COMMENT
16 April 2018

Regulate artificial intelligence to avert cyber arms race

Define an international doctrine for cyberspace skirmishes before they escalate into conventional warfare, urge Mariarosaria Taddeo and Luciano Floridi.

Mariarosaria Taddeo &

Mariarosaria Taddeo is a research fellow and deputy director of the Digital Ethics Lab at the Oxford Internet Institute, University of Oxford, UK; and a Turing fellow of the Alan Turing Institute, London, UK.
Contact

Search for this author in:

Pub Med
Nature.com
Google Scholar

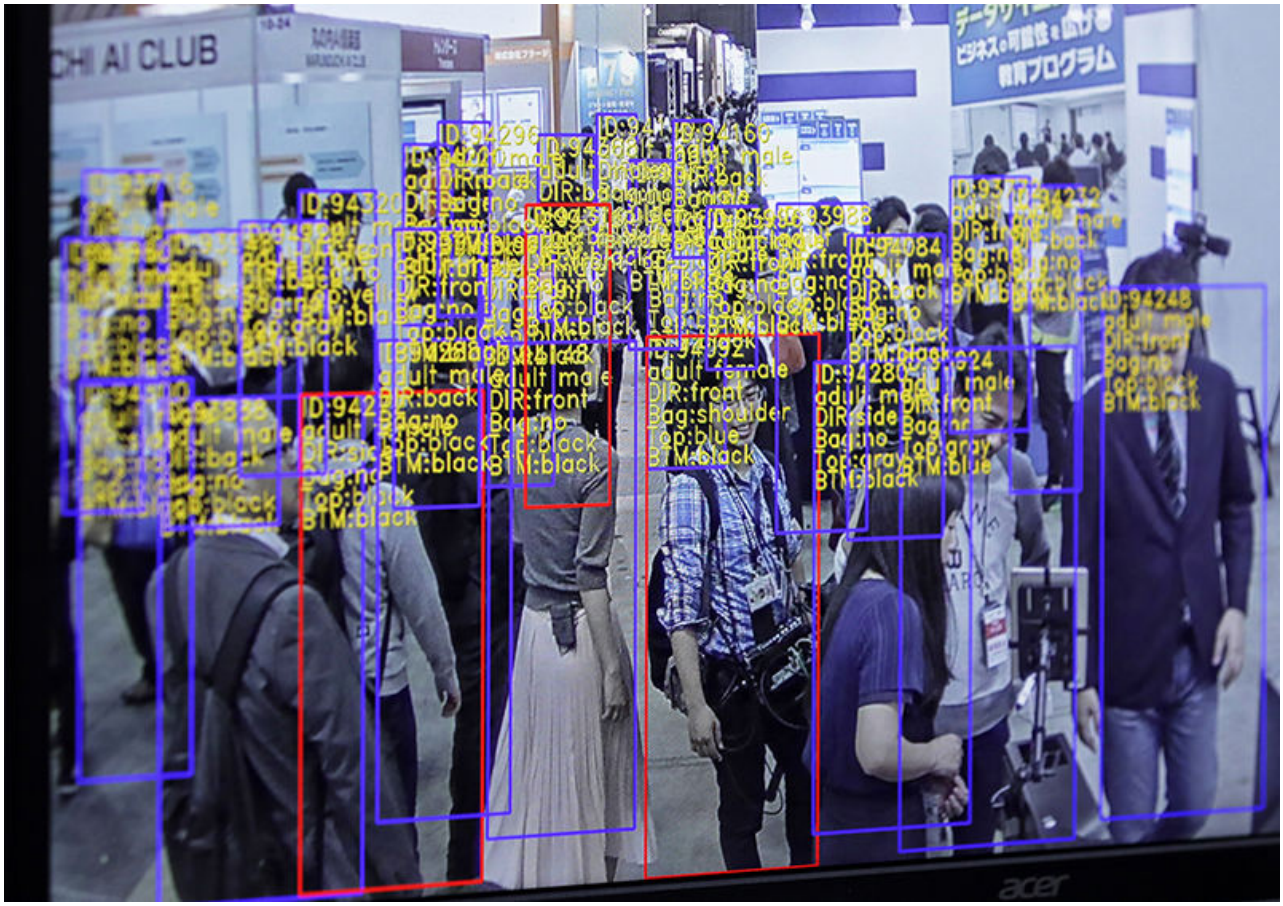
Luciano Floridi

Luciano Floridi is professor of philosophy and ethics of information at the University of Oxford, UK; director of the Digital Ethics Lab at the Oxford Internet Institute; and chair of the Data Ethics Group at the Alan Turing Institute.

Search for this author in:

Pub Med
Nature.com
Google Scholar

PDF version



Object detection and tracking technology on show at the 2018 Artificial Intelligence Exhibition & Conference in Tokyo.Credit: Kiyoshi Ota/Bloomberg/Getty

Cyberattacks are becoming more frequent, sophisticated and destructive. Each day in 2017, the United States suffered, on average, more than 4,000 ransomware attacks, which encrypt computer files until the owner pays to release them[1]. In 2015, the daily average was just 1,000. In May last year, when the WannaCry virus crippled hundreds of IT systems across the UK National Health Service, more than 19,000 appointments were cancelled. A month later, the NotPetya ransomware cost pharmaceutical giant Merck, shipping firm Maersk and logistics company FedEx around US$300 million each. Global damages from cyberattacks totalled $5 billion in 2017 and may reach $6 trillion a year by 2021 (see go.nature.com/2gncsyg).

Countries are partly behind this rise. They use cyberattacks both offensively and defensively. For example, North Korea has been linked to WannaCry, and Russia to NotPetya.

As the threats escalate, so do defence tactics. Since 2012, the United States has used ‘active’ cyberdefence strategies, in which computer experts neutralize or distract viruses with decoy targets, or break into a hacker’s computer to delete data or destroy the system. In 2016, the United Kingdom announced a 5-year, £1.9-billion (US$2.7-billion) plan to combat cyber threats. NATO also began drafting principles for active cyberdefence, to be agreed by 2019. The United States and the United Kingdom are leading this initiative. Denmark, Germany, the Netherlands, Norway and Spain are also involved (see go.nature.com/2hebxnt).

Artificial intelligence (AI) is poised to revolutionize this activity. Attacks and responses will become faster, more precise and more disruptive. Threats will be dealt with in hours, not days or weeks. AI is already being used to verify code and identify bugs and vulnerabilities. For example, in April 2017, the software firm DarkTrace in Cambridge, UK, launched Antigena, which uses machine learning to spot abnormal behaviour on an IT network, shut down communications to that part of the system and issue an alert. The value of AI in cybersecurity was $1 billion in 2016 and is predicted to reach $18 billion by 2023[2].

By the end of this decade, many countries plan to deploy AI for national cyberdefence; for example, the United States has been evaluating the use of autonomous defence systems and is expected to issue a report on its strategy next month[3]. AI makes deterrence possible because attacks can be punished[4]. Algorithms can identify the source and neutralize it without having to identify the actor behind it. Currently, countries hesitate to push back because they are unsure who is responsible, given that campaigns may be waged through third-party computers and often use common software.

The risk is a cyber arms race[5]. As states use increasingly aggressive AI-driven strategies, opponents will respond ever more fiercely. Such a vicious cycle might lead ultimately to a physical attack.

Cyberspace is a domain of warfare, and AI is a new defence capability. Regulations are thus necessary for state use of AI, as they are for other military domains — air, sea, land and space[6]. Criteria are needed to determine proportional responses, as well as to set clear thresholds or 'red lines' for distinguishing legal and illegal cyberattacks, and to apply appropriate sanctions for illegal acts[7]. In each case, unilateral approaches will be ineffective. Rather, an international doctrine must be defined for state action in cyberspace. Alarmingly, international efforts to regulate cyber conflicts have stalled.

We call on regional forums, such as NATO and the European Union, to revive efforts and prepare the ground for an initiative led by the United Nations. In the meantime, computer experts must be transparent about problems, limitations and shortcomings of using AI for defence. Researchers must also work with policymakers and end users to design testing and oversight mechanisms for this technology.

No rules

Right now, the UN process is in deadlock. In 2004, the UN set up the Group of Governmental Experts on Information Security to agree on voluntary rules for how states should behave in cyberspace. Its fifth meeting, in 2017, ended in a stand-off. The group could not reach consensus on whether international humanitarian law and existing laws on self-defence and state responsibility should apply in cyberspace. The United States argued that cyberdefence regulations should build on these laws. Other nations, including Cuba, Russia and China, disagreed. They argued that this would 'militarize' cyberspace and send the wrong message about peaceful conflict resolution. The group failed to deliver its report. It is unclear whether it will meet again, or what will happen next.

International dialogue and action must resume. NATO could pave the way through its forthcoming guidelines, although it is currently unclear what their scope will be.

A military cyberdefence specialist at a conference in Lille, France. Government spending on cyber strategies has soared over the past decade.Credit: Phillippe Huguen/AFP/Getty

Meanwhile, research on AI for cyberdefence is progressing quickly. The United States is in the lead, technologically. It aims to incorporate AI into its cyberdefence systems by 2019[3]. The US Department of Defense (DOD) has earmarked $150 million for research. The US Defense Advanced Research Projects Agency (DARPA) is developing the techniques and

strategies. Steps have already been taken. In DARPA's 2016 Cyber Grand Challenge competition, seven AI systems, developed by teams from the United States and Switzerland, fought against each other. The systems identified and targeted their opponents' weaknesses while finding and patching their own.

The DOD will issue the first US report on AI strategies for national defence in May. There is, as far as we know, no indication of what its approach will be. Previous documents, such as *The DOD Cyber Strategy* from 2015 or the 2016 *National Cyber Incident Response Plan*, did not cover autonomous systems, machine learning or AI. The 2012 DOD directive on 'Autonomy in Weapon Systems' focused on internal procedures for deploying AI but was silent on when the United States would do so in the international arena.

AI is a priority for China, which aims to become a world leader in machine-learning technologies. In July 2017, the Chinese government issued its Next Generation AI Development Plan. Military implementation of AI, on the battlefield as well as in cyberspace, is a crucial part of the strategy. But it is unclear to what degree China plans to deploy AI actively in cyberdefence.

Russia has not released any public documents about its strategies for AI in defence. However, in a video message released in 2017, President Vladimir Putin referred to AI and stated: "Whoever becomes the leader in this sphere will become the ruler of the world." Experts agree that Russia is focusing on developing AI-enhanced tools for its conventional forces. However, since 2014, the Russian National Defense Control Center has been using machine-learning algorithms to detect online threats. Allegedly, Russia has pioneered the use of AI to spread disinformation and intervene in the public debates of other nations, including the 2016 US presidential election and the United Kingdom's EU membership referendum. Although these operations are not part of national defence strategies, they indicate Russia's advanced AI capabilities.

North Korea has a history of cyberspace aggression. It was implicated, for example, in the WannaCry attack in 2016 and in another major breach, against Sony Pictures, in 2014. The country lacks technical expertise in AI but is likely to want to catch up with its adversaries.

The EU is stepping up, too. In 2017, it reassessed cybersecurity and defence policies and launched the European Centre of Excellence for Countering Hybrid Threats, based in Helsinki. The EU has the most comprehensive regulatory framework for state conduct in cyberspace so far. Yet these directives do not go far enough. The EU treats cyberdefence as a case of cybersecurity, to be improved passively by making member states' information systems more resilient. It disregards active uses of cyberdefence and does not include AI.

This is a missed opportunity. The EU could have begun defining red lines and proportionate responses in its latest rethink. For example, the 2016 EU directive on 'Security of Network and Information Systems' provides criteria for identifying crucial national infrastructures, such as health systems or key energy and water supplies that should be protected. The same criteria could be used to define illegitimate targets of state-sponsored cyberattacks.

Regional forums, such as NATO and the EU, must take the following three steps to avoid serious imminent attacks on state infrastructures, and to maintain international stability.

Three steps

Define legal boundaries. The international community needs to agree urgently on red lines that distinguish between legitimate and illegitimate targets. Also needed are definitions of proportionate responses for cyberdefence strategies. International consensus at the UN level will ultimately be required. Until then, guidelines from regional multilateral bodies, such as NATO and the EU, must cover these issues and lead by example.

Test strategies with allies. 'Sparring' exercises should be organized between friendly countries to test AI-based defence tactics. These tests should be mandatory before any system is deployed. They could be in the form of DARPA's Grand Challenge or the simulation exercises routinely run by NATO and the EU. Because AI learns by experience, these matches will improve the strategies of the alliance, while finding and healing weaknesses. Fatal vulnerabilities of key systems and crucial infrastructures should be shared with allies; policy frameworks should demand disclosure. Agreements and regulations with similar sharing and disclosure requirements include the EU Electronic Identification, Authentication and Trust Services Regulation and NATO's Industry Partnership Agreement.

Monitor and enforce rules. The international community needs to agree how to audit and oversee AI-based state cyberdefence operations. 'Alert and remedy' mechanisms are needed to address mistakes and unintended consequences. A

third-party authority with teeth, such as the UN Security Council, should rule on whether red lines, proportionality, responsible deployment or disclosure norms have been breached. Economic or political sanctions should be imposed on states that violate rules. NATO and the EU should enforce the norms within their remits.

The solution is difficult, but it is clear. There is no time to waste.

Nature **556**, 296-298 (2018)

doi: 10.1038/d41586-018-04602-6

References

US FBI. *How to Protect Your Networks from Ransomware* (FBI, 2017).

P&S Market Research. *Artificial Intelligence (AI) in Cyber Security Market* (P&S Market Research, 2017).

US Defense Science Board. *Summer Study on Autonomy* (US Department of Defense, 2016).

Taddeo, M. *Phil. Technol.* https://doi.org/10.1007/s13347-017-0290-2 (2017).

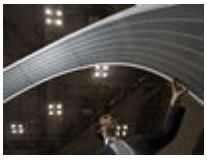
Yang, G.-Z. *et al. Sci. Robot.* **3**, eaar7650 (2018).

Floridi, L. *Phil. Trans. R. Soc. A* **374**, 20160112 (2016).

Taddeo, M. *Minds Mach.* **27**, 387–392 (2017).

Related Articles

Cybersecurity needs women



Use machine learning to find energy materials

The environment needs cryptogovernance

Subjects

Nature

ISSN 1476-4687 (online)