Does the European AI Act bring ethics and law closer together, or does it risk driving them apart?

Alistair Knott

Centre for Data Science and AI, Victoria University of Wellington. Mail: ali.knott@vuw.ac.nz

First some context from me. My background is in AI research - I have been an AI researcher since the mid-90s. In the last ten years or so, I have found myself involved in many discussions about AI governance. Most of these have involved groups I co-lead: the Global Partnership on AI's Social Media Governance group, and more recently, the Social Data Science Alliance.

I would like to argue that the EU's AI Act is a fundamentally ethical project. I'll present my argument in four steps.

1. Powerful new technologies must be governed.

The moral premise of my argument is that powerful technologies must be governed, so they are deployed for the public good. Powerful technologies have large impacts on society. These impacts must be controlled, so they are beneficial on balance. In principle, any form of government could be appropriate, to provide this government. It could involve self-regulation, informal guidelines of use, technical standards, black-letter law, or some combination of these. It could be national or multinational. But there must be some form of government. Something has to act to ensure the technology is used in ways that are beneficial to society.

2. The scale of impacts in AI is unprecedented.

AI in the middle of an old-fashioned scientific revolution. My favourite definition is that of Imre Lakatos, who frames scientific revolutions as 'progressive research programmes', where each new discovery creates the context for further discoveries. What we are seeing in AI at present is an extraordinarily progressive research programme: the field is progressing at such a pace that researchers are hard pressed to keep up. These advancements are precipitating an industrial revolution which is equally clear to see. People are talking about AI in offices, on buses, in student cafeterias, in school staffrooms. It's on the agenda for governments, and in boardrooms. There is no question that AI is a powerful new technology - and therefore, by premise 1, it must be governed.

3. Impactful AI work is being pursued by companies with vested interests, which will not act independently.

The main advances in AI come from large multinational tech companies. These companies have the most data, and the best compute resources, and can attract the best talent. (The most productive engineers are young researchers in their 20s - another sign of a scientific revolution.) To govern AI, there must be governance of these large companies. In principle, implementing principles of governance should be easy. AI systems are computer programs, deployed to very large global audiences. If something is wrong with an AI system, the code can be changed, and the user community will immediately receive an updated version. However, companies' motives are not perfectly aligned with the social good. They are profit-making entities, first and foremost. They certainly need to look after their users, to make sure that they don't go elsewhere. But keeping individual users 'engaged' with an AI product is a very different thing from 'acting for

Forum

the social good'. I'll point out some salient differences.

First, what feels good to individuals is not necessarily good for communities. Drugs are a case in point. It's possible, sometimes even easy, to make users *dependent* on AI. The AI technologies used in social media platforms are a good example. These technologies are set up to 'give users what they want': literally, to monitor what they click on, and then give them more of the same. Social media addiction is a well-known phenomenon: phones and screens capture our attention, and much of this is due to the agency of AI. The new generative AI tools are another case in point. Once users become accustomed to being aided by an AI system, it is hard for them to do without it. Children and students growing up with AI will be dependent on it from the outset. Generative AI companies, fully aware of this, are engaged in a turf war, competing vigorously for market share.

Secondly, users can become 'locked in' to tech products. Social media platforms are perhaps the prime example here. Users work hard to gather an audience on a platform - followers, clients, friends - but they cannot take this audience with them if they choose to leave the platform. The channel of communication they have with their audience is owned by the platform. (It is unprecedented for a communication medium to be owned in this way.) AI systems have a similar property of locking users in. When you have become used to a certain AI interlocutor (friend, assistant, colleague, romantic partner), you don't want to trade in for another version.

Both these kinds of dependence are good for the companies that supply AI products. But they are not good for society more broadly. To shape AI for the public good, therefore, we must sometimes act against the interests of the big AI companies. Given the enormous power of these companies, this action can only happen through the law.

4. The AI Act is the starting point for a discussion about good governance of AI.

The EU is the foremost large jurisdiction in the world working on black-letter AI regulation. The AI Act is the prime example of this - but the Digital Services and Digital Markets Acts are very much part of this effort: what makes social media companies so attractive to users is the AI systems they run on, as already noted, so these three Acts should be taken together.

What we are seeing in the EU is the start of a process whereby light is shone on what is happening inside large AI companies. To give just a few examples: the Digital Services Act will enable auditors, and vetted independent researchers, to gain access to company data, to answer many questions about the social impacts of social media platforms. The AI Act will require companies to disclose what's in their training sets, to fairly enable questions about IP to be considered. The AI Act will also require companies to ensure their AI-generated content is 'detectable' as such.

Exactly how these measures are enforced is still, of course, a very open question. The law is just a starting point for the relevant discussions. But these important discussions would not be happening if there was no law. Without the law, much of the information needed for policymakers to frame detailed rules would simply not come to light. In short, the law brings companies to the table - and requires them to actively engage in discussions about the social wellbeing of their users.