRIS 2019-4/3) and UNESCO (see IRIS 2019-1/8) have also been examining the benefits and risks of AI, and indeed, on 17 May 2019, the Foreign Ministers of the Council of Europe member States agreed to examine the feasibility of a legal framework for the development, design and application of artificial intelligence.

The purpose of the Guidelines is to promote trustworthy AI, and it sets out a framework for achieving this. The Guidelines contain a lengthy definition of AI systems: software systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal.

The Guidelines begin by noting that trustworthy AI has three components which should be met throughout the system's entire life cycle: (a) it should be lawful, complying with all applicable laws and regulations; (b) it should be ethical, ensuring adherence to ethical principles and values; and (c) it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm.

The 41-page Guidelines are divided into three chapters, with Chapter 1 setting out the foundations of trustworthy AI, grounded in fundamental rights and reflected by four ethical principles that should be adhered to in order to ensure ethical and robust AI: (1) respect for human autonomy, (2) prevention of harm, (3) fairness, and (4) explicability. Chapter 2 then puts forward a set of seven key requirements that AI systems should meet in order to be deemed trustworthy: first, human agency and oversight, where AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights; secondly, technical robustness and safety, which requires that AI systems be developed with a preventative approach to risks; thirdly, privacy and data governance, where AI systems must guarantee privacy and data protection throughout a system's entire lifecycle; fourthly, transparency, where the data, system and AI business models should be transparent; fifthly, diversity, non-discrimination and fairness, where unfair bias must be avoided; sixthly, societal and environmental well-being, where broader society and the environment should be considered as stakeholders throughout the AI system's life cycle; and seventhly, accountability, where mechanisms must be put in place to ensure responsibility and accountability for AI systems and their outcomes. Finally, Chapter 3 provides a Trustworthy AI assessment list to operationalise Trustworthy AI which is primarily addressed to developers and deployers of AI systems.

Following publication of the Guidelines, the European Commission will engage in a piloting process during summer 2019 to gather feedback, and the High-Level Expert Group on AI will review the assessment lists for the key requirements in early 2020.

- High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 8 April 2019

http://merlin.obs.coe.int/redirect.php?id=19552

EN

- Council of Europe Newsroom, Foreign Ministers: towards a legal framework for artificial intelligence, 17 May 2019

http://merlin.obs.coe.int/redirect.php?id=19553

EN

Ronan Ó Fathaigh

Institute for Information Law (IViR), University of Amsterdam

The objective of IRIS is to publish information on legal and law-related policy developments that are relevant to the European audiovisual sector. Despite our efforts to ensure the accuracy of the content, the ultimate responsibility for the truthfulness of the facts on which we report is with the authors of the articles. Any opinions expressed in the articles are personal and should in no way be interpreted as representing the views of any organisations represented in its editorial board.