Visiting address
Turfmarkt 147
2511 DP The Hague

Postal address
PO Box 20011
2500 EA The Hague

I www.cybersecurityraad.nl
T +31 (0)70 751 5333 (secretariat)
E info@cybersecurityraad.nl

Ministry of the Interior and Kingdom Relations
Attn: Ms A.C. van Huffelen
PO Box 20011
2500 EA The Hague

Date
25 July 2024

Subject
Informative letter from the Cyber Security Council about classical AI, generative AI and cybersecurity

Your Excellency,

During your visit to the Cyber Security Council ('the Council') on 15 June 2023, we discussed the risks associated with the use of artificial intelligence (AI). Generative AI's implications, in particular, require special consideration. You invited the Council to provide additional information on the risks posed by AI in general, and generative AI specifically.

Generative AI, in particular, can increasingly automate tasks traditionally performed by humans. This also applies to cybersecurity: several large-scale application options exist for both offensive and defensive strategies. However, this presents significant security risks. As a result, regulation and ongoing focus on human decision-making are essential. AI developments also offer many opportunities for cybersecurity, including automated defence through autonomous systems or detecting malicious software. The Council will elaborate on these opportunities and risks and their social consequences in the remainder of this letter.

Analysis

Because of the opportunities that generative AI offers to society, its introduction is often described as being a game changer. However, the field of AI has existed for decades and generative AI builds on previous technological innovations. Unlike other forms of AI ('classical AI'), generative AI is a content creator in its own right. This aspect specifically, along with making the technology available to large user groups, has in fact been the game changer. Appendix 1 to this letter further explains and delineates the various terms according to the most commonly used definitions today.

These developments are paralleled by an escalation in cybersecurity threats. In a broader sense, public values such as data and other forms of privacy, transparency[1], copyright, equal treatment and the functioning of the democratic rule of law could come under pressure. Using responsible AI must therefore be prioritised in applications.

The rapid rise and implementation of generative AI has already sparked public debate about the risks posed by this phenomenon. Recently, it was also the subject of a cabinet session at the PM's official residence (Catshuis), partly at your initiative. The focus is mainly on the risks to these public values, while the cybersecurity implications are still not well recognised and understood by society.

[1] More specifically, AI use involves interpretability, avoiding bias, explicability and fairness.

The speed at which AI applications are evolving today can significantly change the cybersecurity landscape, both in terms of opportunities and risks. Generative AI, for example, can automatically scan networks for vulnerabilities, making it easier to launch large-scale digital attacks. Major consequences for vital infrastructure may arise, especially if human intervention is not swift enough to counter the threat. This could lead to transformative change, with the defensive side becoming dependent on generative or other AI applications to manage these threats.

Specific AI opportunities in cybersecurity

The rise of generative AI has further accelerated the existing interest in AI's cybersecurity capabilities. The opportunities are reflected, for example, in these specific application areas:

a) AI applications enable organisations to automatically detect attacks by detecting anomalies in their network. Generative AI can then automatically generate analyses and distribute 'Indicators of Compromise' so that action can be taken on them.
b) AI can also help analysts in Security Operations Centres (SOCs) combine security alerts from multiple sources to improve attack detection. And going one step further, autonomous systems based on AI make it possible to set up 'smart' SOCs with limited human intervention.
c) Generative AI can be used to develop advisories. These generic advisory messages are sent to organisations or organisational units and provide guidance on mitigating recently identified, widely known vulnerabilities. This can also be a task of 'smart' SOCs, as mentioned in b).
d) AI applications make it possible to automatically detect malicious software, correlate variants and make predictions about new forms and variations of such malware. Organisations can then proactively protect themselves.
e) By using code generators, Generative AI enables developers to write code much faster than before, potentially making it a valuable tool for developing secure software.
f) AI applications can automatically detect and fix vulnerabilities in all types of software implementations.

Conditions for responsible use

The advantage of all these opportunities is that human effort can remain limited. While this is not a definitive solution, it is an important benefit in an era of severe cybersecurity skills shortages. Three observations are appropriate here:

- Generative AI operates (now and in the foreseeable future) on the basis of probabilities and has no concept of what is true or false[2]. This means that generative AI works best in the hands of an expert who can evaluate the output and adjust the results accordingly. Given the current state of AI technology, this is an essential prerequisite for responsible use.
- The quality of the underlying models, the data on which they are based, and their intended use is critical for all applications, defensive and otherwise. Digital attackers do not face this inhibiting effect and can therefore benefit quicker from generative AI.
- Efforts towards the responsible deployment of applications in specific environments, as mentioned under a), b), and c), are underway in both the central government and business sector. The foregoing conditions also apply in this case.
  Note: a potential ban on civil servants using generative or other AI, as recently reported in the news, is at odds with these developments. The Council therefore does not support such a blanket ban.

[2] Popularly known as 'hallucinating' AI, this is an integral part of generative AI; incorrect results in these situations are thus not due to software errors.

Specific AI risks in cybersecurity

Below is a list of the currently known cybersecurity risks of using both generative and classical AI. The implications for cybersecurity are becoming increasingly clear, especially with recent developments in generative AI. The CSAN 2023[3] also discusses this in detail.

a) AI applications facilitate cyberattacks by enabling the large-scale automatic exploitation of existing vulnerabilities. This applies to both newly discovered vulnerabilities (known as zero-days) and known vulnerabilities that have not yet been patched.
b) The latest generative AI technologies allow cybercriminals to apply existing modus operandi more easily, at greater scale, and with higher quality. Examples include generating deepfakes, indistinguishable spam and phishing emails, and automatically creating malware in various variations.
c) Tooling for classical and generative AI is often widely deployed across various organisations and networks. Those environments are not always well-shielded, technically or otherwise, for internal use or segmented from other operational networks. This increases the attack surface, also increasing the risks of intrusion, manipulation and data breaches.
d) The widespread use of AI software also poses general cybersecurity risks. Precisely because of the rapid – and sometimes overly hasty – implementation, the AI software has not been extensively tested and checked for unknown errors in all cases. The authenticity and integrity of datasets and training programmes is also not always guaranteed. Additional safeguards are required in line with the EU-mandated Cyber Resilience Act (CRA) and Digital Services Act (DSA).

Aside from the above risks, some of the opportunities of classical and generative AI for cybersecurity, as mentioned above, also have a downside. As a result, these also create new risks:

e) Regarding opportunities a, b and c: Carelessly using AI in the cybersecurity workflow increases the likelihood of misinterpretations (both false positives and false negatives), erroneous responses, or undesirable decisions. This risk increases if there is over-reliance on these AI applications without active human oversight, while the proper functioning or appropriate application of the AI software itself is not guaranteed.
f) Regarding opportunity e: It is becoming increasingly evident that code generated using AI, particularly generative AI, still contains numerous vulnerabilities. If the overarching use of such modules becomes standard in many software applications, this will increase the risk of mass exploitation. This was previously observed, for example, in generically applied software modules, such as Log4j. Addressing this requires additional cybersecurity controls, such as those arising from the aforementioned CRA and DSA.

International developments

The opportunities and risks of AI, particularly generative AI, are also receiving wide attention internationally. For example, the UN Secretary-General recently established the new High-Level Advisory Body on AI[4]. This group of experts from all walks of life will advise on international AI governance. The chosen interdisciplinary approach and multi-stakeholder strategy also apply to cybersecurity issues in the context of AI. The Council recommends also paying attention to cybersecurity aspects for AI applications and vice versa in this environment.

[3] Cyber Security Assessment Netherlands 2023, National Coordinator for Counterterrorism and Security (NCTV), June 2023
National Cybersecurity Strategy 2023 | National Coordinator for Security and Counterterrorism (nctv.nl)
[4] See: High-Level Advisory Body on Artificial Intelligence | Office of the Secretary-General's Envoy on Technology (un.org)

This should also include considering the AI approaches of global leaders to digital resilience, such as the shaping of strategic AI policy and the necessary regulation that has already started. For example, the US and UK governments recently published new Guidelines for Secure AI System Development[5]. The guidelines provide suggestions and solutions to enable data scientists, developers, managers, decision-makers and risk owners to make informed decisions about the safe design, development, implementation and operation of their AI systems. As a recent survey shows, many countries have now also released national regulations aimed at ensuring privacy[6].

Given the cross-border nature of AI technology, further European cooperation is also necessary. For example, the European Commission recently recommended that Member States conduct risk assessments in four areas crucial for EU economic security, including the AI domain[7]. The Netherlands should also firmly commit to such an AI risk assessment.

The European Union Agency for Cybersecurity (ENISA) has already released several AI publications, focusing on both mitigating cybersecurity risks and enabling secure and reliable AI applications. These publications supplement the European AI Act, which paves the way for regulating AI applications, and the aforementioned CRA, which sets detailed requirements for secure digital products and services. In doing this, ENISA is committed to best practices and standardisation: 'Multilayer Framework for Good Cybersecurity Practices for AI'[8] and 'Cybersecurity of AI and Standardisation'[9].

Additional issues

Concerns have long been raised about the reliability of AI, including in the area of responsible AI. These concerns include the aforementioned authenticity and integrity of the underlying data, datasets and algorithms, and the technical and other measures necessary to ensure such reliability.
In conjunction with the CRA, the European AI Act[10] contains provisions that, while addressing cybersecurity issues relating to AI, are primarily concerned with the implications of specific applications. For example, there are rules about transparency, such as disclosing in advance that you are dealing with a chatbot or that something is a deepfake. Part of this includes providing information about the data used to train the AI model in question. Particularly in the case of generative AI, this requires special attention, including in relation to cybersecurity risks.

A corollary issue to be addressed is the additional knowledge required about AI, particularly generative AI, for application in the cyber domain. This knowledge needs to be developed among cybersecurity specialists (who are already in short supply), but also created through a crossover effect by bringing in broader AI experts. Conversely, we also face a severe shortage of qualified AI experts who actually understand the use of generative and classical AI as a cybersecurity tool.

Alongside plans and ambitions for new AI applications and the need for integrated policy development, there must also be sufficient focus on regulatory implementation. The capacity of regulators to address AI-related issues, particularly in the context of cybersecurity, is a concern. Given the significant increase in their workload,

[5] See https://www.cisa.gov/news-events/alerts/2023/11/26/cisa-and-uk-ncsc-unveil-joint-guidelines-secure-ai-system-development
[6] See KMPG report: Privacy in the new world of AI (kpmg.com)
[7] See Commission recommends carrying out risk assessments on four critical technology areas: advanced semiconductors, artificial intelligence, quantum, biotechnologies | Shaping Europe's digital future (europa.eu)
[8] See Multilayer Framework for Good Cybersecurity Practices for AI — ENISA (europa.eu)
[9] See Cybersecurity of AI and Standardisation - ENISA (europa.eu)
[10] The EU recently reached internal consensus on implementing this new Act, which will enter into force in the spring of 2024.

consideration should be given to allocating additional financial resources to these regulators. The Council makes the following recommendations:

- Ensure good cooperation between regulators. Partly in view of the connection between cybersecurity and AI, it is essential that the designated regulator for each sector also focuses on AI applications. This is preferable to having a single, overarching regulator focus on AI across all sectors.
- Create a broad pool of experts, including cybersecurity specialists, that all regulators can use. This will enhance the crossover effects mentioned above.
- Continue encouraging data-driven work in the area of supervision, which can also lead to greater efficiency here.

Final remarks

In line with the risks identified above and the additional issues, it is crucial to also keep monitoring the cybersecurity implications of new AI applications in the near future. The Council explicitly calls for special attention to setting up 'guard rails' in this area. Rather than solely introducing additional regulation, this involves a continued focus on the importance of human assessment, interpretation and adaptation of AI output, where needed, when making substantial decisions or taking impactful actions, including in cybersecurity.

Technical measures to create digitally secure environments for AI applications should also be promoted, combined with aiming for quality in the AI software itself, the underlying models and the data. Existing regulations need to be supplemented, both to ensure the safety of AI software itself and that of software created with generative AI. Intensifying cooperation with major technology companies, including at European level, is also necessary, to facilitate making digital AI products and services more secure.

By mitigating specific cybersecurity risks, creating awareness and ensuring transparency about how AI is applied, confidence in the use of AI will be able to grow. If continued innovation is encouraged in the areas mentioned, this technology can be responsibly used across all segments of our society.

A copy of this letter will also be sent to the Ministers of Justice and Security, and Economic Affairs and Climate Policy.

Yours faithfully,
On behalf of the Cyber Security Council,

Pieter-Jaap Aalbersberg
CSR co-chair

Theo Henrar
Acting CSR co-chair

Appendix 1: Background on generative AI versus 'classical' AI

The Netherlands Scientific Council for Government Policy [*Wetenschappelijke Raad voor het Regeringsbeleid*, WRR] states that Artificial Intelligence (AI) cannot be captured in a single definition. Generally, the term AI refers to the following: '*The kind of systems that exhibit intelligent behaviour by analysing their environment and taking action with some degree of autonomy to achieve specific goals.'*

The working title 'classical' AI refers here to building and using systems that exhibit some degree of 'intelligent behaviour', often by detecting patterns in large amounts of data, and then learning to recognise similar patterns in new data (machine learning). These forms of AI are generally used as a tool to automate running analyses and identifying connections between data, leading, for example, to predictions, based on which people can make informed decisions. Other applications include self-driving vehicles.

Generative AI represents a significant advancement beyond 'classical AI'. It is a new generation of AI systems (but only one form of AI), enabling new content to be generated for a specific application, based on previously entered data sets. Generative AI models are trained on huge amounts of data, which can include language models (Natural Language Models, NLM), but also image (vision) generation models, source code generation models and audio generation models. Recent developments in generative AI have led to a proliferation of new tools and training programmes (including ChatGPT, Bard, DALL-E, Bloom, etc.).