Summary

The European Evangelical Alliance welcomes the draft AI law the European Commission presented. It is a good start for further negotiations but more needs to be done to protect humanity, both individually and collectively.

Harm is more than just a specific violation of individual fundamental rights. The negative impact of Artificial Intelligence on moral agency, relationships, cognitive skills, dignity, life's opportunities and the environment should be taken into account as well. Therefore, we suggest the adoption of a specific Declaration of Digital Human Rights. This document should include a ban of all biometric identification and categorisation systems.

To increase trust in AI, there should be a comprehensive assessment for all AI, not just AI believed to be high-risk today. Further, the discussion about the future of AI should not be limited to business and developers. Civil Society should be included in shaping the future of our societies.

Authors

The European Evangelical Alliance is a pan-European movement serving the estimated 23 Million Evangelical Christians across the continent. In collaboration with an interdisciplinary group of experts in the field, we have studied the Draft EU AI Law. Our discussions form the basis of our comments included below.

Being Human

As Christians, we believe that all human beings are created in the image of God and are of innate dignity and worth. We are all uniquely gifted and uniquely imperfect. It is this combination of talents and limitations that makes our lives and our societies incredibly colourful and worthwhile. At the same time, preserving these unique human attributes does not always sit comfortably with the adoption of some technologies, Artificial Intelligence included. That's why humans must be masters of AI rather than slaves to it.

Risk

Risk assessment of AI should be carried out in a wider scope of harms to humanity rather than specific consumer orientated safety issues and human rights.

The focus on defined risks (either unacceptable or high) does not adequately represent the range of economic, physical, psychological and/or societal harm than can be caused by AI. AI can be used for great benefit, not least achieving some of the UN Strategic Development Goals, but it can also be used in such a way (intentionally and unintentionally) to cause injustice,

unfairness, and unequitable and/or unpalatable situations to an individual or a group of people. It can also result in marginalisation, and an increase in vulnerability, which is morally reprehensible.

The focus on defined risks rather than a spectrum of potentially harmful outcomes could result in AI uses which are still inadvertently harmful to people, and that could discriminate on grounds of religion or belief, amongst other fundamental human rights.

The categories by which AI is deemed unacceptable or high-risk, leave other types of potentially harmful AI uncovered, or considered as somehow lower risk, and therefore not afforded the protections of the AI governance safeguards proposed by the Draft EU AI Regulations.

Our view is that a more diverse assessment of AI is required on the basis of an outcomes-based approach, as opposed to a mere risk-based approach without sufficient guiding principles. This is necessary to ensure that AI practices are prohibited and fully protected with necessary and proportionate safeguards based on their perceived harm, including environmental harm. Therefore, it is our view that there should be a comprehensive assessment for all AI applications, not just AI believed to be unacceptable or high-risk today – this way any legislation would survive the test of time and technological advances and might be future proof.

We would like to see the following applications of AI added to the high-risk category:

- Research and Development of certain AI applications: these are not explicitly excluded from the regulation, allowing entities to create AI which may be detrimental and harmful to humanity but that, if popularised and profitable, could become mainstream and normalised.
- Any AI which extrapolates and uses in-depth insights into a person that the person may not themselves be aware of, e.g. AI used in Augmented or Virtual Reality to infer certain human emotions or reactions from information which is extrapolated from their use such as skin reaction, eye movement, facial responses, stamina, stimuli, etc. Knowledge of these responses could be used to manipulate, to excessively nudge, or inappropriately interfere with a person's freedom of thought, critical thinking processes, undermining their judgment.
- Augmented AI and transhumanism – The current draft does not include any provision in relation to augmented humans and computer brain interfaces which advance or enhance human capacity and capability. There needs to be greater discussion as to the bioethics and AI ethics of such applications and clarity in law of the boundaries of what is considered safe and morally acceptable, particularly given that this kind of adaptation will have an (in some cases) irreversible impact on a person(s) and impacts on how society views and treats disability, amongst other physical and body traits.
- Any AI which is either a standalone product or as a component part of a device or machine which has a detrimental impact on the environment. We know that machine learning is data hungry, and that data processing requires a high level of energy consumption. Whilst some AI can be used to help reduce carbon footprint and to lower energy consumption, other AI can put a heavy burden on the environment. An environmental impact is high-risk and must also be considered in the impact assessment of AI.

We would like to see the following applications fall within the unacceptable category of risk:

- All biometric identification and categorisation systems – The technology is not sufficiently accurate or precise for the application domains in which such systems are and can be used. These systems can be developed for one purpose, but their use can be adapted and/or repurposed for in almost any public situation, readily allowing for purpose and mission creep (and misuse) without sufficient checks and balances. This is particularly concerning when such systems are used by powerful entities, such as governments, large technology companies, large data companies, and the military. Whether such systems are live in real time and/or post facto, they would still have the same source issues. In addition, such systems are an unacceptable intrusion into the privacy of an individual that is a fundamental aspect of personhood.

Harm

The main concerns of the proposed regulation of AI are to ensure consistency with the EU Charter of Fundamental Rights and the existing secondary Union legislation on data protection, consumer protection, non-discrimination and gender equality. This is too narrow a scope in regard to the impact of AI on humanity since it does not pick up on harms that fall outside of these concerns such as the following:

- Moral agency – There does not appear to be any detail concerning an individual's moral accountability for the actions or inactions of AI. We believe that it harms humanity to delegate moral agency to an artefact.
- Relationships – There does not appear to be any accountability for the impact of AI on relationships, and the ability of AI to devalue, de-humanise, and undermine human-to-human relationships. Furthermore, no account is made of the potential risk of dependency (both emotional and physical) on humanoid, embodied or personified AI robotics, causing concerns over human-to-machine relationships.
- Cognitive skills – There appears to be little or no protection to safeguard against loss of skills and in particular the slow erosion and phasing out of human cognitive skills like critical thinking.
- Dignity – Even though the right to human dignity lies at the very heart of fundamental rights, that respect for human dignity does not appear to be followed through into the draft AI law where AI can be used for human augmentation, for nurturing or caring for the elderly, sick, and inform, or after death in respect of representations of a deceased person. We believe that there is also dignity in work and the proposed legislation does not adequately cover the need to balance efficiency that AI software and robotics can provide with preserving human work.
- Limiting life's opportunities – Whilst it is acknowledged that both private and public sector organisations need to manage day to day operational risks associated with people's actions and behaviours (such as criminal or fraudulent acts, or poor management of finances and credit repayment behaviours), there appears to be little protection or safeguarding against unfair and/or biased outcomes which limit life opportunities based on a person being profiled on the sum of their acts and/or behaviours, or more particularly the sum of the acts and behaviours of people with similar features and attributes to them. This does not just impact on education, but on employability, creditworthiness, law enforcement, migration and administration of justice. No distinction is made between automated and/or autonomous decision

making, or assistive and/or semi-autonomous decision making, particularly where personal data is not used and therefore GDPR does not apply.

- Digital exclusion – There are risks associated with having all pervasive AI adoption without viable non-digital alternatives or options to opt out. The categories of risk as they stand do not include the risk to people being excluded in society and/or exclusionary practices which means certain people or people groups get left behind. We must protect the vulnerable and marginalised, recognising that individual categorisation and optimisation (not just social credit scoring) may cause wider injustice to a people group.
- Environmental impact – As highlighted above, AI can reduce negative impact to the environment, but can also contribute to it. It has a second order impact on the ability of humans to flourish, in particular if this effects food and water supplies, due to global warming.

As fundamental rights are first and foremost individual rights, this lens could easily forego the impact of AI on humanity, human beings in the aggregate, and on society as a whole. Society is more than a mere banding together of individuals and individual rights. Any protections in the draft EU AI Law should have due regard to this nuance whose frequency is likely to be higher where profiling, optimisation and categorisation of people groups occur. Hence, group harms are likely to occur more regularly than individual harms, yet (as it stands) there is no real effective means of those impacted and/or influenced by AI within a group who have suffered collective harm, to appreciate that harm or be aware of others who have been impacted and/or influenced in the same way.

It is important to stress here that harm will not always be tangible and/or visible, but harm will have nonetheless occurred. It may take some time for that harm to transpire particularly where it impacts groups of people who are unknown one to another. Harm can be psychological, physical and economic, however quantifying harm done to a society spread out over time does not fall easily into these determinants which have a much more individualistic feel to them.

Declaration of Digital Human Rights

Given the potential impact of Artificial Intelligence on both individual citizens and our societies, the discussion about the future of AI should not be limited to business and developers. Civil Society should be included in shaping the future of our societies. Citizens and experts together could draft a Declaration of Digital Human Rights. This could include the right not to be manipulated, not to be exploited based on any attribute which might make them vulnerable, or otherwise take unfair advantage of a vulnerability (known to the person or not), not to suffer exclusion, detrimental or unfavourable treatment as a result of an interference or prediction of personal or personality traits, without consent or a lawful, equitable and justified reason to do so, the ability to seek human review and redress etc.

Such a Declaration could also incorporate the other harms to humanity cited above, thus widening the protection given to individuals and societies, effectively providing a right not to be harmed in these ways.

Data

Data (whether actual or inferred) is intrinsic to our personhood, and its use in AI has the power to impact a person positively and negatively. Data cannot be debiased. Whether the bias is introduced through the data, the modelling, the algorithmic intelligent system, and/or the people, the outcome is what has the impact.

Data is not just drawing insight and inference and meaning, it is delving into the depths of a person, their history, behaviours, transactions etc. AI techniques which result in categorisation and optimisation is averaging the uniqueness of human identity.

A variety of recent research findings have highlighted the unrealistic expectations that many had for AI tools used in the battle against the Covid pandemic. A recent article in MIT Technology Review[1] reported that, "many hundreds of predictive tools were developed". Yet, it concluded that "None of them made a real difference, and some were potentially harmful." It is suggested that the fault lay with the data and the way it was used. These findings should give legislators pause for thought in the drive to encourage more data sharing.

At the heart of all concerns over the harms to humanity is the use of data that in an EU GDPR context can be regarded as private. There is a disconnect between recognising the harmful impact that the use of this data in certain AI systems has on humanity and the business and political drive towards economic and technological superiority. In effect, AI has become a new arms race with societies becoming increasingly damaged as a result.

This raises the question of what sort of society do we ultimately want to create? Is it one where human flourishing is defined by economic prosperity and technological advance or is it one that values the uniqueness of human beings and seeks to preserve these attributes, thus taking control of technology that potentially undermines these values and attributes.

Human flourishing does not equate directly with unfettered technological progress and innovation, especially in the area of AI and machine learning.

---

[1] Will Douglas Heaven, Hundreds of AI tools have been built to catch covid. None of them helped, MIT Technology Review, July 30, 2021.