Regulating Artificial Intelligence

Prashant Saran
Former Wholetime Member
SEBI

I was not born when leading scientists wrote a letter urging the US Government not to use the Atom Bomb. But as a child I was exposed to the fear of nuclear holocaust. Since then, I haven't seen any technology evoking such fear as Artificial Intelligence (AI) . I have been following the debate on regulating AI and as a career regulator would like to contribute my bit.

Ever since Generative AI Models such as Chat-GPT and Dall-E became publicly available, there has been much discussion about AI in media. The discussion about AI ranges from existential fear of machines taking over humans in a malicious fashion, elimination of almost all jobs, helping the state become totalitarian, perpetuation and accentuation of existing biases in employment and law enforcement.

In the light of the above, it is seen that governments and international organisations are proposing regulations that will make AI safe while exploiting its potential for betterment. In 2021 EU had proposed AI Act and in June 2023 it has adopted Parliaments negotiating position on the AI Act. The talks will now begin with EU countries in the Council on the final form of the law. EU has adopted a risk based approach classifying the systems into those having unacceptable risks, high risk and limited risk.

Unacceptable risks refer to the systems that are considered as threat and need to be banned. These include those systems that use subliminal messaging to manipulate behaviour and targeting vulnerable groups like children, social scoring and remote biometric identification systems. These are proposed to be outright banned except for criminal investigation and military purposes. High risk systems refer to product safety areas like aviation, toys, medical devices etc. on one hand and eight specific areas such a biometric identification, education, and management of critical infrastructure etc. on the other. All high-risk AI systems will be assessed before being put on the market and also throughout their lifecycle. Other AI systems are classified as limited risk involving disclosures and leaving it to the user whether to use them or not. Generative AI systems like Chat-GPT are subjected to transparency, copyright disclosures and preventing generating illegal content.

The US approach to regulating AI is indicated in a White House Document, Blue Print for an AI Bill of Rights. To summarise, it talks about right of the citizen to be protected from unsafe or ineffective systems, not to be discriminated against by algorithms, having agency over personal data, knowledge and a notice that an automated system is being used that tells how the outcomes might affect the citizen, and the right to opt out have access to a person who can quickly consider and remedy problems. It is believed that the AI Bill of Right will act as set of backstops against potential harm.

Similar approaches have been recommended by various scholars. Other countries, including India have articulated their own vision for a safe and helpful deployment of this emerging technology. Unlike most technologies, the AI industry itself is very keen that the AI should be regulated. In this article, I would attempt to explore how difficult it will be put these high level ideals into regulatory practice and suggest a supplementary approach.

The first thing we notice about these proposals to regulate AI largely focus on protecting an individual from the possible harmful effects of AI. However they don't consider a scenario in which the AI turns rogue and either makes humans its slave or destroys the human race altogether. No less a thinker than Yuval Noah Harari has pointed out to the risk and no less an industry insider than Sam Altman (CEO of Open AI) has mentioned a distinct possibility of AI going rogue. Considering the stature of the these worriers, it will not be prudent to ignore these concerns as a science fiction scenario.

It is easy say that the AI systems should be ethical or unbiased or human centric while drafting a law. There will be huge difficulties in framing actual regulations laying down the practical guidelines. The Collins Dictionary defines ethics as, "a social, religious, or civil code behaviour considered correct, especially that of a particular group, profession, or individual" Now, while framing the regulations it will be important to define what ethical standards should the regulator follow. Whether it should be the ethical standards of western civilisation or the civilisation of the jurisdiction or an amalgamation of the two. Even if it is argued that the AI is a product of western civilisation, the problem will still remain there. A regulator with a puritanical background might devise different tests than one with liberal background. A 1999 paper[1] (of course, the situation has changed by now) argued that the rules of conduct, Cultural factors recognised in respect to a particular class of human actions or a particular group, culture, etc will determine the outcomes. Different cultures have different rules of conduct and therein lies the issue of understanding the roots of ethics across cultures.

"If a US company resorts to bribery, it faces great pressure to hide it, including hiding it in financial statements. In contrast, other countries have a more tolerant or pragmatic view of bribery. As a case in point, at this writing, bribes are explicitly tax deductible in another Western country, Germany. When considering countries that do not share a common cultural heritage, the challenges can be even greater."

The problem will be almost intractable for regulators in non western cultures like India or China while devising the regulations regarding AI operations in their jurisdictions.

Let us turn our attention to the next important thing in the approaches to regulating AI. It concerns biases in the data and algorithms. What is considered as a normal action by one group might be considered as biased by another. To take a hypothetical example, let us consider a jurisdiction that has two communities A and B. Historical

data suggests that out of 1 million members of the group A 10 are convicted of serious crimes every year while the data for group B says that 100 members of group B have been convicted. While zeroing on possible suspects while scanning a large number of presence on scene of crime, the question before the writer of an algorithm will be whether to take the membership of the groups A or B as a variable or to ignore it. One view could be everyone near the scene should be have an equal probability of being the criminal, the proponents of this view will consider it an unbiased approach. On the other hand another view could be that assigning equal probability is ineffective as the shortlisted group will have much smaller number of members of group B than probabilities would require leaving open a possibility that the real criminal may be left out.

Similarly in a college admission algorithm whether introducing affirmative action will be considered introducing bias in the algorithm or it will considered as correction to a historical injustice will be a matter of debate.

Based on the brief discussions above, I am of the view that it will be quite sometime before these contentious issues are debated and settled and in the meanwhile AI engines might as well turn rogue, making the proposed regulations irrelevant. It is necessary to take some rough and ready steps. In this article, I am not suggesting any definitive approach but only suggesting a line of examination.

The most important issue is that of preventing roguish behaviour on the part of algorithms. I don't believe that the immediate threat comes from AI becoming sentient and thereby vengeful. The neural circuitry is still not at that stage where this could be a real threat. More likely it is going to be a coding error that might leads to bizarre outcomes. There is famous example of a pricing algorithm going berserk given by European Commissioner Margrethe Vestager in her speech on "Algorithms and Collusion" on March 16, 2017, described as follows:

"A few years ago, two companies were selling a textbook called The Making of a Fly. One of those sellers used an algorithm which essentially matched its rival's price. That rival had an algorithm which always set a price 27% higher than the first. The result was that prices kept spiralling upwards, until finally someone noticed what was going on, and adjusted the price manually. By that time, the book was selling – or rather, not selling – for 23 million dollars a copy."

It will be helpful if the governments could insist on all algorithms mandating hardcoding of the algorithms so as to put a quantitative limit on the outcomes of an algorithm. For example, security market regulators have put a limit on order cancellation ratio in case of High Frequency Trading algorithms. An AI system that does inventory management could be hardcoded not to give recommendation for a purchase of more than 200% of average inventory level of the last year. I do understand that many AI systems already do it but if the law requires all systems to do it, there will be a reasonable chance of preventing a cascading uncontrolled behaviour on the part a set of algorithms.

So far all the suggestions regarding regulation of AI focus on regulating the designers and implementers of AI systems. No responsibility has been yet cast on the users. The second line of inquiry we could explore is to make users responsible for their actions. Generative AI systems like CHAT-GPT are likely to shape themselves not so much by the design of the Large Language Models (LLM) as by the prompts they receive. The direction in which a LLM proceeds will also depend upon the prompts it receives. Firstly, use of computer generated prompts could be banned outright. There should be code of conduct for the natural person giving a prompt and those prompts that do not fulfil the requirements are disabled by the system itself. I do understand that the LLMs have incorporated some rules in their systems as to prevent some kinds of prompts from being responded to but it needs to be examined in detail by the civil society and formalise these code of conducts.

One could object that it is like making a user responsible for a product performance rather than the manufacturer. We may not recognise or remember but taming of use of fire was the biggest ever technological breakthrough. The manufacturers of fire, for example the matchbox makers are subjected to regulation but the users are also constrained in its use by the Penal Codes all over the world. Same goes for use of dynamite or knives. The users must be made responsible for what they do with AI systems.

Given the pace of development of AI, we could explore the possibility of implementing some rough and ready regulations like the above mentioned two suggestions as the latency for these regulations is likely to be very short. Of course, the existing efforts at regulating the manufacturers of AI systems as suggested by the current proposals should go hand in hand.

---

[1] Ethical Issues Across Cultures Daniel A Pitta et. el. JOURNAL OF CONSUMER MARKETING, VOL. 16 NO. 3 1999, pp. 240-256 # MCB UNIVERSITY PRESS, 0736-3