

BoraCo's submission to the consultation on the "White Paper on Artificial Intelligence - a European Approach to Excellence and Trust"

June 2020

Who is BoraCo?

BoraCo is a small consultancy, consisting of sustainability, risk and technology specialists working with organisations building tomorrow's world. We co-founded the Conscious Advertising Network in 2018 after having witnessed, first hand, the considerable and hard to foresee effects which can be brought about through apparently small, but interrelated risks associated with digital advertising.

Consultation response

Although we welcome the desire to create and develop legislation which ensures that AI is 'trustworthy', we are concerned that:

- The paper's risk based approach is not sufficiently elaborated upon and does not allow for scrutiny
- Risk is too narrowly defined, complex harms caused by frequently occurring, low impact risks are not accounted for, and impact is subjective and not equally shared among the population.
- There is no process outlined for monitoring and reassessing technology for the appearance of 'unknown unknowns'
- Environmental degradation is not adequately considered

Given the above, we are concerned that a risk based approach as it currently stands is not fit for purpose, and advocate the use of a Human Rights approach for regulating AI.

1. The paper's risk based approach is not sufficiently robust and does not allow for scrutiny

Although we believe adopting a risk based approach ensures any regulation is compatible with GDPR, we are concerned that the Commission's 'Theory of Risk' is not yet sufficiently qualified and quantified. The white paper does not currently clarify terms, or define its theory of risk in detail, leaving room for ambiguity which could potentially allow harmful technologies to remain unregulated.

We would strongly recommend that the EC is more transparent about how their risk framework has been constructed and to clarify their risk assessment criteria and model. Our main observations include:

- **The categorisation of likelihood and impact is unclear, more information needed to understand the definitions of:**
 - Low - Med - High
 - Low versus high probability
 - Measurement of impact e.g. fiscal cost, human lives, or jobs
- **Quantification and measurement of impact is not defined. Further information is needed on:**
 - How impact will be monitored to verify the assumptions described
 - Clarification of approach to verifying risk assessment criteria and model, and risk management strategies.
 - How the risk assessment approach will be verified
 - How likelihood is to be assessed
- **There is little provision for identifying, tracking and monitoring complex and interrelated harms:**
 - We reject the idea that AI applications not covered by this whitepaper should be subject to a voluntary code of conduct. In the case of digital advertising, non-binding codes of conduct have been found to be ineffective at best, and social media industry is actively calling for legislation, following failures of corporate governance and codes of conduct.
 - There is little information about how complex harms will be monitored, categorised and acted upon, as, in the case of digital advertising, cumulative effects from seemingly innocuous applications of technology have caused real, far-reaching harms.
 - Further information is also needed on the EC's approach to identifying, tracking and mitigating unknown unknowns. This is vital to an effective strategy.
 - The paper appears to assume 'impact' and 'risk' would be shared equally across the population, and does not currently contain provisions for vulnerable people.

Therefore, we recommend that the EC publishes its quantified and qualified theory of risk, including provisions for vulnerable people and monitoring of 'unknown unknowns' for public scrutiny.

2. Risk is too narrowly defined, complex harms caused by frequently occurring, low impact risks are not accounted for and impact is subjective and not equally shared among the population.

Risk itself is composed of two components - impact and probability. The risk-based approach outlined in the consultation - as we currently understand it - does not attempt to identify or assess risks beyond:

- Material risks include "safety and health of individuals, including loss of life".
- Immaterial risks include "loss of privacy, limitations to the right of freedom of expression, human dignity, discrimination for instance in access to employment".

However, significant risk can be created by lower impact, but frequently occurring incidents. For example, the frequent serving of personalised content online has contributed to filter bubbles and polarisation, as well as discrimination against marginalised groups in areas such as job searches. The reliance on digital advertising as an internet funding model has resulted in complex harms to the environment and society, such as:

- The creation of funding models for hate speech and misinformation, as identified by the UN¹, threatening democracy, marginalised communities and undermining trust in institutions.
- Data and privacy breaches, and excessive data collection as industry standard, practices which has now been exported internationally.
- A decline in the quality of national and local press caused by undermining the funding model of quality journalism.
- Discrimination through personalised advertising in jobs, education and housing.
- Creation of sophisticated online scams which cause trauma and financial losses for many.
- Environmental degradation as a result of the additional data centres needed to serve the above.

Few of the above impacts would have been classified as 'high risk' under the proposed system, yet they have strongly shaped our society and impacted international communities. In our opinion, a focus on direct and material impacts is not appropriate given the scope for complex harms and huge secondary/systemic impacts from the implementation of technology. Particularly those which cannot yet be predicted.

Given the above, we strongly advocate a change to a human rights based approach, with a sliding scale of 'risk' that encompasses:

- *Applications of AI which should be banned. E.g. use of biometric data for mass surveillance.*
- *Applications which require a full, publicly available human rights impact assessment by third parties, which are repeated every 2 to 3 years. E.g. Use of facial recognition in advertising, use of machine learning to target advertising.*
- *Applications which require impact assessments to be conducted every few years, by industry, and audited by third parties. E.g. Use of personalisation engines on Spotify.*

Should the current risk based approach be chosen however, we recommend the Commission builds in measures to monitor and assess 'lower risk' applications of AI technology which may create or exacerbate complex harms.

Case study: a toxic combination of AI and advertising is both inadvertently recommending and funding disinformation and fake news around climate change - a

1

<https://2019unforumbhr.sched.com/event/X8jo/the-economics-of-hate-the-role-of-business-in-ensuring-ethical-advertising>

report by Avaaz² found climate disinformation was being both funded by advertising, and prioritised by YouTube's recommendation engine, while disinformation on the open web has an associated ad funded business model. Responsible advertising strategies do not yet include exclusion of fake news and misinformation as key success factors. Many tech companies have ambiguous stances on climate issues, lobbying for business as usual³, or not being transparent about the footprint of their operations⁴.

4. Environmental degradation is not adequately considered

Environmental degradation is not covered or mentioned, despite significant evidence that the energy and resource requirements of AI training and use require scrutiny and management, given the requirements and obligations set out for EC member states under the Paris Agreement.

If the global IT industry were a country, only the United States and China would contribute more to climate change, according to Greenpeace's #ClickClean⁵ report, and the amount of energy used by data centres continues to double every four years, meaning they have the fastest-growing carbon footprint of any area within the IT sector.⁶ Researchers⁷ estimate that the tech sector will contribute 3.0–3.6% of global greenhouse emissions by 2020, more than double what the sector produced in 2007⁸ and the estimated 2020 global footprint is comparable to that of the aviation industry.⁹

This growth in environmental degradation will be exacerbated by many AI applications, which requires huge amounts of energy to train and operate. Although AI can also be applied to reduce carbon footprints of operations, it comes with a heavy cost of its own. For example:

- OpenAI reported that “[s]ince 2012, the amount of computing power used in the largest AI training runs has been increasing exponentially with a 3.5 month doubling time (by comparison, Moore’s Law had an 18 month doubling period).”¹⁰ As AI relies on more compute, its carbon footprint increases¹¹.
- A study from the University of Massachusetts, reported that training one AI model produced 300,000 kilograms of carbon dioxide emissions¹², roughly the equivalent of

² https://secure.avaaz.org/campaign/en/youtube_climate_misinformation/

³ <https://www.theguardian.com/environment/2019/oct/11/google-contributions-climate-change-deniers>

⁴ <http://www.clickclean.org/international/en/>

⁵ <http://www.clickclean.org/international/en/>

⁶ <https://www.thealternative.org.uk/dailyalternative/2019/5/19/data-centres-energy-use>

⁷

<https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c>

⁸ <https://www.sciencedirect.com/science/article/pii/S095965261733233X>

⁹ <https://www.atag.org/facts-figures.html>

¹⁰ <https://openai.com/blog/ai-and-compute/>

¹¹

<https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c>

¹² <https://arxiv.org/abs/1906.02243>

125 round trip flights from New York to Beijing. Organisations which use and train AI have been accused of not reporting their environmental footprints transparently.¹³

Training and deploying AI has a huge, and growing environmental footprint. The industry urgently needs to map and mitigate its impacts in line with the goals laid out in the Paris Agreement. Organisations using AI should conduct environmental impact assessments and engage in yearly environmental reporting on those impacts.

In conclusion

We believe that the scope of harms being considered by AI developers is currently too narrow, and that human rights impacts and environmental impacts are the way to change this thinking and embed best practice, and an ecosystem of excellence within the EC.

Many industries, such as advertising, stand at the brink of widespread adoption of AI, and have little to no appreciation of how to embed and account for human rights within their operations. Failure to change this thinking risks ingraining excessive data collection habits, inadvertent discrimination, and flawed metric driven decision-making in our technologies and society for years to come. The time for broader consideration of consumer protection, human rights, and environmental impact within AI decision making is now. We consider this, alongside the Digital Services Act, a key moment to ensure that AI is regulated in a way which does not allow the problems of the past to repeat themselves.