A response to **Supporting Responsible AI discussion paper**

This response takes the form of several bullet points, and refers to page numbers of the document where relevant.

* Page 3 states that AI is unique because it can take actions at speed and scale that would otherwise be impossible. This is true, but these characteristics do not make AI unique. Many other computerised solutions can do the same.
* Page 5 states “Machine learning are the patterns derived…”. This is a somewhat exotic definition, whereas machine learning in used to refer to the model of learning, not the patterns.
* Page 7 mentions opportunities are almost impossible to accurately forecast. AI operating in the highly complex system of human society, this is accurate. The same applies to challenges as well, however, a point that is omitted.
* Page 8 contains valuable discussion on bias. However, we need to keep in mind that the bias is often learned from what is inherently a biased society. One cannot simply “correct the bias”; instead, there needs to be a discussion on what sort of society we want the AI systems to reflect. At present, elements of society are “over-correcting” for past biases through activities like affirmative action. There is an implicit notion that these are temporary measures until the world is no longer biased, but baking *them* into AI systems can create systems that perpetuate a different sort of biased world – not necessarily any worse or better than the current one, just differently biased.
* Page 10 should also mention aviation as a target for sector-specific regulation. There is much that that AI domain, and automation more broadly, can learn from how the aviation industry is handling it – it has had to deal with many analogous issues decades earlier.
* Page 14; requirement for explainability is deeply problematic, for some type of AI systems are technically incapable of producing an explanation.
* Further on page 14, there is a statement that adopting ethical principles would “ensure appropriate accountability and governance mechanisms are in place”. Adopting a set of ethical principles does no such thing; they go some way to mitigating the potential negative impacts, but they do not “ensure” anything.
* The table on page 27 does not take into account the lack of transparency as a limitation of voluntary methods. On regulatory side, there is no potential for “stifling” innovation – there IS potential for slowing it down, however, which is different.
* Page 29 has a valid point that regulations can facilitate rather than hinder innovation; this, however, contradicts the earlier point of stifling innovation.
* Page 29 also talks about building explainability into AI systems. It should be stressed this is not always possible. We should explore whether it is always desirable; let’s keep in mind that we didn’t know the science of how steam engines work until 50 years after their invention.
* Page 30 talks about bans; it should be emphasize that banning particular technologies or products, such as ChatGPT, should be seen as a temporary measure until appropriate governance is put in place. Legislation in particular should never talk about individual products as targets.
* The risk table in Box 4 / page 32 is broadly speaking appropriate, but has some issues. The biggest one is that the risk level determination must NOT be left 100% to the organizations developing the system; there is a clear conflict of interest here. A 3rd party must either do the classification, or the government must be allowed to spot-check the classifications and over-rule the organization’s decision. Whistle-blowers must be encouraged, protected and rewarded in this domain.
* Further, AI-enabled recommendation engines should not be classified as a low risk, nor should chatbots.
* A suggestion would be a score sheet for things that are merely recommended. For example, the AI principles are voluntary – what if all products had a public “Ethics Label” with simple tick marks on which principles have been taken into account for this use case, and links to more information for consumers and organizations to find out how, exactly, the principles have been taken into account.