Mozilla's response to the European Commission's Public Consultation on AI

12 June, 2020

Mozilla is the Corporation behind the Firefox web browser and the Pocket "read-it-later" application; products that are used by hundreds of millions of individuals around the world. Mozilla's parent company is a not-for-profit foundation that focuses on fuelling the movement for a healthy internet. Finally, Mozilla is a global community of thousands of contributors and developers who work together to keep the internet open and accessible for all.

We are in alignment with the Commission's objective outlined in its strategy to develop a human-centric approach to AI in the EU. There is promise and the potential for new and cutting edge technologies that we often collectively refer to as "AI" to provide immense benefits and advancements to our societies, for instance through medicine and food production. But the challenge before the EU institutions is to create the space for AI innovation, while remaining cognisant of, and protecting against, the risks. Technological solutionism should be avoided: Problem definitions must be clear, risk should be assessed to encompass the spectrum of potential harms, safeguards must be established and/or strengthened, and any rules or regulations developed should be tailored to suit the specific harm or problem identified.

To that end, the EC's approach should be built around four key pillars:

- Accountability
- Scrutiny
- Documentation
- Contestability

Accountability means ensuring the regulatory framework is sufficient to protect against and mitigate the harms that may arise from certain applications of AI. That will likely involve developing new regulatory tools (such as the 'risk-based approach') as well as reinterpreting and enhancing the enforcement of existing relevant legislation (e.g. consumer protection law). **Scrutiny** means ensuring that individuals, researchers, and governments are empowered to understand and evaluate AI applications, and AI-enabled decisions - to the extent necessary and appropriate - through for instance, algorithmic inspection, auditing, and user-facing transparency. **Documentation** is closely linked to the above, and this principle means that the EU's approach should strive to ensure better awareness and tracking of AI deployment (especially in the public sector), and ensure that applications are developed to allow for documentation where necessary - such as human rights impact assessments in the product design phase, or government registries that map public sector AI deployment. Finally, **contestability** means ensuring that individuals and groups who are negatively impacted by specific AI applications have the ability to contest those impacts and seek redress e.g. through collective action.

The Commission's consultation focuses heavily on issues related to AI accountability. In what follows, we provide specific recommendations on how the Commission could better realise the principle of accountability in its upcoming work. Building on the consultation questions, we provide further insight on:

- ASSESSMENT OF APPLICABLE LEGISLATION
- ASSESSING AND MITIGATING "HIGH RISK" APPLICATIONS
- USE OF BIOMETRIC DATA

ASSESSMENT OF APPLICABLE LEGISLATION

In parallel with the focus on developing safeguards for high-risk applications, we welcome the ongoing mapping by the European Commission of the legal uncertainties and gaps of the existing legislation. We would in particular bid the Commission to take care when assessing the role and responsibility of "software", particularly in the context of the potential reform of the Product Liability Directive, due to its ubiquitous nature.

A systematic application of existing users' rights and a full, harmonised enforcement of the GDPR across the EU for instance may be fruitful in allaying many of the privacy concerns around collection and use of data raised by the current debate on AI (particularly Article 9 addressing use of biometric data, and Article 22 on algorithms). More guidance on the applicability of the EU data protection framework would be helpful here, and we strongly encourage the EU Commission and national Data Protection Authorities to ensure its effective application and enforcement.

Moreover, we encourage the Commission and relevant national authorities to take account of existing rights and protections afforded by the EU legislative *acquis* concerning discrimination, such as the Racial Equilty directive (2000/43/EC) and the Employment Equality directive (2006/54/EC). While these rules largely predate the development of contemporary AI applications, they may provide for protection and redress against discriminatory outcomes. A review of these legislative initiatives and guidance on how they could be applied in the present setting would be desirable. Similarly, the rich body of EU legislation surrounding consumer protection - particularly the Consumer Rights directive (2011/83/EU) and the Unfair Commercial Practices directive (2005/29/EC) - may provide an important baseline for addressing and mitigating harm and discriminatory outcomes in consumer-facing AI applications.

ASSESSING AND MITIGATING "HIGH RISK" APPLICATIONS

AI is an extremely broad concept, being used indifferently to refer among others to driverless cars, new methods for medical diagnoses, chat help services on consumer-facing websites, or content recommendation algorithms. Furthermore, the impossibility of anticipating the development of these technologies implies the need to assess risk broadly and mitigate it through a principles-based approach, and to ensure that safeguards are established to protect against the range of possible harms that may occur in various contexts, both on an individual and collective level.

We encourage the Commission to further develop (and/or clarify) its risk mitigation strategy as it is unclear to us how, by whom, and when risk is being assessed. Further clarifications below:

CONTEXT AND USE ARE CRITICAL COMPONENTS TO ASSESS RISK. Harm is often dependent on the *context* of application deployment (i.e. by whom, for what purpose, and directed at which groups). For example, a facial recognition algorithm, used by an individual to sort personal photos, raises entirely different issues than the same algorithm employed by a government to identify protestors. Similarly, a voice recognition tool used on an individual computer to help a person with disabilities raises different issues from the same tool used by the police to monitor phone calls en masse or capture conversations on the street. It is clear that any policy approach must seek to develop a culture of accountability, where safeguards ensure that the technologies developed and deployed comply with applicable laws and norms. Moreover, regulation must be sufficiently dynamic such that *specific harmful contexts* of deployment can be targeted and curtailed, such as to avoid overly general prohibitions that unintentionally impact benign deployments.

SAFEGUARDS MUST BE COMPREHENSIVE. We understand it is important to start somewhere, and thus looking at applications of AI that could cause the most potential harm should be addressed first. However, this approach should not overlook the spectrum of potential harms that must also be taken into account and safeguarded against. The EU should ensure that these applications do not "fall through the gaps", and should endeavour to address harmful outcomes in dedicated and tailored policy interventions.

REFINE WHAT IS MEANT BY "RISK". The Commission's white paper broadly identifies risk as comprising the domains of fundamental rights and user safety. However, this risk matrix needs significant elaboration, as it does not account for how different risk vectors may interact and be interdependent. Moreover, it is unclear how various risks are to be practically assessed, and how regulators and regulated entities should approach instances where measures to offset one risk may have the effect of heightening another risk . Furthermore "risk" could translate to a variety of harms or undesired outcomes, from bias, to opaque decision making, violations of privacy, and so on. Rarely can risks in these different domains be judged in like-for-like terms.

It must also be acknowledged that harm is unevenly distributed within societies. Populations that are marginalised, due for instance to social or economic class, race, gender, religious affiliation, people with disabilities, LGBTQ+, and so forth are more vulnerable to, and will be impacted far more than, those who are not in these categories.

Thus we strongly encourage further elaboration on this aspect in order to develop a comprehensive mitigation strategy (or strategies). Failure to do so could result in significant legal uncertainty for developers and companies on the one hand, and the possibility for "check-box" compliance that runs contrary to the spirit of the regulation on the other.

RISK BASED APPROACH IS ONLY ONE OF SEVERAL OPTIONS TO MITIGATE HARM. The current paradigm assumes that high risks can be mitigated, and that there are already established protections for low-risk applications. It may be that some applications and uses of AI technologies are simply incompatible with a "human-centric" AI ecosystem (see more in this article). It is therefore important to look not only at potential harms through the lens of "risks", as a comprehensive mitigation strategy must incorporate the spectrum of options, where "risk" may be appropriate in certain applications, and "red lines", or so-called moratoriums, could be explored for some applications deemed unacceptable as they are inherently incompatible with the Charter of Fundamental Rights.

(HIGH) RISK CANNOT BE SELF-ASSESSED. It should also go without saying that risk assessments, particularly in higher risk areas, cannot be self-assessed. We recognise that the risk-based approach in the GDPR is important to provide a good balance between innovation, flexibility, and protection, but this should only be for "low" risk applications, such as those that may be covered in the DSA framework (e.g. content recommendation systems). As for those that eventually are deemed to be "high risk" and would require market authorisation: there are numerous approaches to this in different sectors, from medicine to gambling. At the very basic level, it should be ensured that any prior authorisation is not a simple tick box exercise.

RISK ASSESSMENT MUST INCLUDE DIVERSITY. Mozilla strongly believes that in order to develop inclusive technologies that work for everyone, no matter of gender, race, or class, or anything else, it is crucial that those developing and designing these technologies be diverse. We believe the same should hold true for entities empowered to undertake risk assessments. Bias and inequalities have a tendency to mirror the technology that we build. Therefore it is critical to consider how different people are impacted by technology and regulations in different ways. An example is the way in which facial recognition technologies often misgender Black women at higher rates than any other intersection of gender and race (see: http://gendershades.org/). If the EU is committed to developing a truly human-centric AI ecosystem, recognising the importance of diversity, and building these perspectives and protections into the regulatory approach will be absolutely crucial.

RISK SHOULD BE ASSESSED THROUGHOUT LIFECYCLE OF APPLICATIONS. Finally, we strongly recommend that it is specified *when* risk will be assessed. For instance, in the ideation phase; upon release into the consumer market; or after its deployment? By definition, machine learning applications learn and develop as they collect more data. So it is also possible that harms may materialise *after* market authorisation and deployment. Risk can also be introduced well before an application is fully developed, e.g. even at the data collection phase. For instance, if an application collects census information such as area code, this can reveal race and/or class (as proxies), and thus engender discriminatory outcomes with respect to protected classes. In this case, bias and other potential harms must be explicitly addressed in order to protect against them, and it will be necessary to have information about race and class in order to assess whether or not the algorithm is in fact biased. Ensuring safeguards will mean that there is adequate oversight and impact assessments throughout the applications' entire lifecycle, among other important assessment criteria and oversight.

USE OF BIOMETRIC DATA

The collection and use of biometric data for the development and implementation of AI applications is a key component of the debate on AI regulation in Europe. As underscored in the GDPR, the collection and use of biometric data comes with significant privacy risks and should be carefully considered where possible in an open, consultative, and evidence-based process.

Biometric data is unique to each individual, and individuals should have the ability to reveal only specific relevant attributes associated with their identity depending on the legitimate requirements of the context. As a matter of course, AI applications harnessing biometric data should conform to the existing legal standards governing the collection and processing of biometric data in the GDPR. Without comprehensive safeguards around the use of biometric data (especially for facial recognition, e.g. in public spaces) the EU's strategy for "human-centric" AI will be futile. In addition to a comprehensive enforcement of the GDPR across the EU, AI applications using biometric data should therefore be subjected to a high threshold for deployment.

Besides questions of enforcement and risk-mitigation, we encourage the Commission to explore edge-cases around biometric data that are likely to come to prominence in the AI sphere. In this regard we are particularly cognisant of the case of applications that depend on speech recognition - a field of research that Mozilla is heavily engaged in. Our findings to date have suggested that, while speech is biometric in that it concerns physical and biological features of a person, speech in some cases may not necessarily constitute a *unique identifier* of a person e.g. a person's voice can often fluctuate considerably depending on their mood, time of day, environmental considerations, etc. In any case, the unique identifiability of biometric data is a primary reason why it requires enhanced protection and limitations on processing, and we would encourage the Commission to explore (in collaboration with the research community and data protection experts) whether and to what extent that condition may or may not apply in the field of voice recognition applications.

Thank you for this opportunity to share our views in the public consultation and we look forward to continuing to collaborate with the EU Institutions to develop a strong framework for the development of a trusted AI ecosystem. For any further information, please feel free to contact us at brussels@mozilla.com.