PRIF SPOTLIGHT

PEACE RESEARCH INSTITUTE FRANKFURT / LEIBNIZ-INSTITUT HESSISCHE STIFTUNG FRIEDENS- UND KONFLIKTFORSCHUNG

AI FOR ARMS CONTROL

// How Artificial Intelligence Can Foster Verification and Support Arms Control

Time and again, modern technology has enhanced arms control - with satellites, surveillance planes or more potent sensory equipment – to detect traces of forbidden substances. More recently, uncrewed vehicles found their way into the arsenals of arms control inspectors, enhancing verification. A very promising yet more difficult application will be to make use of artificial intelligence (AI) in arms control. However, many people have mixed emotions when it comes to AI, and exaggerated hopes as well as unjustified fears dominate the debate. The aim of this paper is to dispel reservations and, based on small projects, show how AI can be used in a reasonable way to enhance arms control and verification without getting caught up in hype.[1]

Khanh Tran: Cyber specialists via https://cybervisuals.org/, Creative Commons Attribution 4.0.

by Niklas Schörnig

Never has the dictum “arms control is in crisis” been more fitting than it is today. Now, at the beginning of the new decade, there are hardly any relevant arms control regimes left intact. The general mood is one of mistrust of arms control partners, but increasingly also of the international institutions that monitor agreements and prohibitions. It is thus all the more important to use every opportunity to foster legally binding arms control agreements and offer new and more reliable solutions in order to overcome mistrust and reservations. One starting point would be to develop new and improved verification techniques.

While new technologies and more advanced weapon systems pose challenges to arms control, time and again modern technology has also enhanced arms control itself,[2] with satellites, surveillance planes or more potent sensory equipment for detecting traces of radiation or chemical agents serving as prime examples. More recently, “emerging technologies” – especially drones or other uncrewed vehicles –have not only had a tremendous impact on military planning and warfare but are already beginning to enhance verification. The most promising, yet significantly more difficult, application will be to make artificial intelligence (AI) in general and machine learning in particular useful in arms control. This text offers insight into where AI can be of actual help and what has to be taken into account to avoid pitfalls and disappointments. It seeks to inspire arms control experts to develop ideas where AI can be of help in their particular field as well. These ideas may be related to verification in a strict sense, but also to the application of AI in a broader arms control context, or to checking compliance with export guidelines and restrictions set by an export control regime.

Many projects are already in a proof-of-concept phase, demonstrating the potential of AI to support arms control and verification measures in the years to come. Institutions such as the International Atomic Energy Agency have also been debating the potential of AI for the specific purpose of verification for some years

now. Finally, the use of AI in export control – technically not verification – is also very promising, starting with the analysis of X-ray images of containers all the way to the analysis of shipping and trade routes.

The idea of using algorithms for arms control is not new. More than 30 years ago a SIPRI volume on "Arms and Artificial Intelligence" dedicated an entire section (Part IV) to "Applications [of AI] in arms control analysis."[3] However, the AI described in that book was limited, as AI was severely restricted by processing power and the lack of affordable computing memory.

Today's AI is heavily reliant on machine learning (ML), both supervised and unsupervised, where the computer learns to categorise large volumes of training data and establish patterns and subsequently to apply what has been learned to new and previously unknown pieces of data. In contrast to so-called "expert systems" developed in the 1960s, 70s or 80s, these patterns enable computers to identify specific objects in pictures or videos, translate text from one language into another, come up with new and previously undetected categorisations, or even provide solutions to specific problems. Over the last few years, AI has proven itself capable of mastering increasingly complex problems, especially in controlled environments where the rules of the "game" are clear and no surprises lie in wait.

The astonishing way in which Google's AlphaGO defeated GO master Lee Sedol in 2016, to the surprise of all human experts, is a case in point here, as the AI had learned new, unprecedented moves just by repeatedly playing against itself.[4]

However, to be of help in arms control contexts, the AI does not need to "beat" humans. It is sufficient for the AI to humans in the best way possible. While most observers agree that AI will not replace human inspectors any time soon, many ideas are already under consideration where AI can at least assist inspectors, ease their workload, and support them in a significant manner. The following examples show where this could be the case.

Translation and analysis of text

Inspectors very often have to work in the context of foreign languages. English often serves as the lingua franca and some inspectors and other arms control experts acquired at least some proficiency in Russian during the Cold War. The multilateralisation of arms control, however, increases the need to communicate with people from many other countries. But experts who can communicate about very technical matters in Chinese, Farsi or Korean, let alone other languages,

The need for verification in arms control

Arms control is based on the notion that cooperation between antagonists is possible even in areas as sensitive as national security – under certain circumstances. When actors have a unilateral incentive to deceive or cheat as they benefit from an opponent's compliance (for example by arming, when an opponent does not arm), specific mechanisms have to be agreed upon in advance to reduce the likelihood of cheating. These mechanisms are called "verification" or "safeguards," and they are often conducted by specially trained human inspectors.

Many traditional arms controllers agree that "verification needs to be built into an [international arms control] agreement" (Keir/Persbo 2020: 16). However, even dense verification measures offer no 100% assurance against any undetected violation of the treaty. The aims of verification are less ambitious: first, making it significantly more costly to cheat and thus reducing incentives to cheat, and second, acting as an early warning mechanism for detecting actual violations of a treaty before their impact reaches a severe level, ensuring that – due to the early warning – there are enough options to react.

AI and machine learning

According to IT specialists, "artificial intelligence," or AI for short, has made significant progress with the invention of so-called machine learning (ML) algorithms two decades ago. In contrast to older, deterministic variants of AI (e.g., "expert systems"), machine learning algorithms learn to classify different objects within a dataset, either based on human feedback (supervised learning), or completely independently (unsupervised). In a supervised scenario, an algorithm would, for example, be given a dataset of photos showing cats and dogs, pre-classified by a human. The software would "look for" (statistical) similarities and create a model of its own which is capable of identifying a dog or a cat in a new picture. In an unsupervised scenario, the AI would identify clusters on its own, without any prior human input. Based on this method, AI can, for example, identify pictures showing the same person or location within a large database – for example a smartphone gallery.

As there is currently no significant research on AI that is not based on machine learning, and given that the learning is based on complex algorithms, the terms "AI", "ML" and "algorithm" are used interchangeably in this text for the sake of simplicity.

are rare. AI supported translation services like Google Translate or DeepL, on the other hand, have improved tremendously over the last few years and, despite occasional minor errors, many translations have reached a satisfactory level of accuracy. Many users have become accustomed to these services without actually being aware that a complex and powerful AI is being used in the backend of the application. For arms controllers, the ability to instantly translate text might be especially helpful when assessing material such as newspapers, government statements or social media in the public domain, or material gathered in other ways in a language unknown to the technical arms control expert. Algorithms could be specifically trained and optimised for technical language, automatically categorising relevant and irrelevant material and thus making valuable material accessible which has not been accessible until now. More ambitious, yet no longer in the realm of science fiction, are projects aiming at Babel Fish-like qualities, such as the US Defense Advanced Research Projects Agency's (DARPA) *Broad Operational Language Translation* (BOLT) programme. As in the famous Douglas Adams novels, a specific device could be used to simultaneously translate a foreign spoken language into the user's native tongue. But translation is not the only relevant AI application when it comes to text. In a project not related to arms control, the AI helps to transform text from PDfs or images into processable text, adapting and correcting based on different types of layout,[5] and making more dated scans far more usable than in the past.

U.S. Army Sgt. Quran T. Williams unloads a Talon IV robot. The robot is used for reconnaissance and detection in chemical, biological, radiological, and nuclear environments. (New Jersey National Guard photo by Mark C. Olsen via flickrr, license CC BY 2.0.*)*

Analysis of images and film

Verification often entails the analysis of pictures such as images captured by drones or satellites. Has a certain installation been enlarged, or do we see signs of current activity? Surveillance videos could show the entrance to a restricted area and it might be relevant who accessed it over the previous few days. Experts agree that image classification and interpretation has made tremendous advances over the last decade[6] and in some cases experts are already supported by algorithms when they interpret aerial footage. But more applications seem possible. In one project analysts trained an AI to identify running nuclear facilities based on Flickr images.[7] Another project trained AI to distinguish between unproblematic copper mills and proliferation-relevant uranium mills,[8] while yet another AI learned to identify small arms contraband on X-ray pictures of shipping containers. While all these projects are still in a proof-of-concept phase, they show what can be achieved in the future.

But even more applications are possible: Google Vision, for example, can "interpret" the content of pictures and indicate what it recognises, usually objects or places. But it can also rate things such as "vintage" or "retro" clothing. In the future it could be possible to have an image of an unknown device analysed by the algorithm, suggesting its potential use. This would be, for example, helpful for customs staff when confronted with unknown devices – relevant for export controls. Finally, AI can help by analysing hours and hours of video footage and flagging only relevant activity for the human analyst to consider.

Sensory data other than text or images

In the arms control realm, inspectors often rely on measuring devices to determine the presence or absence of hazardous substances or activities. The Comprehensive Test Ban Treaty Organization, for example, monitors seismic activity to detect banned nuclear tests. The high art of experts now is their ability to distinguish between weak and/or distant earthquakes and human-made events such as a nuclear test. As early as 2010 Russel, Vaida and Le Bras argued that machine-learning algorithms "could improve the detection and localisation of low-magnitude events, provide more confidence in the final output, and reduce the load of the human analyst."[9] Other conceivable applications might include the use of seismic or acoustic sensors to monitor movement of military vehicles in peacekeeping operations, with the AI suggesting the number and kind of vehicle.[10]

About the Author

Dr Niklas Schörnig is a senior researcher and head of PRIF's "Research on Emerging Technologies, Order and Stability (rETOS)" research group. His research focusses on emerging military technology, the future of warfare, and arms control.

Contact schoernig@hsfk.de

Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung
Baseler Str. 27–31, 60329 Frankfurt am Main
Postvertriebsstück D 43853, Entgelt bezahlt, ISSN-2512-627X

Conclusion

This – incomplete – list of examples shows that AI has much to offer to arms control experts and verification inspectors. While many projects are still in the proof-of-concept phase, AI is making tremendous advances in the civilian realm and in the work of international institutions such as the IAEA, which closely monitors developments. One problem, however, is that many traditional arms controllers are still unfamiliar with the new approach and are cautious and hesitant as a result. Some caution is definitely in order at this stage. To mention only a few problems: The data on which the algorithms are trained needs to be very well curated. Many examples show that the choice of the training data has a tremendous impact on how well the algorithm performs. However, international organisations often have excellent datasets which could serve as a starter.

It must also be kept in mind that, at the moment at least, AI is still a black box. Even for the programmer of a machine-learning algorithm it is often not clear what correlation the AI uses when establishing the pattern it needs in order to run. As a direct consequence, "explainable AI" has become an important aim and could help to ensure trust in the results produced by the algorithms.[11]

These limitations and hurdles show that, at least for the foreseeable future, AI should only be used to support human inspectors. We are a long way from autonomous AI verification and this form of AI should not be the ultimate aim, especially as there is always the possibility of using complex algorithms for espionage or manipulation. Creating algorithms within multilateral teams in impartial international institutions is one possible solution in this area. But verifying software code is definitely an important issue as well.[12]

But AI can help significantly improve the verification systems we have and weaken the position of those who argue that no reliable verification is possible. Arms control experts should therefore look deeply into the question of how AI can support verification and arms control in their field and not avoid dialogue with AI specialists.[13]

Acknowledgement:
This PRIF SPOTLIGHT is based on and draws heavily from the results of the workshop "Governing the Opportunities and Risks of AI for International Peace and Security: What Role for the EU" co-organized by SIPRI and PRIF on September 8–9, 2020. The workshop was generously supported by the German Foreign Office. The author is indebted to all the participants who shared their knowledge and experience under Chatham House Rule.

References and further reading:
hsfk.de/spotlight0122-ref

DOI 10.48809/prifspot2201

PRIF SPOTLIGHT
The Peace Research Institute Frankfurt (PRIF) is the largest institute for peace research in Germany. PRIF sets out to analyze the causes of violent international and internal conflicts, carrying out research into the conditions necessary for peace and working to spread the concept of peace.

V.i.S.d.P.: Karin Hammer, Öffentlichkeitsarbeit (HSFK), Baseler Straße 27–31, 60329 Frankfurt am Main, Telefon (069) 959104-0, info@hsfk.de, www.hsfk.de.
Design: Anja Feix · Layout: HSFK · Druck: Druckerei Spiegler