

EU Draft AI Guidelines

ARTICLE 19 written submissions on the European Commission's High Level Group on Artificial Intelligence

ARTICLE 19 is an independent human rights organization that works around the world to protect and promote the rights to freedom of expression and access to information. Among others, the work of ARTICLE 19 focuses on the nexus between human rights and technology. As such, we actively contribute to the debate about how to regulate Artificial Intelligence by providing inputs, policy guidance, and by participating in various fora across the Internet governance and standards development landscape.

Introduction: Rationale and Foresight of the Guidelines

- ARTICLE 19 contests the major role that the draft attributes to ethic and urges to replace it with an approach focusing on the protection of human rights. In the section dedicated to “The Role of AI Ethics”, the draft states: “The goal of AI ethics is to identify how AI can advance or raise concerns to the good life of individuals, whether this be in terms of quality of life, mental autonomy or freedom to live in a democratic society.” ARTICLE 19 believes ethics is insufficient to protect individuals from the harm that AI can inflict on them. Therefore, the role of ethics is minor, and certainly less important than the role that legal provisions shall have for this purpose.
- The scope of the guidelines does not include an awareness of dual-use AI technologies. However, in many instances this kind of technologies raises tremendous challenges for individuals' fundamental rights. Various countries have already in place norms regulating the import/export of these technologies, and some companies are already calling for States to provide regulatory guidance concerning development. ARTICLE 19 believes that the guidelines shall duly consider dual-use AI explicitly and establish rules aimed at creating safeguards and guarantees for individuals' rights.

Chapter I: Respecting Fundamental Rights, Principles and Values - Ethical Purpose

- ARTICLE 19 that the session on “Fundamental Rights and Human Beings” could be more efficient if an explicit mention on specific Articles of the Treaties and of the Charter would be included where relevant. This will enhance legal certainty and facilitate the identification of relevant case law where needed. In particular, section 3.2 could include explicit reference to the rights and freedoms that compose the

- “freedom of individual”, such as, for example, freedom of expression, privacy, non-discrimination.
- Section 3.4 raises challenges. Various AI systems are behind intellectual property protections, how can one guarantee access to information and knowledge? ARTICLE 19 believes that sometime distinctions can be legitimate. In addition ARTICLE 19 notes that workers and consumers are not “minorities”.
 - ARTICLE 19 suggests that the citizens’ rights mentioned in section 3.5 shall be equally guaranteed where data are used for automated decision making by both States and private actors.
 - Session 4 on “Ethical Principles in the Context of AI and Correlating Values” contains a list of high level principles: beneficence, non-maleficence, autonomy, justice, explicability. The majority of these principles have been already established and made applicable by different branches of EU law, such as the General Data Protection Regulation. ARTICLE 19 believes that a vague reference to them risks undermining their legal status, diminishing legal certainty and suggesting somehow that the enforcement of these principles depends on the inclinations or goodwill of companies that develop AI. Instead, reference should be made to the relevant rules that guarantee these principles and/or implement them.
 - Session 5.1 “Identification without Consent” states: “Differentiating between the identification of an individual vs. the tracing and tracking of an individual, and between targeted surveillance and mass surveillance, will be crucial for the achievement of Trustworthy AI.”. ARTICLE 19 calls for a deeper analysis of the distinctions, as well as of the safeguards and the possible shortcomings. More in general, ARTICLE 19 notes that AI identification technologies interfere with privacy, right to anonymity and potentially freedom of expression of individuals, and recalls that the interference shall be subject to the three-party test of legality, proportionality and necessity.
 - ARTICLE 19 calls for the inclusion of the right to know when one is subject to a decision/interaction with AI in session 5.2.

Chapter II: Realizing Trustworthy AI

- ARTICLE 19 suggests that section 1 should focus on how to establish accountability, and on who is to be held accountable, rather than on possible remedies. Possible remedies could be suggested in a separate session to be added.
- With regard to section 2, dealing with data governance, ARTICLE 19 warns on the fact that the efforts to combat bias shall in no case lead to the violation of fundamental rights. In addition, on issues concerning collection and purpose limitation, ARTICLE 19 recommends to insert references to the relevant provisions of the General Data Protection Regulation.

- Concerning section 3, ARTICLE 19 notes that not all systems are, or should be, intended for all. ARTICLE 19 therefore suggests having a qualifier, which makes explicit that the State-deployed systems intending to serve all should allow all citizens to use the products or services, regardless of their age, disability status or social status, or better, in a way that is not discriminatory to protected attributes.
- The section dedicated to good governance shall include remedies, redress mechanisms and due process guarantees. ARTICLE 19 also considers that guidance about which kind of remedies companies could design would be useful in providing legal certainty.
- The section dedicated to “Traceability and Auditability” (p.22), both elements are presented as a mere option. ARTICLE 19 believes this call shall be reinforced. In addition, ARTICLE 19 suggests to add interpretability among the parameters.
- In the section dedicated to “Regulation”, ARTICLE 19 calls for an explicit reference to due process when dealing with redress and remedies.
- ARTICLE 19 believes that the section dealing with “Standardisation”, shall contain the recommendation that standard setting bodies include human rights impact assessment in their considerations.

Chapter III: Assessing Trustworthy AI

- No comments on this part.

General Comments

- The guidelines focus on ethics, rather than on fundamental rights and existing legal frameworks. The draft suggests a new concept of "ethical purpose" that should include fundamental rights, principles and values. Nevertheless, ethics and law remain two different concepts. ARTICLE 19 believes there is no added value in introducing a nebulous concept like “ethical purpose” which, on the contrary, appears to create confusion and undermine existing legal rights and duties.
- The draft explicitly takes into account privacy, non-discrimination and human autonomy, and dedicate to them specific provisions. Nevertheless, there are other fundamental rights, such as freedom of expression that might be strongly affected by AI. ARTICLE 19 calls for an explicit mention that all fundamental rights affected by AI deserve adequate protection.
- The draft guidelines are based on the concept of “Trustworthy AI”. ARTICLE 19 notes that trust is a relationship between peers, where the trusting party, while not knowing

for certain what the trusted party will do, believes any promises being made. Therefore, one cannot trust AI, and AI cannot be trustworthy. For AI, accountability has to be used. AI can be accountable. Trustworthiness should be for institutions that hold AI accountable.

- The text repeatedly refers to “data subjects”. ARTICLE 19 urges the HLEG to remember that data subjects are human beings with all their rights as guaranteed by EU Charter. Therefore, when discussing about safeguards for data subject, the approach has always to be: do the planned legal provision/remedy adequately guarantee the protection of people’s fundamental rights?
- In the executive summary, it states that “...on the whole, AI’s benefits outweigh its risks, (...)”. This assumption is not proven, and it should not be taken as a reason to relax the approach on regulating AI.
- The draft contains numerous expressions which indicate that compliance with what prescribed or recommended is left to the goodwill of AI developers. This interpretation is corroborated by the focus on ethics rather than on legal obligations. Examples are, among many: “Moreover, keep those requirements in mind when building the team to work on the system, the system itself, the testing environment and the potential applications of the system” (p.3). “Keep those requirements in mind” is a weak expression that does not create any duty on the developers. “Strive to **facilitate the auditability** of AI systems, particularly in critical contexts or situations.” (p.3). “Strive to” guarantees too large margin of manouvre for developers. “Be mindful that there might be fundamental tensions between different objectives” (p.3). Being mindful does not imply the requirement for any specific action. The draft suggests to “Communicate and document these trade-offs”: however, it is not clear to whom and why they should be communicated, nor how they should be documented. All in all, the softness of this language undermines legal certainty. ARTICLE 19 urges to opt for a stronger approach and to avoid vagueness around what companies are called to do, not to do, or to comply with.
- The draft states “This guidance forms part of a vision embracing a human-centric approach to Artificial Intelligence, which will enable Europe to become a globally leading innovator in ethical, secure and cutting-edge AI.” (p.3). ARTICLE 19 strongly believes that Europe shall commit to lead on a human rights friendly AI, not on an ethical one. On this regard, ARTICLE 19 recalls that the EU Charter of fundamental rights establishes the obligation, for the EU institutions as well as for the member States when applying EU law, to guarantee full respect of the fundamental rights listed therein. This remains valid concerning any actions that EU institutions undertake with regard to AI.
- The concept of “human-centric” approach is based on human values (p. 4). ARTICLE 19 urges this concept to be based on human rights instead, and to therefore take into due account the existing international legal framework to protect them.