CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY

Stakeholders' Consultation

Comments on the "Draft Ethics Guidelines for Trustworthy AI" by the High-Level Expert Group on Artificial Intelligence.

The European Commission appointed the High-Level Expert Group on Artificial Intelligence (AI HLEG). The AI HLEG has the objective to support the implementation of the European strategy on Artificial Intelligence. This will include the elaboration of recommendations on future-related policy development and on ethical, legal and societal issues related to AI. In January 2019, the Commission asked stakeholders for comments on the AI HLEG's "Draft Ethics Guidelines for Trustworthy AI". CISPA submitted the following comments and remarks in the Stakeholders' Consultation.

We welcome the European Commission's efforts to assess the transformative societal effects of Artificial Intelligence. The establishment of the High-Level Expert Group on Artificial Intelligence (AI HLEG) and its recommendations are the crucial starting point for the discussion on "Trustworthy AI made in Europe". We very much welcome the "Draft Ethics Guidelines for Trustworthy AI" (draft hereafter) and hope that the recommendations will fuel stakeholder engagement.

While we welcome the draft, we would like to respectfully comment and discuss the draft's scope and results drawing on our expertise in the areas of information security and data protection regulation and technology.

Introduction: Rationale and Foresight of the Guidelines

The aim of the draft is to outline a human-centric approach to AI by ensuring the ethical purpose of AI as well as its technical reliability and robustness. We agree with and support these two main areas of focus for the document and appreciate that the AI HLEG sees the draft as a "living document that needs to be regularly updated over time to ensure continuous relevance". While the draft is intended to foster reflection and discussion, we are concerned that the document lacks **incentives** for stakeholders developing, deploying or using AI to practically apply the draft's recommendations. As all recommendations are referred to as voluntary and suggestions, the draft's impact is unclear.

We understand that the AI HLEG will address questions of policymaking and potential regulation in its second draft (due in May 2019). While the draft states that respecting fundamental rights and complying with applicable regulation are a prerequisite for the AI's ethical purpose, the draft does not touch on the actual **applicable regulation**. The draft points out that "it should be noted that no legal vacuum currently exists, as Europe already has regulation in place that applies to AI", but unfortunately does not elaborate on what this regulation encompasses with regard to AI. Many of the non-committal guidelines in the draft are actual hard and enforceable legal requirements by European and national law. We are concerned that the draft might lead to the impression that the design, application and use of AI in Europe is mostly unregulated when that is far from the case. We respectfully suggest for the AI HLEG to consult with experts in the field of EU data protection and technology regulation to ensure that applicable legal requirements (such as Art. 22 and 25 para. 1 GDPR) are reflected in the guidelines.

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

On page 5 the AI HLEG introduces a rights-based approach to AI ethics with the "additional benefit of limiting **regulatory uncertainty**". What would constitute this regulatory uncertainty is not clarified in the draft. We consider it of utmost importance to take stock of the existing regulation and drafts of upcoming regulation to assess and explain for which application fields and to what extent missing regulation or "regulatory uncertainty" actually exist.

Also on page 5, the draft introduces the principle of autonomy as a core example of a fundamental right derived principle leading to informed consent as a value. While this is a helpful example to understand the relationship between fundamental rights, principles and values, in the context of AI it could be misunderstood as informed consent being the primary or the only legal base for AI use. In recent years with advancing digitalisation newer regulation such as the GDPR have acknowledged that **informed consent** (meaning knowing and understanding all consequences of data processing) has become less and less realistic in many circumstances of complex technologies. Art. 6 GDPR mentions informed consent as only one of several legal bases for data processing. Therefore, more paternalistic technology design regulation independent of consent has been introduced such as "Data Protection by Design" and the regulation of tracking technologies in the upcoming ePrivacy Regulation. Many AI application fields are equally or more complex and, hence, may not lend themselves to use cases for the primary of informed consent. We ask that the AI HLEG would discuss the suitability of autonomy (e.g., by informed consent) and mandatory technology design for different application fields considering the already existing regulation.

We appreciate the list of "families" of **fundamental rights** on page 7. However, we are surprised to not see Art. 7 ("Respect for private and family life") and Art. 8 ("Protection of personal data") of the Charter of Fundamental Rights of the European Union and Art. 8 ("Right to respect for private and family life") European Convention on Human Rights being mentioned as these have a crucial relation to data-driven AI technologies and are linked to several of the mentioned families of fundamental rights.

The derived **Ethical Principles** (page 8-10) are certainly helpful principles for designing AI systems. However, we are concerned that they are too vague and open to interpretation by those designing and operating the AI systems to being able to introduce meaningful rights and protections for the individuals subject to these AI systems. This is why existing regulation should not be out of scope for the draft. It would be beneficial, if the draft would also offer any guidance on how to address conflicting human rights or principles when designing or using AI systems. The focus on utilitarian arguments of collective good seems to be excessive considering the jurisprudence of the European Court of Justice and the European Court of Human Rights.

Critical concerns raised by AI

In 5.1 the draft raises concerns with regard to AI systems using biometric data. We would like to encourage the AI HLEG to reconsider the listed examples of **biometric data use** (listed are lie detection, micro expressions, voice profiling) as all of these face serious criticism from the scientific community as being not sufficiently based on evidence and scientific methodology.

This raises another concern for AI systems that is not yet addressed in section 5. The use of AI decision making based on not scientifically proven assumptions (correlation vs causality) or **pseudoscientific applications**. Risks to ethical AI should not only encompass the risk of being unjustly identified but also the risk of being subject to unethical AI systems using not scientifically recognised assumptions or mathematical-statistical methods.

We would also like to point out that targeted or mass surveillance for law enforcement purposes are not subject to the GDPR but to the Directive (EU) 2016/680 and its national implementation laws.

We are unaware whether the AI HLEG will issue a report on technical guidance. But since the section 5 of the draft mentions anonymisation (it warns against insufficient de-identification) we would like to suggest to include information on **Privacy Enhancing Technologies** (such as Differential Privacy, Private Learning and Federated Learning) and encourage AI developers and users to consider more privacy-friendly designs for Machine Learning.

We appreciate that the AI HLEG explicitly mentions increased risks in application scenarios with (informational, organisational, or legal) power asymmetries. This is why we would like to stress the need to discuss mandatory "**contestability**" in addition to transparency of AI decisions.

Chapter II: Realising Trustworthy AI

We strongly suggest to either add "**Security**" (meaning IT-Security) as an additional requirement or replace No. 8 "Robustness" with Security. From our point of view Robustness is a subcategory of IT-Security not the other way around. This is why the requirements under No. 8 miss crucial security requirements for AI: confidentiality and integrity (not only resilience to attacks), security against misuse by insiders, resilience and robustness against tempering with the learning process (adversarial learning), real-time guarantees, and intervenability.

We would like to suggest that the section on "**x-by-design**" approaches on page 19 explicitly encourage the use of Privacy Enhancing Technologies and a data protection by design approach (mandatory under Art. 25 para. 1 GDPR).

We think that it would be beneficial to extend No. 1 "**Accountabilty**" to also address accountability in complex distributed or federated processes.

Chapter III: Assessing Trustworthy AI

The proposed questions are a useful first step for self-assessment. It could be beneficial to advance them into a framework for an "**Ethics Impact Assessment**".

With regard to No. 8 Security/"Robustness" we would like to suggest to switch from asking about specific attacks to the more commonly used IT-security approach of defining the **attacker model** and its capabilities against whom you want to defend your system.

General Comments

We thank the AI HLEG for the opportunity to participate in the Stakeholders' Consultation.
While we very much welcome the "Draft Ethics Guidelines for Trustworthy AI", we are concerned that the draft gives a misleading impression with regard to the extent of existing regulation or legal uncertainty.

We are also concerned that IT-security requirements that enable lawful and ethical use in the draft are limited to robustness and therefore miss crucial security requirements.

We would like to encourage the AI HLEG to consult with experts in the field of EU data protection and technology regulation as well as IT-security experts to ensure that applicable legal requirements and the state of the art in secure system design are reflected in the guidelines and other upcoming documents.

Ninja Marnau

Senior Researcher
CISPA Helmholtz Center for Information Security