ETNO Response - Stakeholders' Consultation on Draft AI Ethics Guidelines

January 2019

Introduction: Rationale and Foresight of the Guidelines

ETNO welcomes the draft "Ethics Guidelines for Trustworthy AI" launched by the European Commission's High-Level Expert Group (HLEG) on AI.

We are delighted that the draft AI ethics guidelines place European citizens at the heart of AI development and use ("human-centric" AI), respecting fundamental rights, applicable regulation, and core principles that underpin the ethical purpose for AI. Several ETNO members have made public commitments to ethical principles governing the development and use of AI technologies. Our members have robust data governance programs, whose policies and procedures are also generally applicable to uses of data in AI applications and solutions.

We support the guidelines' vision to create a culture of "Trustworthy AI made in Europe", which will not only protect and benefit individuals and the common good, but also enable Europe to become a globally leading innovator in AI, as it will generate user trust and facilitate AI's uptake. The establishment of a European approach to AI to foster competitiveness in the EU should be particularly emphasised. European values, enshrined in digital ethics, can represent a competitive advantage for the development of Trustworthy AI. Our understanding is that the guidelines are intended to help seize this potential.

ETNO fully agrees with the acknowledgment that "no legal vacuum currently exists, as Europe already has regulation in place that applies to AI" and that the guidelines will not imply any form of regulatory intervention. The development, deployment and use of AI technologies are subject to a robust horizontal (and, in some areas like privacy, sector-specific) legislative framework that protects the fundamental rights and integrity of European citizens. Tightening the existing legal framework could stifle the European AI ecosystem rather than nurturing it, and ultimately let other regions of the word like China and the United States dictate the rules of the game.

Nevertheless, we recognise that some elements of the existing framework (e.g. cybersecurity) may need adjusting to the new challenges brought by AI. In this respect, ETNO agrees with the statement that "different situations raise different challenges". Different AI-based systems may have a different impact on the rights of individuals at any stage of their life cycle. Therefore, we recommend embedding a clear "risk-based approach" in the guidelines and any possible future initiatives on AI, recognising that the requirements and methods for achieving Trustworthy AI should vary depending on the specific AI system's application.

We also encourage the HLEG to ensure that the document does not contradict EU law by introducing novel terminology or by reinterpreting specific, well-established legal concepts and obligations especially related to the General Data Protection Regulation (GDPR). Furthermore, we also suggest that the guidelines clarify what terms like "wellbeing and the common good" mean according to the EU understanding based on fundamental rights.

Finally, although it is clear that the guidelines will be voluntary and non-binding, it is less clear what the practical implication of their formal endorsement by stakeholders will be. Most notably, it is unclear whether any benefits or duties will be attached to the formal adoption of the guidelines, and how stakeholders' compliance with them will be scrutinized. It is also unclear how endorsing the guidelines will affect existing self-regulatory initiatives, such as guidelines and codes of conduct, already implemented by individual organizations. We ask the HLEG to elaborate in further detail on the concrete functioning and effects of the future mechanism for endorsement. This is crucial for ETNO, as many European telecom operators have already launched their own guidelines, manifestos, dedicated work streams or committees.

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

ETNO supports a fundamental rights-based approach to AI ethics, underpinned by the families of EU's citizen rights described in the document.

However, we have some remarks about the four identified principles that rest on fundamental rights (beneficence, non-maleficence, autonomy, and justice). Our main concerns are as follows:

- <u>The Principle of Beneficence: "Do Good"</u>

  We encourage the HLEG to recognise commercial uses of AI technology as legitimate and beneficial. AI applications that increase efficiency and productivity have real positive impacts on society. A narrow application of this principle bears the risk of restricting companies' freedom to innovate. It could have undue adverse effects on the innovation capabilities of economic actors whose primary mission is not necessarily to improve collective wellbeing. It would also cause uncertainty with regard to existing applications that pursue legitimate business goals, but that do not clearly contemplate the "Do Good" principle.

- <u>The Principle of Non maleficence: "Do no Harm"</u>

  Technology is a tool, not an end in itself. It is arguable whether a technology can be inherently "good" or "bad", or whether in principle all technologies can be regarded as ethically neutral and what determines their positive or harmful impact is their specific use. Therefore, any principles related to Good or Harm can only apply to the specific application and business model. Therefore, transparency regarding the application and business model of an AI system is more important than the transparency of that system's technological aspects.

- <u>The Principle of Autonomy: "Preserve Human Agency"</u>

  ETNO supports the principle of autonomy, noting that it should recognise that different uses of AI call for different degrees and types of autonomy. This principle is largely reflected in the GDPR, whereby data subjects have the right not to be subject to a decision based solely on automated processing (Art. 22) and have a right to object to most forms of processing of their

data at any time (Art. 21). It is then important that the principle of autonomy as described in the guidelines be consistent with the existing legal framework. For instance, footnote 13 could be interpreted as an extensive right to object to any AI-based data processing in the working environment, beyond the letter of Art. 88 GDPR on processing in the context of employment. The flexibility and balancing of interests inherent in the GDPR are a valuable reference in this context.

- The Principle of Justice: "Be Fair"

  Besides the concept of fairness and the importance of redress mechanisms and remedies (which are already provided for by the GDPR), we support the concept that human agents are ultimately responsible for AI-based decisions and their impacts on individual rights. Identifying the person(s) and/or role(s) responsible for a given system should be part of every developer and implementer "accountability" mechanisms.

- The Principle of Explicability: "Operate Transparently"

  We agree that AI systems should be as transparent as possible for users, to provide them an understanding of how decisions affecting them are taken. However, we recommend applying the proportionality and risk-based approach principles to explicability, whereby the degree of insights required would depend on the complexity of the system as well as on its impact on individuals' rights.

Furthermore, we would like to comment on the statements that "informed consent is a value needed to operationalise the principle of autonomy in practice" and that "in order to ensure that the principle of explicability and non-maleficence are achieved the requirement of informed consent should be sought". We would like to remind that, according to the GDPR, user consent is just one of six legal bases for processing personal data. With regards to automated decision-making, the data subject has a right to object when the processing produces legal effects or significantly affects him or her.

Therefore, we recommend not considering consent as the panacea to ensure the respect of the ethical principles at hand. Depending on the context, other legal basis may be equally or more suitable for ensuring transparency, explicability, non-maleficence and accountability of an AI system. Identification without consent per se is not unethical and does not automatically imply a threat for individuals.

Moreover, regarding "the usage of "anonymous" personal data that can be re-personalized", the potential for re-identification depends on the technical means used to anonymise or pseudonymise personal data as well as the way data is clustered, packaged and processed thereafter.

As to "Covert AI systems", we are not convinced that these systems represent a critical concern as such. That will depend upon the function the system provides and the context in which it operates. Appropriate transparency measures towards users of AI systems are key; nonetheless it should be considered that, in some context, individuals that know they are interacting with a machine will behave in a different way that hinder the objectives of the system (e.g. in medical research).

Finally, we believe that providing examples of potential longer-term concerns at this stage could be premature and could fuel unfounded worries. For instance, Artificial Moral Agents (AMAs) should not per se pose a threat as long as these have been trained within a given and acceptable ethical framework; on the contrary, AMAs might well be considered one of the few technology principles for developing ethical AI in practice. Additionally, whether self-improving Artificial General Intelligence (AGI) is possible is still a matter of speculation. These are subjects that could be considered in an eventual follow-up phase of discussions.

Chapter II: Realising Trustworthy AI

We think that this Chapter (and the guidelines more in general) should clarify where a distinction can be drawn between professional AI systems (i.e., used for businesses and public institutions) and consumer AI systems. The ethical frameworks and the measures to make a system Trustworthy may differ accordingly. There is a big difference between realising Trustworthy AI for a professional user (e.g., pilot, robotics operator, flight controller, etc.) and doing so for a regular person using an AI-based app for e.g., tax declaration or social security, though some applications could be less distinct, such as public sector use of AI in sentencing guidelines.

ETNO would like to raise some comments regarding the identified requirements of Trustworthy AI.

- Accountability: In our understanding, accountability goes far beyond redress and compensation for wrongdoings. Accountability is a much broader principle that requires an organisation to demonstrate respect of individuals' rights and compliance with applicable regulation and standards, as well as to be held responsible for its activities and their effects. Therefore, accountability mechanisms may include self-regulation instruments such as codes of conduct.
- Data Governance: Data governance is a broader concept than what is reflected in this requirement. An organisation's policies, procedures, data protection officers, and training programs related to the use of data should all be relevant when assessing its approach to Trustworthy AI. Furthermore, we agree on the importance of datasets quality, but we are concerned that pruning biases away before engaging in training may in fact cause other, unintended biases to emerge. It may be preferable to identify the biases in the datasets before training, but to correct them *ex post* after the processing of the datasets has occurred. Particular attention should be given to the practice of data labelling. We also have doubts about the description of how anonymisation should not hamper a proper division of datasets for training and test. Anonymisation is not per se linked to which data is used in training and test, as long as the same data is not used in both sets; two different pictures can easily be split so that one ends up in training and the other in testing, which has nothing to do with the process of anonymization. Finally, this section could elaborate on the legal grounds available for processing personal data.
- Design for all: We agree that AI systems should in principle be accessible by all citizens. We would also note that some AI-based products and services may target one or some specific groups (e.g., age-specific or gender-specific) while not barring everyone else from technically accessing that system. "Positive discrimination" is not automatically in contradiction with this requirement; for instance, an AI-based product may be specifically designed for disabled people.
- Governance of AI Autonomy (Human oversight): We welcome the risk-based approach attached to this principle. We believe that a clear designation and communication of the person(s) and/or role(s) responsible for a given system should be a key part of good governance.
- Non-discrimination: As already mentioned, positive discrimination is not necessarily unethical and may even be necessary to reach an objective. For example, medical researchers may need to study a component of the population that have specific characteristics, and use AI to extrapolate this sample by excluding the rest of the population.
- Respect for (& Enhancement of) Human Autonomy: It may be difficult for an AI system to protect citizens from abuses by design. Systems should include processes to avoid their misuse, but it is very hard to prevent any governmental or business abuses that depend on the actual usage of the

technology.

- Respect for Privacy: We suggest that the guidelines highlight the importance of effective technical and organisational measures that mitigate the privacy risks for individuals, such as the "pseudonymisation" of personal data.
- Robustness: Security and resilience to attacks are fundamental prerequisite of robust AI systems. We suggest that the guidelines expand on what mechanism could be implemented to ensure high cybersecurity standards for AI systems (e.g., "security- by-design"). We recommend assessing the relevance of the regulatory framework for operators of critical infrastructure (i.e., Directive on security of network and information systems) for AI systems.
- Transparency: We reiterate that explainability should be guided by the principle of proportionality and the risk-based approach.

With regard to the technical and non-technical methods to achieve Trustworthy AI, we have the followings remarks about the technical methods described by the guidelines:

- Architectures for Trustworthy AI: Trustworthy AI should not only be ensured by "formulating rules, which control the behaviour of an intelligent agent, or as behaviour boundaries that must not be trespassed", but also through mechanisms enabling operators to deactivate and stop AI systems at any time.
- Traceability & Auditability: The meaning and the objectives of traceability and auditability for the purpose of these guidelines should be clarified, bearing in mind the context and the application (professional vs. consumer) of an AI system. Producers and developers of AI should keep track of the decisions made and the information fed to the system also in order to enhance the quality of decisions.
- Codes of Conduct: The headline is misleading, as there is more to ensuring an organisation's adherence to ethical principles than just codes of conduct. We suggest renaming the section "Corporate Governance".

Additionally, we would like to suggest further technical methods to achieve Trustworthy AI:

- Responsibility: As already mentioned, AI systems should have a responsible person or role that takes decisions regarding the system and monitors its operations. Responsibility should be present at every stage of the system's lifecycle.
- Pseudonymisation: Pseudonymisation of personal data enables data processing in a privacy-friendly manner but, contrary to full anonymisation, it preserves the necessary identifiers that allow to repeatedly merge large amounts of data from various sources over time while eliminating the direct link between data and data subject. The EU has embraced pseudonymisation as a privacy-friendly technique in the GDPR.

Chapter III: Assessing Trustworthy AI

We do not have specific comments to this Chapter.

General Comments

The draft guidelines mention that the HLEG will elaborate on four use cases in the final version of the document (Healthcare Diagnose and Treatment, Autonomous Driving/Moving, Insurance Premiums and Profiling and law enforcement). Even though telecommunication services are not contemplated in the list of use cases, ETNO would like to provide the HLEG with some elements describing the role of AI in our industry. We identify three main clusters of use cases enabled by AI:

1. Network Operations: As providers scale up their infrastructure by adopting network virtualization, software defined networks, cloud-based applications and 5G, AI becomes particularly crucial for efficiently operating the network. Network security and predictive maintenance of networks are just two of the most important use cases enabled by AI.
2. Customer Relationship: AI is key for enhancing CRM and customer experience. As much as many other sectors, telecoms are increasingly using customer service applications that rely on chat bots, virtual assistants, and personalized content and offerings in real time.
3. New Products: AI systems are important for the development of new, data-driven services and products. Several telcos are investing in the creation of "platform ecosystems" for their clients, largely powered by AI. An example is data management platforms offered by telcos, where their customers can store, share and use data in a secure and privacy-protective manner.

On a separate note, we would like to comment on the definition of AI that is provided in the addendum to the guidelines. This definition does not seem fully accurate and would deserve further consideration. For instance, the goal an AI system is tasked with meeting may not be necessarily complex and an AI system is not necessarily designed by a human, but by another machine.

We recommend that a revised definition of AI features the following criteria:

- exclude software systems based on traditional and determined algorithms that are clearly not based on AI;
- take into account that the AI algorithm takes decisions as a consequence of the application of advanced analytical techniques (i.e., machine learning and deep learning) to solve problems;
- require strict ethical scrutiny of an AI system only when its purpose may constitute a risk to individuals' fundamental rights (risk-based approach).