MINDS
MOBILISATION DES IDÉES NOUVELLES EN MATIÈRE DE DÉFENSE ET DE SÉCURITÉ

Security-Policy NET Nexus of Emerging Technologies

# NOTE FOR NATIONAL DEFENCE: AI-Cyberspace, Cybersecurity and Ethics

**Authors:** M. R. Nematollahi [1] and K. 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 unintended use of AI.

- It is essential to scrutinize the human-AI relationship, as well as the nature and 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 solutions to collaborate and operate as equals to humans.

- AI solutions require to be interpretable, transparent, and explainable for humans to be able to understand the AI solution's intentions in order to 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 on the level of autonomy of AI products, and as we approach to fully automated or autonomous AI systems, the ethical question regarding AI solutions become even more crucial.

- The AI powered cyberspace and the increase in the sovereignty gap that has already caused by misuse of cyber technologies and cyber-enabled exercises of influence by non-state actors, will add to the many challenges for the strategic autonomy problem that has already existed in the digital age.

- Powering cyberspace by the AI systems from the point of view of a 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", to just name a few.

- Even with effective transparency, AI systems due to lack of security by design, autonomy by design and privacy by design remain vulnerable to cybersecurity threats such as the data poisoning and bias injection.

- AI systems have the potential to improve the porous defences and to facilitate and enable reduction of cyber-crimes. AI systems can improve cyber-security through improving the system robustness, system resilience, and system responses. However, they also ironically facilitate the escalation process of cyber-attacks and exploitation of existing potential vulnerabilities.

- Although AI systems can be used for automating software testing and can bring self-healing capabilities for their production, yet, they are still imprudent to grant such autonomy to AI as this may increase the ethical risks.

## CONTEXT

- Although AI offers countless advantages to human society such as saving time, money and lives, as well as providing individuals with prospect of a more-customized future, it also introduces threats to human autonomy, agency and capabilities. In other words, they raise 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 some tasks.

- Some experts believe that too much dependency on AI will in the long run erode humans' capabilities 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 erosion of socio-political structures and possibility of great loss of lives.

- AI tools are mostly employed and supervised by companies and entities who only seek profit and power, and human values and ethics have been widely overlooked. It is crucial that one attempts to ensure that technology matches human values at all levels, such as government, business, 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 super-human performance will definitely allow increasingly concentrated 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 the AI to carry 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.

- The severity of risks of AI systems 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 has been given an opportunity that is the degree of freedom by which it is allowed to perform by the human trustor, as well as the authority delegated to it. 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 are raised again and even become bolder as AI and cyberspace and cybersecurity merge in a 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, nowadays, it refers to much broader criteria such as the economy, society, and the democracy.

- In a more general manner strategic planning defines 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 faced new challenges in strategic autonomy in the digital age due to the above reasons of possible misuse. This is while adding the functionalities of AI makes it even more challenging.

- AI lacks transparency in how it approaches the problem of decision-making and temptation of granting responsibility of an operator onto 'the system' can results in extensive monitoring for members of society that could be intrusive and coercive in the sense that it promotes 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.

- AI interaction with cyberspace goes beyond what it has been discussed, and can be used for shaping the future of the cyberspace as well. In this regard, among the many questions, the questions regarding the cybersecurity are of outmost importance.

- AI has the potential to improve the porous defenses and to help reduce cyber-crimes. AI can improve cyber-security through improving system robustness, system resilience, and system responses. However, as AI enables better targeted, faster, and more impactful attacks through recognizing and detecting the vulnerabilities of systems, it ironically facilitates the escalation process of attacks and exploitation of potential vulnerabilities.

- AI can take software testing to a new level by creating the capability of self-testing and self-healing in software design, which enables them to verify and validate software and to become more robust. However, granting full autonomy to the AI in this sense is imprudent as it may increase ethical risks.

- AI also has been employed for threat and anomaly detection to improve system resilience. For this purpose, 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 increases the risk for 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-edged sword.

- Machine learning algorithms are trained and developed by using large datasets and heavily depend on data collections in order to progress in their recognition and detection capabilities, which ultimately makes them vulnerable to unintended bias or intended biases and data positioning by adversaries.

- By the growth of AI and 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 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 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.

- Even effective transparency fails to address data input, storage and transmission, as it is still impossible to clearly describe how neural network-based AI 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 work is needed in information exchange covering the entire AI system 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 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 requires private-public collaboration and inter-governmental work at the UN to refine ethics in cyberspace.

- As discussed, the AI systems from cybersecurity perspectives can bring benefits to this field as well as risks, therefore it is like a double-edged sword. With the pervasive distribution and fast-paced execution of AI systems, unforeseen consequences increase and AI advantages become less measurable.

- Developing and enforcing regulations and policies is 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 capability to convene international policies and norms regarding responsible state behavior and compliance in the cyberspace.

## References

- 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
- Timmers, Paul (2019), “Ethics of AI and cybersecurity when sovereignty is at stake” Minds and Machines 29:635-645, published online, Springer Nature
- Taddeo, Mariarosaria (2019), “Three ethical challenges of applications of artificial intelligence in cybersecurity”, Minds and Machines 29:187-191, published online, Springer Nature
- Gearheart, Frank (2020), “The ethical use of machine learning in cybersecurity”, 14-ISSA Journal
- https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/