AI and Ethics: Moral Concerns

Jeroen van den Hoven

University Professor Ethics and Technology

Delft University of Technology

Bridging the Gap, AI and Fundamental Human Values

Roundtable Meeting Article 17 13:00-14:30

November 15 2023, Library of the European Parliament

I would like to start of with a remark that was made by Commissioner Vestager on May 31 of this year in the context of talks she had with some of the CEO's of Generative AI companies. Generative AI, ChatGPT and LLMs had taken the world by surprise. Many were enthusiastically experimenting, others were either blissfully unaware or slightly worried by the miraculous performance of Large Language Models. Vestager said about generative AI "I think if we all are honest with ourselves about other technologies, including social media, *we probably wish we had not done things* (…). You know, we *could have* but *should we have*? And so let's work together to get this right, because the stakes are a whole lot higher."

That observation and analogy make a lot of sense.

After more than two decades of social media in our lives, it is clear that there are very serious problems with social media. The business model of the protagonists of the platform attention/ economy is crystal clear and the silicon valley approach of 'move fast, break things and apologize later', and 'innovation in the regulatory grey zone', has now been laid bare. Many recent investigations – e.g. regarding the activities of Cambridge Analytica - conjure up a rather gloomy picture. The list of negative externalities is long. Society now needs to deal with the consequences of what has proliferated online: False Information, Misinformation, Disinformation, Malinformation, Propaganda, Systemic lies, Conspiracy theories, Deep Fakes and Hallucinations. The accumulative harms come back to haunt us - as was the case in climate change.

WE have seen the numbers of Merchants of Doubt, Influencers, Meddlers, Trollsters steadily grow. "Conspiracy entrepreneurs" have crowded out serious attempts to understand the world. Advanced digital technology has given a helping hand to all of them. Future AI applications will pour rocket fuel over this problem (and they potentially can do that to many of our problems). As Barack Obama already observed in 2020: "If we do not have the capacity to distinguish what's true from what's false, then by definition the marketplace of ideas doesn't work. And by definition our democracy doesn't work. We are entering into an epistemological crisis".

What is important to realize in all of this is that it has become very clear that Big Tech has successfully off-loaded formidable cost to society and kept the remarkable profits for themselves and their shareholders. It was probably naïve to have expected otherwise. It was also naïve to expect that for-profit companies, driven by a relentless logic of quarterly revenues, would be concerned with the delicate fabric of society, our communication, interactions, the role of civility in discourse,

the importance of tolerance in democracies, ideals of conviviality, the fragility of the moral development of children. Trust in society, truthful communication and the appreciation of moral values in human lives; they are never in our excel sheets, but we know how much we need to do to bring them back once they are gone . One or two corporate ethical committees will not repair the damage, nor will they prevent future disastrous outcomes.

So the first problem with AI is not so much AI itself, nor its use. AI refers to a set of powerful technologies and techniques that can be designed and used for good, but that can also be designed and used malevolently, or recklessly, or negligently or self-servingly. The EU spearheaded the **responsible** use of AI for over 5 years now. Gradually state actors all over the world – and the UN - have risen to the occasion and joined the moral and global governance AI conversation. Many have come to appreciate the wielding of soft power by Europe and the so-called 'Brussels effect' that they saw at work in the way the EU successfully set Global standards for data protection by means of EU law, in the form of the GDPR. It is of the utmost importance that the primacy of democratic politics is fully reinstated in the age of AI.

This will be very difficult since it is very late in the day, and the network effects, the lock ins, path dependencies and winner take all effects are very powerful. Effectively bringing our public values, our fundamental rights to bear upon this extremely dynamic and complex world of Big Data and AI will require determination and hard and detailed work, and contributions from many disciplines and parties. In the meanwhile there are also other forces to be reckoned with. Powerful geo-political blocks in the world are onto AI big time. They have been identified as *EU's system's rivals*. They use AI to run society with completely different sets of values, a different image of human beings, different socio-economic models and often radically different ideas about human rights, rule of law and democracy. Europe's challenge is to solve our greatest problems and societal challenges regarding sustainability and climate, health, equity, social justice, human security, living conditions, immigration, **while at the same time remaining true the charter of fundamental rights and the European Convention of Human Rights and a conception of human dignity and respect for persons that is enshrined in it.** Our Innovations *in and with* AI need to be **responsible**, or the European project will be jeopardized at its core. This calls for what the European Commission has successfully promoted as Responsible Innovation with AI. Not in a naïve modus of finding simple technical solutions to social and political problems, but fully cognizant of the wider political context and the higher stakes that commissioner Vestager referred to in the quote I gave at the beginning.

Before I discuss what I take to be some of the deeper moral concerns regarding AI. I want to indicate that the literature about the more or less obvious ethical issues of AI's unreliability, its biases and discriminatory effects, its propagation of inequities, its intransparency and unexplainability, its shaky data protection and its often infirm governance are rapidly growing. They all have to be dealt with and are all in scope of the EU's AI act that requires Human agency and oversight, Technical robustness and safety, Privacy and data governance, Transparency, Diversity, non-discrimination and fairness, Societal and environmental well-being, and Accountability.

The first deep problem of AI I see is that it may be extremely hard to restore an adequate level of confidence and trust in our epistemic institutions, as represented by e.g. journalism, science, politics. I think the most dangerous technologies are in a sense the social or cognitive ones that prevent people from having a clear view of the world and the needs of others. AI is a special one in

that category. These technologies are often conducive to dehumanization and invite people to turn each other into self-obsessed and unthinking biased individuals. These technologies are potentially like fog machines that create the conditions that make it easy to renounce, deny or turn a blind eye to our common humanity and our human responsibility. They are technologies of bad faith. And those who work on them are often instrumentalized by others, naïve, culpably complicit, or masters in orchestrating plausible deniability.

The way AI brings about epistemic chaos is therefore I think one of our most serious threats. Generative AI and Chatgpt just mimic descriptive and referential use of language. Representation of the world is a happy coincidence, not even a goal function, let alone a virtue, a commitment or a passion.

A second problem that I already referred to is the demise of democracy and the withering of a public sphere that is conducive to democratic culture and democratic habits. As generative AI will gain more and more ground, it will – in its free range version – make it more and more difficult for non-experts to separate interesting, relevant and true information from trivial and dangerous rubbish. If we want to reclaim democracy and the type of discourse, mutual understanding and tolerance that forms it's life blood, we will have to bring it about by design. Not only by inoculation against desinformation and trolling, not only by target hardening in the face of doing battle with AI bot net armies, but also by experimenting constructively with AI powered and supported forms of democratic deliberation.

A third deep problem is concerned with the conditions of human responsibility. We more or less agree about which conditions need to be satisfied for attributing moral responsibility to an agent: freedom and no-coercion, knowledge, intention, certain mental and moral capacities and control. Coercion, ignorance, lack of control, can provide one with valid excuses, assuming that they are not self-caused. Responsible innovation is not about new technology with the enigmatic property of being responsible. A lazy chair is also not lazy itself. It has been designed to accommodate lazy people ,or those who feel or behave lazy-like. Responsible AI similarly should be designed to stimulate, support and accommodate people who are or strive to be morally responsible. Therefore AI needs to be designed to help us in achieving our moral responsibility by optimizing the conditions for responsibility. If no special efforts are made however it will not do so, quite the opposite seems to be true. It can mess with the conditions for responsibility. For example, if systems are fully autonomous (as we will see more and more often in armed conflicts in the future) how can we be said to have meaningful human control such that there is moral responsibility for untoward moral outcomes. If weapons system are like recommender systems in online shops which suggest 'you liked this target you may also like that target' and can autonomously engage a target, human moral responsibility seems to have evaporated. Or, our knowledge dependence on AI systems may have taken on such forms in certain settings, that it has become impossible to overrule an AI system without thereby taking a moral risk that one cannot justify at that very moment of non-compliance. The human being working with the system is in that case effectively reduced to a component of the system.

A fourth set of issues regards the immodesty to which AI systems may give rise when it comes to knowing persons and their fate. The vast amount of data -including brain data- we have on people may give rise to immodest claims to know them - better than they know themselves: a form of big data hybris. We know that a couple of your likes on social media will give some people reason to think that they know you better than a friend or partner. The epistemic authority attributed to AI

systems will make it difficult – if not impossible – to disagree in the AI's identification of an individual, not in a forensic or administrative sense, but in a moral sense. Respect for persons requires an acknowledgement that the whole persons **cannot** be known in full, not with all big data of the world and not with the most powerful AI. We need to acknowledge the privacy of mental life and its irreducible subjectivity. We owe other persons the attempt to see them from their point of view, i.e. including their own perspective and what it is like for them to live their life . We also owe it to them to provide them with the tools with which they can make most sense of their experiences. This is an important aspect of what respect for human dignity and the human person in the age of AI and big data mean, I suggest.

Finally, I would like to bring to the fore the crucial insight with strong European roots that our human rights and ethics need to be present at the right time, at the right place and in the right format for them to have a chance to make a difference. If we do not bring our ethics to bear upon AI, effectively, continuously, transparently, carefully, then others may do it for us, ineffectively, haphazardly, self-servingly and insidiously.

I started with a quote from commissioner Vestager, I want to close off with a quote from Paul Nemitz, senior legal advisor of the commission: "In order to protect and strengthen Western Liberal democracies in the Age of AI and the core trinitarian idea of 'human rights, rule of law and democracy' we need " a new culture of technology and business development …which we call human rights, rule of law and democracy by design". All the ethical ideals that we pursue in law and politics, technology and economics, the ethical principles and values we are committed to, we need to design for them, explicitly, demonstrably, systematically, continuously, transparently, inclusively. This is the ideal of Responsible Innovation with AI. Failure is not an option. Europe has to make its ethics work in the Age of AI.