BRIEFING NOTES

BN-69-Space and Cyberspace-Aug2021

AI, CYBERSECURITY, CYBER SPACE, AND ETHICS AND PRIVACY

Authors: Mohamadreza Nematollahi[1] and Kash Khorasani[2]

1 Graduate student, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada

2 Professor, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada

SUMMARY

- The explosive expansion and social integration of artificial intelligence (AI) has raised numerous concerns and dilemmas regarding human autonomy and safeguarding society against AI.
- It is essential to scrutinize the human-AI relationship, as well as the nature and the role of trust in it, this is while there still remains a gap between the present form of AI technology and the requirement for AI to collaborate as equal to humans.
- AI solutions require to be interpretable, transparent, and explainable for humans to be able to understand the AI solutions' intentions and develop a mutual predictability and shared understanding.
- One needs to ensure that technology matches human values at all sectors, such as government, business, academia, and individual choices.
- One needs to properly assess and decide the level of autonomy of AI products, and as we approach to become fully automated, the ethical questions regarding AI solutions and autonomous AI systems become more crucial.
- AI powered cyberspace, increase the sovereignty gap already caused by misuse of cyber technologies and a cyber-enabled exercise of influence by non-state actors that will add to many challenges of strategic autonomy problems already existing in the digital age.
- Powering cyberspace by AI from the view point of risk management approach to strategic autonomy sheds light on a great many ethical issues such as the "erosion of individual autonomy, unfair allocation of liability, the fallacy of human in the loop, the contestable ethics of mass surveillance and of trading off individual casualties versus collective protection".
- Even with effective transparency, AI due to lack of security by design, autonomy by design and privacy by design remain vulnerable to cybersecurity threats such as data poisoning and bias injection.
- AI has the potential to improve the porous defenses and help reduce cyber-crimes. AI can improve cyber-security through improving system robustness, system resilience, and system responses. However, it also ironically facilitates the escalation process of attacks and exploitation of potential vulnerabilities.
- Although AI itself can be used for automating software testing and can bring self-healing capabilities for productions, yet, it is still imprudent to grant such full autonomy to AI as it may increase ethical risks.

CONTEXT & CONSIDERATIONS

- While AI offers countless advantages to human society such as saving time, money and lives, as well as providing individuals with the prospect of a more-customized future, it also introduces threats to human autonomy, agency and capabilities. In other words, it raises concerns regarding the possibility and consequences that computers exceed and undermine human intelligence on tasks such as complex decision-making, reasoning, learning visual acuity, pattern and speech recognition, to name a few.
- Code-driven tools diminish human agency by sacrificing independence, privacy and power over choice, as they become more prevalent and complex. At the same time, such autonomous systems can reduce or eliminate the need for human involvement in certain tasks.
- Some experts believe that too much dependency on AI will in the long run erode humans' abilities to think independently or to interact effectively without relying on automated systems. Also, some experts go as far as believing that the expansion of code-based machines and the accelerated growth of autonomous military applications and the use of weaponized information, could result in the erosion of sociopolitical structures and possibility of great loss of lives.
- AI tools to certain extent have been employed and supervised by companies and governments who only seek profit and power, and human values and ethics have been widely overlooked. It is crucial that we attempt to ensure that technology matches human values at all levels, as in governments, businesses, academia, and individual choices.
- Although AI will be used to make world a better place by eliminating poverty, improving health, and providing better education, its superhuman performance will definitely allow increasingly concentrated accumulation of wealth and power, leaving many behind.
- The issue of trust is of utmost importance for social interactions. In this context the terms trustor and the trustee play key roles. These roles are applicable to humans and to AI in carrying out a given task.
- For each trustee, automation level defines the level of capabilities in performing the task, while autonomy of the trustee defines the opportunities, and is somehow proportional to the level of trust on the trustee.
- Severity of risks of AI are proportional to the level of autonomy of AI solutions, and the level of autonomy should be aligned with the level of automation of each product or solution.
- AI systems have been given an opportunity on the degree of freedom by which they are allowed to perform by the human trustor, as well as the authority delegated to them. This could be equivalent to an opportunity for the trustee to defect. In other words, "if the intent of the AI is not aligned with that of the human, the AI is likely to make decisions that disappoint the human and cause the human to suspect the intent of the AI, leading

to a situation of human mistrust regardless of the AI's level of automation and level of autonomy".

- Ethical questions and dilemmas have been raised again and even become bolder as AI and cyberspace and cybersecurity merge in an emerging and disruptive technology (EDT) developing context.
- Through a misuse of cyber technologies and a cyber-enabled exercise of influence by non-state actors, 'cyber' has created a 'sovereignty gap' which consequently disrupts and alters the balance of power in the traditional state-based system of international relations.
- Strategic autonomy is a tool to sovereignty. Traditionally, strategic autonomy was mostly regarded as a military and defense domain term. However, it recently refers to much broader criteria such as the economy, society, and democracy.
- In a more general manner strategic planning is defined to be "the ability, in terms of capacity and capabilities, to decide and act upon essential aspects of one's longer-term future in the economy, society and their institutions".
- Policy makers face new challenges in strategic autonomy in the digital age due to the above reasons of possible misuse. This is while adding the AI emphasis has made it even more challenging.
- The AI lacks transparency in how it approaches the decision-making problem and temptation of granting responsibility of an operator onto 'the system' can result in extensive monitoring of members of society. These could be considered intrusive and coercive in the sense that they promote the feeling that their sense of being in control is becoming more and more elusive. Under such circumstances one's sense of autonomy turns out to be fragile and insubstantial.
- The AI interaction with cyberspace goes beyond what was discussed, and can be used for shaping the future of the cyberspace as well. In this regard, among the many prevailing questions, the questions regarding the cybersecurity are of outmost importance.
- The AI has the potential to improve the porous defenses and can help reduce cyber-crimes. AI can improve the cyber-security through improving system robustness, system resilience, and system responses. However, given that AI does enable better targeting, faster, and more impactful attacks through recognizing and detecting the vulnerabilities of a system, it ironically facilitates the escalation process of attacks and the exploitation of potential vulnerabilities.
- The AI can take software testing to a new level by creating the capability of self-testing and self-healing in software design, which enables the AI software to verify and validate itself and to become more robust. However, granting full autonomy to the AI in this sense is imprudent as it may increase the concerned ethical risks.
- The AI also has been employed for threat and anomaly detection to improve the system resilience. For this purpose, the AI often tracks and monitors human data and their device interactions. It also monitors their behavior and generates biometric profiles. Such

extensive monitoring and comprehensive data collection adversely impact human subjects' privacy and increase the risk for the breach of data confidentiality through cyber-attacks. It also creates a mass-surveillance effect which is mostly undesirable. Therefore, in this sense AI behaves as a double edge sword.

- Machine learning algorithms are trained and developed by using huge datasets and are heavily dependent on large data collections in order to achieve their recognition and detection capabilities, which ultimately makes them vulnerable to unintended bias or intended biases and data positioning by the adversaries.
- Due to the growth of AI and the rapid advancements in this field "citizens will face increased vulnerabilities, such as exposure to cybercrime and cyberwarfare that spins out of control, and the possibility that essential organizations are endangered by weaponized information".
- The worst-case scenario might be the occurrence of unrestricted military development in which machines take over and destructive weapons will become more readily accessible. Therefore, it is vital and imperative that particular AI ethics be developed and guaranteed.

RECOMMENDATIONS:

- It is essential to scrutinize the human-AI relationship, as well as the nature and the role of trust in it. Research suggests that there still remains a gap between the present form of AI technology and the requirement for AI to collaborate as equal to humans. Thus, "the research community is still developing solutions for AI to be, interpretable, transparent and explainable to allow humans to understand the intention of an AI system and develop mutual predictability and shared understanding".
- Proper level of autonomy for each AI solution should be evaluated and be used proportional to the AI automation capabilities and risks.
- Success in integrating AI with society highly depends on ensuring accountability and developing transparency solutions and regulations.
- Even effective transparency fails to address data input, storage and transmission, as it is still impossible to clearly describe how a neural network-based AI system reaches a decision. Therefore, since it lacks security-by-design, autonomy-by-design and privacy-by-design it remains vulnerable to cybersecurity threats such as data bias and data poisoning.
- Practical research work is needed in information exchange covering the entire AI chain, from pre-AI data capture to AI processing to post-AI explainability of algorithms.
- For the global common good route to be kept open, ethical guidelines of AI as applied to cybersecurity must be examined and re-evaluated. It is also necessary to address the issue of ethical certification for weaponized AI and transparency of automated decision-making, which require private-public collaboration and intergovernmental work nationally as well as internationally to refine ethics in cyberspace.

- As discussed above, AI systems from cybersecurity perspective in the sense that what they can bring to this field and what are the risks, are like a double-edged sword. Therefore, as with the pervasive distribution and fast-paced execution of AI systems unforeseen consequences increase and AI advantages become less effective.
- Developing and enforcing regulations and policies are essential to ensure proportionality of responses to AI benefits and risks, legitimate targets, and responsible behavior both in private and public sectors. It is also necessary to establish an authority with the ability to convene national and international policies and norms regarding responsible state behavior and compliance in cyberspace.

REFERENCES

[1] Hussein A. Abbass (2019), "Social integration of Artificial Intelligence: Functions, Automation, Allocation logic and human-autonomy trust" Cognitive Computation 11: 159-171, published online, Springer Nature

[2] Timmers, Paul (2019), "Ethics of AI and cybersecurity when sovereignty is at stake" Minds and Machines 29:635-645, published online, Springer Nature

[3] Taddeo, Mariarosaria (2019), "Three ethical challenges of applications of artificial intelligence in cybersecurity", Minds and Machines 29:187-191, published online, Springer Nature

[4] Gearheart, Frank (2020), "The ethical use of machine learning in cybersecurity", 14-ISSA Journal

[5] https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/