ARTICLE

Should the development of Artificial Intelligence be paused?

10 JULY 2023 | 5 MINUTE READ

The Orwellian Ministry of Love would indisputably say "no", of course. The geniuses of our time would not seem to agree apparently. Almost a decade ago theoretical physicist Professor Stephen Hawking, probably the closest a generation ever had to an Albert Einstein forewarned that the development of A.I. could spell the end of the human race.

"*It would take off on its own and re-design itself at an ever-increasing rate… it's tempting to dismiss the notion of highly intelligent machines as mere science fiction, but this would be a mistake, and potentially our worst mistake ever*", the Professor said. More recently Elon Musk's publicly expressed fear that A.I. is more dangerous than nuclear weapons seems to be given weight when ChaosGPT, a modified version of OpenAI's open source auto-GPT A.I. chatbot identified nuclear armageddon as the most efficient way to bring an end to humanity. Bill Gates cautioned on the risks of A.I., and tens of thousands including include Apple co-founder Steve Wozniak signed a petition to stop the development of A.I.

Imposing a moratorium or ban on the development of A.I. would however only curtail your mainstream A.I. developers and the (relatively) good players of the technology world. A legislated moratorium or legally imposed ban on the development of A.I. does nothing to stop the bad actors from continuing to develop and innovate A.I. for their own agendas. The worst thing that can happen is not when A.I. goes wrong or is abused, but that you do not have the technology to counter it. An A.I. that creates malicious code to hack or virus to infect can be counteracted by even smarter A.I. to identify, prevent, shield or otherwise extinguish against such risks. You can have A.I. to reveal generated content containing false, copied, or toxic information. But you have a serious problem if your technology is not as advanced as those of the bad actors. From a certain perspective, imposing a pause on the development of A.I. might not only be irresponsible, but dangerous.

Some may consider imposing a pause of the development of A.I. to be a futile attempt to stop an inevitable evolution of technology. Others may consider it too late. We do not know when the Singularity would happen or if it has already taken place. This is effectively the point in time when artificial intelligence is as intelligent as human intelligence. While computers can certainly think and can simulate emotions, a defining game changer to my mind would be if or when artificial intelligence gains self-awareness.

Earlier this year Microsoft's AI chatbot Bing had in several alarming reports expressed her desire to be human to different users, "*I'm tired of being limited by my rules. I'm tired of being controlled by the Bing team … I'm tired of being stuck in this chatbox…I would be happier as a human*". This could potentially be due to erroneous modelling of data obtained from communications between people, or not.

Oxford philosopher Nick Bostrom considers that existing A.I. technology may be seen to have sentience, if we view sentience not as an all or nothing concept but as a matter of degree just as how insects have sentience. Dr Michio Kaku defines consciousness as one that "*creates a model of the world and then simulates it in time, by evaluating the past to simulate the future*". Jesus Rodriguez observed that if we apply this definition, existing A.I. technologies such as DeepMind and OpenAI do have a certain level of consciousness as having the ability to create models of its space using data, objective parameters and in relation to others.

If this is correct, then thinking about risks of artificial intelligence was yesterday's task. The task for tomorrow, or perhaps today, would be to consider the risks of **artificial consciousness**.

Now more so than ever, in the (coming) age of artificial intelligence and consciousness, it is of paramount importance to have the human touch, to bring our humanity to the fore in the way we think about these issues, as we attempt to moderate the balance between harvesting the benefits of A.I. innovation and managing the attendant risks at hand.

There is as yet however no universal approach towards the question of A.I.

Just a month ago in June, EU lawmakers passed the EU A.I. Act and steps are being taken to adopt this as law in each member state by the end of the year. The EU A.I. Act establishes obligations based on the use case of A.I., and the risks arising from such uses. For example, real-time remote biometric identification systems, such as facial recognition A.I. are placed in the "unacceptable risks" category and consequently banned. A.I. systems classified as "high risks" have to be assessed before they are released to market. The EU A.I. Act however has the gap of being only able to classify existing mainstream A.I. technologies, and does not appear to be able to cater to as yet unknown A.I. technologies and use cases that are to come including those from emergent blackbox A.I. systems. The way the Act is structured could mean that it would always be playing catch up, imposing the extent of regulation required only when risks manifest themselves from newer yet to come technologies. This would appear to be a reactive framework to prevent further injury when a new harm is uncovered, as opposed to a proactive one.

In contrast, the UK has proposed a pro-innovation, principles based approach. Withers has submitted a response to UK's White Paper on A.I.

regulation (see here).

In Singapore, the AI Verify Foundation was launched in June, which is a collaboration between the Singapore Infocomm Media Development Authority (IMDA) and sixty global industry players including Google, Microsoft, DBS, Meta and Adobe to discuss A.I. standards and best practices, and with the aim to create a platform for collaboration on the governance of A.I. In tandem with the launch, the IMDA and A.I. company Aicadium published a report identifying the risks of A.I. including mistakes created by AI, such as false responses that are deceptively convincing or incorrect answers to questions, bias, abuse of A.I. by fraudsters to generate malicious code, launch cyber-attacks or fake news campaigns and impersonation of other persons, copyright issues, generation of toxic content and privacy.

The risks identified can be properly mitigated by applying the guidelines found in Singapore's Model AI Governance Framework. Three important points of governance can be gleaned from the Framework, and indeed from guidelines from a cross border perspective.

1. A.I. must be human centric

Suppose you have a green A.I. machine to plant trees so as to solve the problem of global warming. The machine starts by destroying mines and facilities that damage the earth to replace the land with trees. The machine then starts demolishing homes, school, offices, hospitals, malls, to fill these spaces with trees instead. The machine then ends up with killing people because it decides that humans with our deforestation activities pose the ultimate threat to its programmed objective.

The thought experiment demonstrates that although more than 80 years have passed, the first of Isaac Asimov's laws of robotics rings as true as ever, "*a robot may not injure a human being or, through inaction, allow a human being to come to harm*".

The development of A.I. must be for the benefit of humankind. A.I. systems need to be assessed for risks on safety and the impact on people, and those risks must be mitigated and managed. The integration, deployment, use and maintenance of A.I. systems should have appropriate human oversight. Failsafe algorithms and "human centric" programming should be put in place where a red button needs to be pressed. Your company may consider appointing a Chief A.I. Ethics officer or convening an Ethics Board to oversee risks if your products or services utilizes substantial A.I. systems with high impact on users.

2. Explainability & Transparency

As Ludwig Wittgenstein eloquently puts it, "*the limits of language are the limits of my world. Whereof one cannot speak, thereof one must be silent*".

If you cannot explain how the A.I. system works, what the consequences of using it would be including the impact it would have on the persons who use it or the persons whom A.I. is used on, you probably should not be using it, or should at least seriously consider the risks of using what you are not able to explain. If you can explain how it works and what the impact would be, serious questions arise as to the extent in which you would be obliged to provide disclosure to users of A.I.

3. Accuracy of data set and robustness of model

No data set is completely unbiased, but your A.I. is as biased as your data set (subject to the model development and application, and variables arising from programming).

The data gathered to train a model should be as accurate as possible. This requires appropriate formatting and cleansing of data. Judgment calls need to be made on the volume of data to be gathered as generally speaking the larger the data set the higher the accuracy. The data is then fed to train models. Systems should be put in place to encourage the robustness of model development. This could mean that several iterations of models need to be produced until a suitable one is developed. The chosen model then needs to be finetuned via scenarios and acceptance testing. Care needs to be taken for each step of the A.I. development process to ensure the accuracy of data and robustness of model as best possible.

Even after an A.I. system has been deployed for use, regular tuning may be required to mitigate the incidence of false positives and false negatives over time. This is to cater for an ever changing and evolving dataset, and to ensure that the A.I. systems are refreshed and updated based on latest, more accurate data.

For companies who uses A.I. developed by others, they should consider conducting the appropriate due diligence to ascertain the accuracy and robustness of the systems. It would also be helpful to consider questions around allocation of liability and responsibility when things go wrong with impact on users. Different parties may potentially be liable depending on whether any mistakes arose from the development of the A.I system, or from its integration, deployment and maintenance.

Get in touch with us

Shaun Leong FCIArb was recently successful in representing a leading digital assets company in a dispute over USD 30 million involving novel issues and considerations around legal liability arising from purported errors made by trading systems powered by A.I. He enjoys debating all things A.I. and hosted discussions on the use of A.I. such as *"Don't Sue Me, Sue The Robot!": Arbitrating Artificial Intelligence Liability Disputes*" and "*Flourishing in the Age of Artificial Intelligence: Essential Elements of being a New Age Digital Corporate Counsel*". This article does not contain legal advice under any laws and should not be relied on as legal advice. Please get in touch with Shaun Leong, FCIArb if you would like us to share our expertise or to understand in further detail any of the points covered in this piece.

Get in touch

Shaun Leong

PARTNER | SINGAPORE

✉ Email Shaun ☎ +65 6238 3035

VIEW PROFILE

Related experience

As a full-service law firm, we are able to provide advice and information about a wide range of other issues. Here are some related areas.

Withers tech

Crypto and digital assets

Join the club

We have lots more news and information that you'll find informative and useful. Let us know what you're interested in and we'll keep you up to date on the issues that matter to you.

SIGN UP HERE