Artificial Intelligence and national security

Julien Barnu,
advisor on industry and digital technology, Secrétaire Général de la Défense et de la Sécurité Nationale (SGDSN)

Abstract:
Artificial intelligence (AI) is likely to become a key technology for national defense and security, in particular for cyberdefense and intelligence operations. AI's development in these areas entails taking under consideration aspects such as the need to structure operational data and organize the access to them as a function of their level of confidentiality. It will also be necessary to address the crucial question of assessing how much trust we can place in AI systems.

Cédric Villani (2018) cited defense and national security as priority sectors for developing and using artificial intelligence (AI) in France. For him, this sector could "*represent a comparative advantage for France*".[1] Herein, I would like to make a few remarks, far from exhaustive, about specific characteristics of national security in relation to AI. A few examples will be mentioned from among the assignments of the General Secretariat of Defense and National Security in matters of cyberdefense and intelligence.[2]

Implications and limits of AI for cyberdefense

AI is likely to be a key technological component in the cyberdefense, a field that can, however, be used to illustrate several of its inherent limitations. Let us start by looking at this new technology's implications for the successive phases of cyberdefense in the case of an attack.

Prior to an attack, for the protection of information systems, AI will likely augment the capability of the state and of firms to conduct diagnostic analyses or even adopt low-cost corrective measures, since it will automate the search for vulnerabilities and defects in a network's configuration. Likewise, it will serve to automate assessments of the security of digital products. The overall impact of this technological progress on cybersecurity is subject to discussion however. For one thing, attackers use these same AI tools to massively and automatically detect flaws in a network and in software and digital devices. For another, human flaws and shortcomings (using infected flash drives, opening contagious or malicious attachments to e-mail, etc.) are still (apart from technical defects) the major gateway through which attackers penetrate networks. Furthermore, the widespread use of AI will probably enable assailants to automatically make ever more credible counterfeits of the contents they catch; they need but browse the social networks or penetrate electronic messaging systems. This use of AI to better exploit human flaws might more than make up for anything positive that AI brings to network security. The development of AI in cyberspace might, therefore, have a globally negative impact on a country's cybersecurity.

[1] VILLANI C. (2018) *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne* (Paris: Prime Minister's Office), 235p., available via: https://fichiers.acteurspublics.com/redac/pdf/2018/2018-03-28_Rapport-Villani.pdf; English translation: *For a Meaningful Artificial Intelligence: Towards a French and European Strategy* available at https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

[2] This article has been translated from French by Noal Mellott (Omaha Beach, France).

The phase for which AI holds the most promises is probably the detection of attacks. Current detection technology mainly relies on “signatures” for qualifying the attacker: the IP address of the server from which the attack was launched, malicious code, etc. These signatures, once known, are integrated in detection procedures, such as antivirus software. These procedures cannot, therefore, detect new attempts to launch attacks with unknown characteristics. A solution for this is being worked out by integrating AI (in particular its learning techniques) in detection systems so as to discover anomalies in a network and, therefore, the *modus operandi* of attacks of a previously unobserved sort. Major work in France and elsewhere is being conducted on using AI to detect cyberattacks.

To uses AI to respond to cyberattacks, several caveats must be made. First of all, we cannot rely on a technical analysis, even one made by AI, to identify the perpetrators and blame them for an attack. Of course, AI will help us associate an attack more quickly with its *modus operandi* (*i.e.*, the set of tools and procedures characteristically used by threat actors); but such an association does not provide sufficient grounds for attributing the attack. Many weapons used for attacks are for sale on the Internet and are easy to reuse. Some high-level assailants are even known for introducing in their software fake evidence (that, for example, misleadingly points toward a certain country) so as to cause attribution errors. Once again, by automating the introduction of fake evidence in software and making it ever more credible, AI will probably bolster an assailant’s ability to cover his tracks. Work by human beings and, in particular, information of human origin, will undoubtedly remain just as crucial as it already is for attributing an attack.

For responding to an attack, it is even hard to imagine that AI signals a breakthrough. The decision to respond implies that the attack has been attributed. That is a political decision. Furthermore, the nature of the response will still have to be evaluated case by case as a function of many factors, in particular geopolitical. Nor will the response necessarily take place in cyberspace. It might be diplomatic, political, economic or even military.

A misunderstanding about AI and big data

A major challenge for national intelligence services is to optimally process the volumes of heterogeneous data collected: from online connections, intercepted voice messages, images, satellites, etc. To take up this challenge, the bricks of artificial intelligence can naturally be posed to improve currently used data analytics. This can, for example, help us detect trends or singular events that are hardly perceptible by human beings. In this field, AI’s role, at least in the short term, has often been overestimated. The major difficulty now encountered is (even before we imagine deploying such tools) to have a sovereign capacity for curating and consolidating the data collected and providing analysts with the electronic means for more efficient data-processing.

Contrary to what is taken for granted, the success of software (*e.g.*, the programs offered by Palantir) is based not on AI (which is, in fact, very slightly integrated in the software) but on this market leader’s ability to rapidly aggregate data in various formats and from various origins, and provide analysts with the means for visualizing the data, for crossing and analyzing them as they have been trained to do. The strength of such firms does not stem from any technological advance in AI but, instead, from the developments in computer science that they have undertaken in close association with customers in various domains (intelligence services, banks, industry, etc.). Thanks to this approach, these firms have available a constantly growing array of rapidly usable applications that they can adapt, as need be, to new clients — a sort of virtuous circle.

The lag of Europe and France does not, therefore, result from any slack in AI. French firms have difficulty proposing competitive products because they lack agility and marketing savvy. Evidence of this is their inability to make customers' needs their own and to work closely with clients so as to make offers adapted to their line of business. Although AI-based applications will, of course, be added onto data-processing software, what counts in the short run is not so much a firm's technological achievements in AI as a better approach to customer relations. This is needed for a plausible national offer to emerge.

AI specific to national security: The question of data availability

In many sectors, the small volume of available data, which are mostly in the hands of big, foreign high tech firms, has set back the development of a sovereign AI. In contrast, defense and national security have an abundance of data: images from satellites, electronic communications, metadata captured from telecommunication networks, videos and sound tracks, computer data collected for cybersecurity purposes, etc. More than in other sectors however, these data, their access and use are subject to major restrictions. The principles of confidentiality and the "need to know" severely restrict access to sensitive data collections. French law tightly regulates access to the data collected by the country's intelligence services.

Given this context, how to obtain the sets of training data that can be used for AI, even in national security? The answer implies reviewing the governance of data governance in major government ministries and drafting a clear policy on data management and use (with provisions about making the data available for training algorithms and for testing algorithms on real data).

This problem swells insofar as the development of AI implies a paradigm shift in defense. Defense used to be a field where inventions were made and then transferred toward the civilian economy. Nowadays, defense has to rely on unfamiliar high tech firms and imagine new methods for working together. Defense must make its operational data lakes available to innovative firms while strictly overseeing access. It must also use open-source software programs and adapt them, or else invest in designing its own AI systems even though the latter might, at least for the time being, be less effective than what foreign high tech firms are selling.

Trust in AI, a crucial national security issue

AI algorithms sometimes yield results that human beings deem aberrant; and AI systems sometimes unexpectedly fail. According to a former director of the US Defense Advanced Research Projects Agency (DARPA), "*The problem is that when they're wrong, they are wrong in ways that no human would ever be wrong.*" Furthermore, AI learning techniques risk being biased, involuntarily (due to unrepresentative training data) or voluntarily (because a third party has tried to modify the training data to influence the system). These techniques might also yield obscure results that are hard to explain. Major international studies have been devoted to the oft asked question about the explicability of AI. These limitations raise questions about how much we can trust AI, questions that are even keener when AI is put to use for defense and national security. In fact, these questions are so important that AI cannot be used for many military purposes. We cannot yet predict whether these limitations will be overcome or not.

Given the impossibility of guaranteeing trust in AI, a requisite “level of confidence” has to be set for each AI application in defense and national security. This approach forces us to have a sovereign capacity for evaluating AI systems so as to gauge the level of confidence or certify a system — a procedure on par with the certificates of security delivered by the National Cybersecurity Agency (ANSSI: Agence Nationale de la Sécurité des Systèmes d’Information). However the certification of AI (of its learning techniques in particular) is still a topic for research. Certification procedures have not yet been defined. This requirement represents a major challenge and limitation with regard to using AI in defense and national security. Without substantial advances in this respect, AI will remain restricted to uses that help human operatives gain time, but without disruption and without upsetting balances of power.