Lernende Systeme

GERMANY'S PLATFORM FOR ARTIFICIAL INTELLIGENCE

Artificial Intelligence and IT Security

White Paper by Jörn Müller-Quade et al.
Working Group IT Security, Privacy, Legal and Ethical Framework

Executive Summary

Technologies based on Artificial Intelligence (AI) are increasingly permeating all spheres of life. The ways in which they can improve the security of IT systems, together with the security of AI systems themselves, are essential for enabling citizens, businesses, politics and public authorities to reap the benefits of advancing digitalisation.

AI systems will play an important role in boosting IT security in the future. For example, machine learning methods can be used to improve the capability of intrusion detection systems, or to distinguish between normal and suspicious activities in networks. AI systems can effectively assist IT staff with security issues and thus counteract the effects of the lack of specialist workers in this field in the short term.

However, AI technologies also offer potential for dual use in the realm of IT security. The machine-learning methods that can detect previously unidentified security loopholes in networks or software systems can also be applied by attackers, who are able to use AI methods and processes to optimise their strategies for attacks or to develop new threats. Although the risk of such threats should not be overplayed, it does provide additional motivation for gaining an upper hand in the technology used in this field of AI application and for heightening developers' and users' awareness of this potential for dual use.

AI systems are increasingly being integrated into processes where security and data protection are crucial. The AI systems themselves therefore need protection against attacks. Their resilience against potential manipulation needs to be increased, and the appropriate protective measures need to be implemented.

The new impetus that AI systems inject into the realm of IT security gives rise to different areas that require action – From helping SMEs to acquire skills in AI and IT security to developing and designing the systems in questions. The authors of this paper are outlining initial approaches to develop solutions for these areas of action.

The approaches outlined below provide an initial estimation by the authors of this paper and will be developed further by Plattform Lernende Systeme.

General areas of action

- The future use of self-learning systems in security-critical applications seems to warrant particularly special care and the **integration of specific protective and defensive measures**. This should involve attending to fundamental security issues and spurring on entirely new defensive concepts to specifically protect learning algorithms and enable long-term, secure use of AI. In relation to this, the methodology of demonstrable security could also be used in connection with AI systems.

- AI systems should be **equipped with a technical fallback level** – in case of malfunctions, attacks on the system, or the system itself demonstrating security-critical behaviour. The reliable operation of the entire system cannot be allowed to be endangered by any such occurrences.

- With regard to the potential for dual use, it is advisable to investigate different ways of providing optimum **protection against AI systems being hijacked**. The first step could be to develop social and legal standards.

- However, it seems rather unlikely that **highly specialised applications** such as side-channel analysis will appear as finished products on the market any time soon. In fact, users of such applications (e.g. in industry, universities or public authorities) should **hone their own expertise**, as required.

- In light of the networked devices including AI components being used by individuals or public authorities, the current focus on operators of critical infrastructures should be expanded to also consider the principles of **security by design** and **security by default** in the research and development conducted on systems with AI components.

- The trustworthiness required from AI systems, including in security-critical contexts, calls for a focus on research and industrial policy with regard to the aspiration of **digital sovereignty**.[1]

Areas of action for politics and public authorities

- Efforts towards **training and attracting IT security experts** need redoubling to remedy the shortage of specialists in this field. This might include covering how to use AI systems for IT security in specialist

1 Digital sovereignty entails "full control over stored and processed data together with the autonomy to decide who has access. It also entails the ability to independently develop, change and check technological components and systems, as well as to add further components. Digital sovereignty is therefore both an important basis for reliable systems and also an absolute prerequisite for independent state action." (Plattform Innovative Digitalisierung der Wirtschaft, 2018: 3.)

training and continuing education and continuously updating curricula to keep pace with state-of-the-art technology.

- A **basic understanding of IT** security should also be suitably integrated into disciplines where this issue is gaining importance with the advance of digitalisation, such as in mechanical engineering.

- For SMEs, it makes sense to create or increase the offerings in existing competence centres, which can be used to expand **skills in using AI systems for IT security** – including the relevant consultancy services. A navigation system for IT security in the context of AI could assist SMEs and help make existing offerings clearer and more accessible.

- With regard to **integrating AI systems into public administration** and the services offered to citizens, the security of these systems plays an important role. For example, if public authorities use chatbots for their services, measures need to be devised and implemented to prevent attackers from overloading the AI components of these services to gain access to personal data.

- The use of AI systems highlights the need for **coherent, preferably global IT security policies**. As a result, international initiatives should also promote responsible action by states in this sector at a global level, such as determined efforts to prevent attacks by hackers emanating from their own territory.

Areas of action for businesses

- Large businesses should purchase and operate commercially available tools such as **intrusion detection systems with AI functionality**, while smaller ones could buy in services for this purpose. It is important to provide the appropriate offerings for SMEs in this regard.

- Acquiring technical abilities and skills for handling AI in the realm of IT security could well be critical to the success of a business. Marketable **AI-assisted security solutions and their continuous further development** are important prerequisites for IT security in German industry.

- Companies should check their **IT security measures and skills** with regard to the future use of AI in this context and potentially undertake efforts to accumulate the appropriate skills.

- In connection with AI, any attacks on the IT systems running in offices and production could in future become more targeted and far more sophisticated. Companies could therefore augment conventional analyses of weak spots and threats with **AI-assisted and self-learning surveillance systems**.

- Just as in terms of IT security, technology is also advancing in AI systems. This therefore calls for a revolving **examination of the intelligent defence measures already in place**, including any AI that is used.

Areas of action in the field of research

- Research is required into how AI systems can be used to support and improve IT security, but also into the security and protection of the AI systems themselves. This calls for instigating or expanding appropriate research activities in this area. This can be supported by **strengthening the networks linking institutions** focusing on AI and/or IT security while also boosting these skill levels in Germany.

- The potential uses that AI systems can be put to for citizens, businesses or public administration in various contexts can only be harnessed if the systems are protected as effectively as possible against manipulation, particularly attacks on their prognoses or learning processes. Intensifying the research into **AI systems' resilience against manipulation** can be expected to generate significant progress.

- The security of AI systems can be increased particularly by conducting **tests for special cases**. Research into techniques specifically aimed at automatically generating unusual input and thus simulating potential attacks could play an important role in this regard.

- It must also be ensured that personal data is protected during the use of AI systems, particularly in areas of application where sensitive data is used for the systems' learning processes, such as in medicine. Continuing research and **development in the field of data protection-compliant learning algorithms** that prevent or hinder the extraction and reconstruction of personal data from the models used in self-learning systems can help encourage the use of AI systems in different areas of application by ensuring they uphold data protection.

- In IT security, as in other areas in which AI systems are used, the **ability to explain the decisions taken by the systems** can be key to their viability. This applies to areas where the people who interact with these systems need to be able to trace and assess the factors that influenced their decisions, particularly in the case of complex neural networks. Research into potential means of achieving this should be spurred on to ensure these systems can be used securely and transparently.

Imprint
Editor: Lernende Systeme – Germany's Platform for Artificial Intelligence | Managing Office | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de | Follow us on Twitter: @LernendeSysteme | Status: April 2019 | Image credit: matejmo / iStock

This executive summary is based on the white paper Artificial Intelligence and IT security – Current situation and solution approaches, Munich, 2019. The authors are members of the working group IT Security, Privacy, Legal and Ethical Framework of Plattform Lernende Systeme. The original version of this publication is available at: https://www.plattform-lernende-systeme.de/publikationen.html

SPONSORED BY THE

Federal Ministry of Education and Research

acatech
NATIONAL ACADEMY OF SCIENCE AND ENGINEERING