RSiS S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES
Nanyang Technological University, Singapore

RSIS COMMENTARY

www.rsis.edu.sg

No. 103 – 28 July 2023

The Crucial Need to Improve AI Governance

By Xue Zhang

SYNOPSIS

Unless properly managed, ChatGPT, an artificial intelligence (AI) chatbot, poses threats to privacy and cybersecurity, and might even lead to the rise of superintelligent AI, which could become a menace to humanity's survival. This commentary reviews existing approaches adopted by governments and highlights the roles AI companies and users can play, in improving AI governance.

COMMENTARY

ChatGPT (Generative Pre-trained Transformer) has taken the world by storm with its proficiency to deal with various complex tasks, including comprehending and composing text in the human language. But ChatGPT – created by an American AI research laboratory OpenAI – has also raised concerns about the threats and risks that it poses, which present challenges for AI governance. What are these concerns?

Firstly, ChatGPT, an AI language model, was trained using databases scraped from the Internet. A massive 570GB of data or 300 billion words were fed into it for model training. Given that more data will provide for a better model, OpenAI's hunger for data created a "privacy black hole".

Italy was the first western country to block ChatGPT, albeit temporarily, over privacy concerns. This led other European countries, such as France, Ireland and Germany, to deliberate whether greater control was required for AI chatbots. In Canada, an investigation into OpenAI was launched in response to a complaint over personal data breaches.

Secondly, although AI has long been a concern for cybersecurity, ChatGPT has brought about new risks. ChatGPT's sophisticated natural language ability enables AI-generated phishing scams to become more convincing and effective. It can also be made to write malicious code for hacking purposes. Furthermore, ChatGPT can be manipulated to provide biased and distorted perceptions and be used as a tool for foreign interference and terrorist propaganda.

Thirdly, the rise of superintelligent AI, with the potential to break free from control, may create a significant threat to human beings. If it is not properly constrained, superintelligent AI could pursue goals not in line with human values, leading to destructive consequences, or even be a threat to humanity's survival.

Should AI's Development Be Stifled?

In view of the risks and threats posed by AI and the critical need to be in control of it, a group of tech leaders including Elon Musk signed an open letter in March 2023 calling for a six-month pause on AI development. But should AI's further development be stifled?

It is often the case that the development of technology is accompanied by unpredicted or unintended side effects. The invention of the Internet, for instance, has improved many facets of our lives, including revolutionised ways of communication, collaboration and entertainment, as well as easier access to information and essential services. However, it has also raised concerns over cyber privacy and security, as well as new issues such as internet addiction, cybercrime, cyberbullying and online falsehoods.

However, the wheel of history cannot be turned back or stopped. The proper development, deployment, and use of AI applications can automate monotonous administrative tasks, streamline processes, reduce labour cost, improve efficiency, and enable workers to focus on creative and high value work. It can also greatly benefit society in various areas such as in transportation, healthcare, research and education.

US tech giants such as Microsoft, Meta, Apple, etc., have scrambled to pursue AI dominance while China's leading tech corporations such as Baidu, Tencent, Alibaba, etc., have announced plans to develop ChatGPT rivals. The chatbot is also back in Italy after OpenAI addressed regulators' demands for privacy protection.

This is the way to go. Instead of giving up a technology that is transforming many aspects of life for the better, it is time for us to improve AI governance.

AI Governance: What Governments Have Done So Far

Governments' approaches to AI governance have been different. The European Union's (EU) Artificial Intelligence Act of 2021 positions it as a leader in AI regulation. Following a risk-based approach, the Act differentiates the uses of AI based on low or minimal risk, high risk and unacceptable risk levels. Several concerns about the Act, for example, regarding its enforcement, lack of flexibility, and its neglect to ensure meaningful accountability and transparency, were noted in analyses conducted by various institutions and organisations.

Moreover, ChatGPT, as a powerful language model without intended purpose and scale of adoption, has presented significant challenges to the current AI Act approach, in terms of the unpredictability of potential risks, the feasibility of categorising generative AI systems based on their risk levels, as well as concerns over private risk ordering.

In contrast to the EU, a higher level of hesitancy was noted in the US in its introduction of legal regulations on AI. Steady progress has since been made following the passing of the National AI Initiative Act of 2020. In addition to existing laws related to automated decision making, which can be repositioned to cover AI, the Blueprint for an AI Bill of Rights published in October 2022 outlines five principles to safeguard the individual's rights.

In May 2022, to ensure a balance between tech innovation and government regulation, the Singapore government launched the world's first AI Governance Testing Framework and Toolkit – AI Verify, which was offered to organisations for self-assessment and self-regulation. However, local experts have pointed out that such ethical framework and principles may not be properly translated to developers, besides the difficulty of ensuring the compliance of all organisations.

In China, there has been a notable increase in the attention policymakers give to AI governance. In the development plan for the new generation of AI released in 2017, the State Council laid out a framework for formulating laws, regulations, and ethical norms to promote AI development, and for establishing an AI safety supervision and evaluation system.

In January 2023, the Chinese government started enforcing its provisions on deep synthesis administration. In April, the Cyberspace Administration of China published draft measures for the administration of generative AI services, with substantial responsibilities placed on service providers. China's approach allows it to tailor regulatory requirements to particular technical capabilities more precisely.

Historically, regulation has often struggled to keep up with the pace of technology advancements. It is crucial, although challenging, to ensure regulatory adaptability for AI due to its rapid progress, significant impacts across various domains, as well as interaction with other emerging technologies.

A forward-thinking mindset, an enduring commitment, and timely reactions to new threats and risks posed by AI, are essential for planning, evaluating, and regulating AI's further development and penetration into society. Rigorous enforcement also guarantees that laws and regulations are complied with, and those who violate them are investigated, prosecuted, and punished.

Improving AI Governance: What AI Companies and Users can Do

AI companies, besides adhering to responsible and compliance practices, can mitigate potential threats and risks by taking the necessary precautions during the development process to safeguard human values including the need to ensure transparency and accountability, and respect for human rights.

As part of corporate AI governance, continuous staff training is crucial to ensure that AI innovations are created, implemented, and adopted ethically and legally. AI companies may consider outsourcing AI governance auditing and employee education to third parties.

Users themselves can also play a part in AI governance. AI legislation can help to regulate ethics, transparency, and security in data gathering and handling, but it cannot prevent users from voluntarily granting AI systems access to their private data or from sharing their personal information with AI applications like ChatGPT.

Privacy protection also relies on users being able to fully understand the potential risks associated with the disclosure of personal data and make informed consent accordingly.

Users should also be wary and not be too trusting in AI. To ensure proper and ethical use of AI, users should be aware of the inherent limitations in AI systems and the need to adopt a critical eye to cross-check facts, challenge incorrect assumptions, control biased views, and correct falsehoods.

ChatGPT is a neutral app created for the benefit of society, but it could be trained and implanted with biases and even manipulated to work against ethics and human values. Given the ongoing AI gold rush, i.e., the current period of intense interest and investment in AI, there is a need to prioritise AI governance through international cooperation among stakeholders, including policymakers, technology companies, user communities and the public.

Xue Zhang is a Research Fellow at the Centre of Excellence for National Security (CENS), S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore.