Confronting the longterm risks of Artificial Intelligence

Posted at: 18/10/2023

Context:

Risk is a dynamic and ever- evolving concept, susceptible to shifts in societal values, technological advancements and scientific discoveries. For instance, before the digital age, sharing one's personal details openly was relatively risk free. Yet, in the age of cyberattacks and data breaches, the same act is fraught with dangers.

Risks associated with AI:

1. Our understanding of Artificial Intelligence (AI)-related risk can drastically change as the technology's capabilities become clearer. This underscores the importance of identifying the short and long term risks.
2. The immediate risks might be more tangible, such as ensuring that an AI system does not malfunction in its daytoday tasks. Long term risks might grapple with broader existential questions about AI's role in society and its implications for humanity.
3. Addressing both types of risks requires a multifaceted approach, weighing current challenges against potential future ramifications.

Over the long term:

1. Yuval Noah Harari has expressed concerns about the amalgamation of AI and biotechnology, highlighting the potential to fundamentally alter human existence by manipulating human emotions, thoughts, and desires.
2. One should be a bit worried about the intermediate and existential risks of more evolved AI systems of the future — for instance, if essential infrastructure such as water and electricity increasingly rely on AI.
3. Any malfunction or manipulation of such AI systems could disrupt these pivotal services, potentially hampering societal functions and public well being.
4. Similarly, although seemingly improbable, a 'runaway AI' could cause more harm — such as the manipulation of crucial systems such as water distribution or the alteration of chemical balances in water supplies, which may cause catastrophic repercussions even if such probabilities appear distant.
5. AI sceptics fear these potential existential risks, viewing it as more than just a tool — as a possible catalyst for dire outcomes, possibly leading to extinction.

The Evolution to Human Level:

AI that is capable of outperforming human cognitive tasks will mark a pivotal shift in these risks. Such AIs might undergo rapid self improvement, culminating in a super-intelligence that far outpaces human intellect. The potential of this superintelligence acting on misaligned, corrupted or

malicious goals presents dire scenarios.

Ethics and AI:

1. The challenge lies in aligning AI with universally accepted human values. The rapid pace of AI advancement, spurred by market pressures, often eclipses safety considerations, raising concerns about unchecked AI development.
2. The lack of a unified global approach to AI regulation can be detrimental to the foundational objective of AI governance — to ensure the long term safety and ethical deployment of AI technologies.
3. AI Index from Stanford University reveals that legislative bodies in 127 countries passed 37 laws that included the words “artificial intelligence”.

International Collaboration:

1. There is also a conspicuous absence of collaboration and cohesive action at the international level, and so long term risks associated with AI cannot be mitigated. If a country such as China does not enact regulations on AI while others do, it would likely gain a competitive edge in terms of AI advancements and deployments.
2. This unregulated progress can lead to the development of AI systems that may be misaligned with global ethical standards, creating a risk of unforeseen and potentially irreversible consequences. This could result in destabilisation and conflict, undermining international peace and security.

The Dangers of Military AI:

1. Furthermore, the confluence of technology with warfare amplifies long term risks. The international community has formed treaties such as the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) to manage such potent technologies, demonstrating that establishing global norms for AI in warfare is a pressing but attainable goal.
2. Treaties such as the Chemical Weapons Convention are further examples of international accord in restricting hazardous technologies.

Conclusion:

Nations must delineate where AI deployment is unacceptable and enforce clear norms for its role in warfare. In this evolving landscape of AI risks, the world must remember that our choices today will shape the world we inherit tomorrow.