February 1, 2019

ITI Recommendations on the EU High Level Expert Group Artificial Intelligence Draft Guidelines

Introduction: Rationale and Foresight of the Guidelines

The Information Technology Industry Council (ITI) is the global voice of the tech sector. ITI members are global companies with complex supply chains around the world, we understand the importance of consumers being able to use technology seamlessly across borders every day. As both producers and users of privacy protecting and enhancing products, technologies and services, ITI hopes to see the most effective approaches applied to promote the beneficial development and use of artificial intelligence (AI) globally.

General Comments

We thank the European Commission High level Expert Group (HLEG) for their work on these guidelines and for taking a constructive approach that focuses on practical suggestions that AI developers and users can benefit from. We are happy to see that the Commission and the HLEG sees AI as a net positive for society and support the view that AI cannot be approached with one-size-fits all rules and prescriptions - but requires a constant discussion and iteration. We hope the HLEG will continue working this understanding into future steps of the work plan. Our general recommendation regarding the guidelines is that elements of it would benefit from further collaborative development, so they can be implemented and offer real world value. We have pointed out these areas and offered suggestions in the comments that follow.

In addition, as AI is not developed in regional siloes, we recommend avoiding references to 'AI made in Europe', which are in contradiction with the global perspective the European Commission has endorsed. Products and services are the combination of components developed in different locations and are part of a global ecosystem. Many AI tools can be accessed via cloud computing and will work in combination with European and non-European elements.

The Commission and the HLEG should aim to promote the ethical development & use of AI globally via collaborative engagement with its international partners. To realise ethical AI, the HLEG should consider the inclusion of a section on global governance. AI and technology as a whole are often built and applied across borders. They are part of an ecosystem, where different components might stem from different regions in the world. For this reason, the EU must

Global Headquarters
1101 K Street NW, Suite 610
Washington, D.C. 20005, USA
+1 202-737-8888

Europe Office
168 Avenue de Cortenbergh
1000 Brussels, Belgium
0032 (0)2 380 7764

info@itic.org
itic.org

maintain a dialogue with other geographies when it comes to the responsible development of AI. The most reliable way for Europe to ensure trustworthy AI for its citizens is to collaborate to promote a shared understanding and common norms across geographies. Europe should not miss the opportunity to shape the global debate on AI governance.

Glossary and Executive Summary

We appreciate that the HLEG notes the incomplete nature of the glossary (pg. iv) and expects to further complement it. Below, we lay out our views about some of the current definitions.

Artificial Intelligence (AI)

We find that the definition for AI (pg. iv) is not in line with that generally used by the community of AI practitioners and the current state of art in AI technology. In particular, the statement that AI systems are "designed by humans" and are "deciding the best actions to take (according to pre-defined parameters) to achieve a given goal" appears outdated and ignores the existence of machine learning systems that are in fact not completely pre-defined by humans.

There is a discrepancy in the definition of 'trustworthy AI' in the glossary section and the one used in the executive summary. We find the one in the glossary section is superior in that it makes clear that fundamental rights and regulations should be complied with during the development, deployment and use of AI (it is not the AI system itself that does all these things). We also endorse the acknowledgment that no legal vacuum currently exists, as Europe already has regulation in place that applies to AI. We believe building trust also means demystifying some of the unfounded concerns around the technology and educating the public on what AI is and how it can be used, and we recommend including these points.

Bias

The definition of bias (pg. iv) is not aligned with its actual scientific meaning in statistics, instead overly focusing on the human element. It also overstates the risks compared to the advantages of AI, only once mentioning the potential for AI systems to support less biased decisions. We advise the following changes:

"*Bias is a prejudice for or against something or somebody, that may result in unfair decisions. It is known that humans are biased in their decision making* ***and that unfair bias permeates our societies****. Since AI systems are designed by humans* ***and rely on data****, it is possible* ***that their results are biased*** *even in an unintended way. Many current AI systems are based on machine learning data-driven techniques. Therefore, bias can* ***manifest itself*** *in the collection and selection of training data. If the training data is not inclusive and balanced enough, the system could learn to make unfair decisions. At the same time, AI can help humans to identify their biases, and assist them in making less-biased decisions.*"

Endorsement Mechanism

We commend the HLEG for acknowledging that a domain-specific ethics code – however consistent, developed, and fine-grained future versions of it may be – can never function as a substitute for ethical reasoning itself, which must always remain sensitive to contextual and implementational factors, and that different situations raise different challenges. Given the Guidelines' voluntary nature, we urge the HLEG to clarify the meaning and implications of the formal stakeholder endorsement process (pg. 2), as the current understanding may suggest some form of legal compliance and raises the question whether the guidelines may subsequently be referenced elsewhere (e.g. endorsement as a requirement in procurement procedures).

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

Voluntary Rights Based Approach

We are concerned that throughout this chapter, the voluntary nature of the Guidelines is not properly reflected. In particular it speaks of "**governing** the "ethical purpose" (pg. 5) and "identifies the **requirements** for trustworthy AI" (pg. 8) rather than providing guidance. There are also various instances where it is insinuated that the technology sector has a negative attitude towards their customers or individuals. The paragraph on respect for human dignity, for example, suggests that it be a requirement that "people be treated with respect due to them as individuals, **rather than merely as data subjects**" (pg. 9). This is not a fair representation and ignores that AI is actually used and developed by our companies to better fulfil individual needs.

Finally, the language used in this chapter often overlooks that AI systems do not only "hold the potential to improve the scale and efficiency of government […] services" but they "**are already improving**" them.

Implementing Measures

When discussing the Principle of Autonomy to "Preserve Human Agency" (pg. 9), a right to opt out and a right of withdrawal are contemplated, but not qualified according to the use case. These rights need to be qualified for instances where opting out might cause harm to others or prevent an authority from performing its duties for the common good. such a 'right' can't be horizontal - it must vary according to the use cases.

Similarly, when contemplating citizens' "**right to be informed of any automated treatment of their data** by government bodies" and "**systematically be offered to express opt out**" (pg. 7) the Guidelines overlook that most, if not all, government service provision will in future entail some degree of automatic processing of data and it is unclear how citizens would be informed of each of these. Secondly, and of greater concern, is that it is entirely unclear how an opt-out would work in practice. Would citizens, for example, have the right to have their

tax declarations or social benefits allocations checked manually? This would undermine any smart governance systems and may result in uneven (and unfair) distribution or provision of public goods and services, not only for the citizens who opted out but also for those who did not.

Furthermore, the Guidelines discuss rights associated with "direct or indirect" AI decision making- "a right to knowledge of direct or indirect interaction with AI systems, a right to opt out and a right of withdrawal". How this would work in the real world, even with today's use of technology, is entirely unclear. Ultimately, the only choice for the individual may be to not use the service at all. A similar lack of clarity exists around what is envisioned as being "**effective redress** if harm occurs" (pg. 10), and around why this should be singled out from a general right to redress.

In addition, while measured transparency is indeed a key element in creating trust in AI systems, the Guidelines require "AI systems [be] intelligible by human beings at varying levels of comprehension and expertise" (pg. 10). This should only be required in **qualified cases depending on the application and based on agreed procedures**.

The Guidelines also require "both technological and business model transparency" (pg. 10). Not only is the concept of 'business model transparency' unprecedented, it is also unworkable since business models change over time and a system developed for one purpose could end up being used for another. These Guidelines should also make explicit that they similarly do not aim to imply disclosure of source code or any other information that would threaten industrial property or trade secrets.

The Guidelines also refer to 'informed consent' (pg. 11), with reference to the GDPR though its meaning in this context is not clarified. We suggest the definition be specified, but with the additional consideration that GDPR allows data processing based on reasons other than consent, like legitimate interest. In many instances, a blanket right to refuse being subject to AI technology, would neither be possible nor desirable as it could go against the benefit of the user, against the rights of others or impede the functioning of public institutions. We suggest removing this concept altogether and replacing with the focus of these guidelines: trust.

Finally, when contemplating Ethical AI, purpose and context go hand in hand. One must also be mindful of what practices are harmful and not harmful, lawful and unlawful. For example, the section on Identification and Consent (pg. 11) does not consider that not all identification processes create a danger for the individual and many are actually beneficial.

Chapter II: Realizing Trustworthy AI

The HLEG notes that an agreement had not been reached on operationalizing the principles laid out in the first section and therefore sought input from the consultation. While the high-level principles and values laid out in the previous chapter are uncontroversial in and of

themselves, we are concerned that they are not sufficient and further developed thinking is required on realising the goals stated therein. Many of our members have devoted thought and resources to developing more comprehensive internal guidelines, built specifically for developing and deploying intelligent systems.

Accountability

This chapter aims to offer guidance on implementing trustworthy AI, but the description of "accountability" (pg. 14) is extremely narrow. It excludes preventative and systemic accountability processes within organisations developing or deploying AI systems, i.e. measures to ensure things do not go wrong in the first place and that, if something goes wrong, there is a procedure to follow and people in charge to address issues. Accountability should also include providing to the ability to contest AI output and provide feedback on why a certain output is right/wrong.

Data Governance

The data governance section is silent on many established best practices in data governance and handling, and instead is largely focused on the quality of data sets. Given this, this section should be retitled "data quality and governance". It should also devote some attention to the traceability of data sources, how data undergoes transformation, and maintaining documentation on the quality and nature of data, including considerations of potential re-identification of individuals. This section assumes that biases can be "**pruned away before engaging in training**" (pg. 14) while this may not always be possible and contradicts a later assertion that "data always carries some kind of bias". Machine bias can be introduced at various stages, be it due to characteristics of the AI's connectivity, the system's technical architecture or design, or through training bias (which is impossible to eliminate completely). Thus, bias cannot be removed or prevented, but it can be assessed, documented, mitigated and/or disclosed. One key consideration in implementing trustworthy AI is making this information available in a meaningful manner and determining what level of bias is acceptable for which application. We, therefore, suggest reframing to say that datasets inevitably contain biases, and one has to prune these away **to the maximum extent possible** before engaging in training. We also caution that suggestions to "always keep record of the data that is fed to the AI systems" (pg. 15) may not be always compatible with EU data protection laws.

Governance of AI Autonomy

We commend the HLEG for their balanced approach in recognizing that assuring properties such as safety, accuracy, adaptability, privacy, explicability, compliance with the rule of law and ethical conformity heavily depends on specific details of the AI system, its area of application, its level of impact on individuals, communities or society and its level of autonomy (pg. 15).

Respect for (& Enhancement of) Human Autonomy

The autonomy principle is discussed at length in the fundamental rights section and is a much more expansive concept than what the draft has aimed to operationalize here. As discussed

in Chapter I, human autonomy comes into play in many more contexts than B2C personalization online. Personalization is not only more complex, but also this kind of personalization could in fact augment human autonomy, rather than compromise it. In the right kinds of applications, AI could enable people to exercise much more precise preferences than would be otherwise practically feasible. We suggest the HLEG refer to the autonomy principle as discussed in Chapter 1 and simplify this section to say, "systems that are tasked to help the user, must respect their right to human determination, ensuring that the overall wellbeing of the user as explicitly defined by the user her/himself is central to system functionality" (pg. 17).

Respect for Privacy

The section on "respect of privacy" appears underdeveloped given the importance of this subject, and portrays data controllers as nefarious actors looking to "take advantage" (pg. 17). This is unwarranted, since as the HLEG has acknowledged, pre-existing regulations including the GDPR provide robust protections in this area.

Robustness

The section on "robustness and accuracy" (pg. 17) should further emphasize maintaining transparency about the level of confidence with which predictions are made or the level of uncertainty involved in those predictions.

Transparency

The "transparency" section overly focuses on the perception that one has to look into the "black box" (pg. 18). Given that this type of transparency may not always be possible due to the complexity of systems or their nature (e.g. self-learning systems), it is important to focus on the input and, even more, on the **output stage** to foster transparency.

We also suggest the draft modulate the requirement of providing information about decisions concerning data sources, development processes, and stakeholders depending on the **impact** of the model on human beings. We also note that the term "human data" (pg. 18) is unclear and leaves little room for the nuance encouraged elsewhere in the guidelines.

Traceability and Auditability

The "traceability & auditability" section (pg. 20) provides no parameters for what "transparent" and "understandable" could mean, which is key to having practicable guidance.

Standardization

The "standardization" (pg. 21) section appears to overly focus on standardizing the design of AI systems, rather than APIs and interfaces. It is unclear what the desired goal is for this kind of standardization, especially given that the nature of AI makes it difficult to imagine a horizontal standard that would be meaningful across applications and sectors.

Codes of Conduct

We suggest broadening this section (pg. 22) to include other modes of self-regulation.

Chapter III: Assessing Trustworthy AI

In absence of a generalized model, the suggested approach to developing a set of guidelines involves assessing one use case at a time. We are generally supportive of this contextual approach, though some of our members have advanced a generalized model for assessing AI systems (independent of their nature) organized around types of biases. They have identified three types of biases that are detectable during assessment of any AI system that will lead to untrustworthy AI system operations: 1) biases related to data used to build the intelligence of an AI system, 2) human biases injected into knowledge bases gathered and used by an AI system, and finally 3) biases related to an AI system learning from another AI system. The HLEG might consider this level of abstraction in assessing AI systems.

* * *