NORDE
NORSK RÅD FOR DIGITAL ETIKK

Response to the hearing on the European Commission's Proposal for AI Regulation

- On behalf of The Norwegian Council for Digital Ethics (NORDE)

AI regulation should enhance ethical aspects of AI systems. The Norwegian Council for Digital Ethics responds in this document to the national hearing regarding the Proposal for AI regulation (hereafter the Proposal). In our response to the hearing, we suggest that more attention should be given to formulations and definitions, in order to secure trustworthy AI.

To summarize, we propose that more attention is given to the following:

1) Best practice: Suggesting a supervisory authority at the European level - learning from the already implemented Norwegian Universal Design model.
2) Risk stratification: Incentives to seek the most cutting-edge innovation
3) Practical implementation: Securing diversity and anti-discrimination
4) Avoiding discrimination and securing safety, by clarifying definitions
5) Identifying the structures of power to protect the freedom of the individual

On the following pages, we elaborate on these issues. If you would like us to clarify or elaborate further on any of these issues, we are at your disposal.

Kind regards,

Norwegian Council for Digital Ethics (NORDE)
Leonora Onarheim Bergsjø
Diana Saplacan
Inga Strümke
Cathrine Bui
Ishita Barua

NORDE
NORSK RÅD FOR DIGITAL ETIKK

Response to the hearing on behalf of NORDE

1) Best practice: Suggesting a supervisory authority at the European level - learning from the already implemented Norwegian Universal Design model

A model for implementation in practice of the AI regulation proposal can be learned from the Norwegian model on the Equality and Anti-Discrimination Act and from the Norwegian Regulation on Universal Design of ICT (see regulations that legislate universal design of information and communication technology (ICT) in Norway). These have the Norwegian Digitalization Agency as a supervisory authority.

Similarly, at the European level, a supervisory authority who ensures that the regulation is followed should be established, which can assist with advice regarding the implementation of the regulation in practice. We suggest that such authority should be created not only at a national level since this might create conflicts of interest for the respective member state. One such example is the case of Ireland, where many of the tech-giants are based due to national tax privileges and regulations. Conflicts of interests should be avoided, in which an authority at the European level could assist.

2) Risk stratification: Incentives to seek the most cutting-edge innovation

The Proposal is based on a risk-stratification system for different AI systems, with different regulations for the different risk categories. The regulation could also benefit from clarifying what the potential gain would be if the risk is significantly reduced or eliminated, and guidelines to follow in order to reduce the risk. **Developers and manufacturers of AI systems should be given an incentive to seek the most cutting-edge innovation without being stifled in their endeavor.**

Otherwise, the risk-stratification could end up mainly keeping the innovators and product developers from making products that would fall into this category, to avoid regulation and red tape. In medicine, **many high-risk AI systems are also the ones that could result in the most favorable clinical outcomes**. For instance, robot-assisted surgery, which will most likely be deemed high-risk, can result in both life-threatening and life-saving outcomes, depending on the system's success.

Clinical decision support (CDS) is currently not deemed high-risk, but most such systems will likely be re-categorized as high-risk within the new framework. Clinical decision support is another AI system with great life-saving potential, allowing clinicians to make better informed clinical decisions. Categorizing it as high-risk, will cause the need for earmarked resources to reduce the risk. **Added focus on risk-reducing steps, alongside the risk-stratification, can serve to avoid stifling the development of important medical devices for many struggling companies that would otherwise take the risk of those innovations.**

3) Practical implementation: Securing diversity and anti-discrimination

The diversity aspect

We recommend added focus on issues regarding anti-discrimination and diversity. For instance, AI systems using language for communication, e.g. chatbots, and personal voice- and service assistants, have proven difficult for natural persons to use if they don't understand which *keywords* to use in order to communicate efficiently with the AI-based system (see Verne et al, 2021). This represents a potential challenge for natural persons with low digital literacy. This is especially important when integrating such systems into public systems or those intended to be used by the general public.

Vulnerable groups and consent to automated data collection

The regulatory aspects regarding vulnerable groups and consent to automated data collection are unclear. For example, AI-based systems are part of "care" (physical) robots to be used in home and healthcare contexts. Such robots are often equipped with different sensors, actuators, and cameras. If such a robot is placed, e.g., in the home of a care recipient, mechanisms limiting its data collection to the care recipient and no one else, should be in place. The regulatory aspects regarding how the robot differentiates among the care receiver, formal and informal caregivers, visitors, or children that did not consent to automated data collection, are unclear.

On responsibility and ownership of data, security, safety, and privacy issues

The Proposal should to a greater extent discuss and define accountability. Returning to the example of (physical) care robots, e.g. robot assistants, these have both software and hardware components, which are likely produced by different manufacturers. The distributor, provider, and user can also be different legal instances. Who is responsible for potential harm caused by the robot, and who owns the data it collects? Both researchers, companies and end users could benefit greatly from further clarifications.

4) Avoiding discrimination and securing safety by clarifying definitions

Aspects that were left open for interpretation by the GDPR, and that have not been clarified by established practice, such as what is meant by an 'explanation' of AI systems, should be clarified in the proposed regulation.

On the risk of intersectional discrimination

Instead of listing protected groups explicitly (gender and ethnicity), and thereby leaving out other current or future groups that require protection (e.g. LGBTQ+), **the regulation should refer to the body or regulation that defines, updates, and lists current protected groups.** An introduction to intersectional discrimination can be read here.

In addition, **the regulation should clearly define and address the kinds of discrimination that are made possible by AI systems, e.g. machine learning. Advanced data correlation-based models can give rise to intersectional discrimination, and thus circumvent regulatory requirements for protecting specific groups**.

AI systems can also derive protected attributes, meaning that excluding these from the data is not enough. The regulation should address this concretely enough for implementation.

The tension and balance between different fairness definitions should be addressed in the regulation. If AI-systems are required to be non-discriminatory against *all* races and *all* genders, or other groups, without differentiation between historically dominant and marginalized groups, this creates a mathematically impossible situation.

The Proposal does not define whether AI systems should aim to decrease historical inequalities, or whether preserving the current status is preferable. A study by Wachter, Mittelstadt, and Russell found that certain fairness metrics preserve the biases of the status quo, while others do not. The AI regulation should provide at least guidelines for such circumstances.

On the concept of safety

We suggest that the concept of safety in the context of AI should not be limited to physical safety, but also include mental/cognitive safety. This should be specified in the proposal, and the concept of safety clearly defined, or even re-defined, in the context of AI.

The Union should work in close cooperation with international standard bodies that aim to develop new standards regulating AI-based systems.

In the current version of the Proposal, it is not clear what kind of safety the current Proposal refers to.

- "Safety" is defined in the Proposal as: *"safety component of a product or system means a component of a product or of a system which fulfills a safety function for that product or system or the failure or malfunctioning of which endangers the health and safety of persons or property"*.
- The proposal also states that *"the extent to which the outcome produced with an AI system is easily reversible, whereby outcomes having an impact on the health or safety of persons shall not be considered as easily reversible"*.

Current standards, such as those used by the International Standards Organization, refer mostly to safety as *physical* safety (see ISO/DIS 21260 on safety of machinery, or ISO/TR 8124-9-2020 on safety of toys). Such a definition is inadequate for AI-based systems.

5) Identifying Structures of Power to Protect the Freedom of the Individual

We suggest that the Proposal go further in protecting individuals from the power of governments and big tech companies alike.

In its current form, there is a risk that the Proposal's intention to "increase people's trust in AI", will end up benefiting and protecting AI companies from public scrutiny. Modern democracy builds on the notion of freedom as “absence of arbitrary power to interfere”, and this is exactly the kind of freedom which might be threatened by non-transparent AI systems.

Assigning more responsibility to end-users is, according to our view, a mistake: Requiring, for instance, users to flag deep fakes on behalf of platforms implies that content platforms are not liable for spreading fake content or fake news.

For example, we argue that *notifying* users about emotion recognition systems is not sufficient. Opting out of platforms using AI systems today, would mean exclusion from important services and social arenas. Consequently, users should have the **right to opt-out** of AI functionalities without receiving negative repercussions from governmental institutions, and without having to opt-out of a platform entirely (e.g. opt out of addictive features on social media).

Moreover, the AI regulation should ensure that academic research is not skewed by companies that wish to obscure the damaging impacts of AI. A study by Abdalla and Abdalla found that the biggest AI actors are using the same lobbying tactics as when the tobacco industry tried to hide its negative health effects in the 1950-90s. AI companies and research institutions should be required to divulge funding relations, and audits need to expose practices for ‘ethics washing’.

We suggest that companies developing AI systems should be required to implement whistleblower protections that trump other confidentiality agreements, to ensure that employees are not fired for exposing unethical AI systems.

Last, but not least, we are deeply concerned that the proposed exceptions for law enforcement to use ‘real-time’ remote biometric identification systems in public spaces, might open up for full-time surveillance without the consent of citizens.

We in the Norwegian Council for Digital Ethics, thanks KMD for the opportunity to respond to this hearing.

Kind regards,

Norwegian Council for Digital Ethics (NORDE)
Leonora Onarheim Bergsjø
Diana Saplacan
Inga Strümke
Cathrine Bui
Ishita Barua