The Security and Military Implications of Neurotechnology and Artificial Intelligence

15

Jean-Marc Rickli and Marcello Ienca

Contents

Abstract

This chapter aims at taking stock of technological advances in artificial intelligence (AI) and neurotechnology and looks at the security and military implications of these technologies in light of their current capabilities. AI and neurotechnology hold a great transformative potential due to their ability to read,

J.-M. Rickli
Geneva Centre for Security Policy (GCSP), Geneva, Switzerland
e-mail: j.rickli@gcsp.ch

M. Ienca (✉)
Department of Health Sciences and Technology (D-HEST), Swiss Federal Institute of Technology, ETH Zurich, Zurich, Switzerland
e-mail: marcello.ienca@hest.ethz.ch

O. Friedrich et al. (eds.), *Clinical Neurotechnology meets Artificial Intelligence*, Advances in Neuroethics, https://doi.org/10.1007/978-3-030-64590-8_15

modify, simulate and amplify human cognition in a variety of domains and in response to a variety of cognitive and analytical tasks. Furthermore, both technologies are rapidly proliferating outside traditional supervised settings (e.g. the clinics and academic research) onto multiple and unsupervised domains, a phenomenon that can be labelled "horizontal proliferation". Among these domains, their co-optation into the military sector and subsequent weaponization are of particular concern from an international security perspective. For each technological category, five security-relevant issues are discussed: data bias and accountability, manipulation, social control, weaponization and democratization of access. We argue that, in light of their disruptive potential and rapid proliferation, both neurotechnology and artificial intelligence urge global governance responses that deal with their accessibility, their proliferation, their dual-use nature including how easily these technologies can be repurposed and obviously the ethics and values that should accompany the development and use of these technologies. These responses should be inclusive and comprise all the different stakeholders (governments, private sector, scientific community, civil society and tech companies) and be very versatile as these technologies and applications evolve rapidly.

15.1 Introduction

Cognition is a major driver of complex information processing, knowledge acquisition and adaptive behaviour in both biological organisms and artificial systems. In the last century, parallel advances in the mapping and functional understanding of the human brain, on the one hand, and the processing of information in artificial systems (e.g. computers), on the other hand, have led to the development of a variety of technologies that assist, augment or simulate human cognitive processes or that can be used to achieve cognitive aims. These technologies, sometimes referred to using the umbrella term "cognitive technology" [1, 2], can be classified into two main categories: (a) technologies that monitor, assist or enhance cognitive processes in biological organisms—human beings included—and (b) technologies that simulate (aspects of) natural cognitive processes through artificial systems. The first category, commonly referred to as neurotechnology, encompasses devices that interface biological nervous systems to monitor, assist or enhance the cognitive processes executed by those systems. The second category, commonly referred to as artificial intelligence (AI) (or cognitive computing), encompasses systems and devices that artificially simulate cognitive functions typically executed by biological nervous systems—especially human brains—such as learning, planning, reasoning and perceiving the environment. In the last decade, these two domains have increasingly converged due to a twofold trend. First, artificially intelligent features have increasingly been embedded in neurotechnologies in order to better extract, classify and decode neural signals. Second, following trends such as neuromorphic artificial

intelligence, artificial cognitive systems have been inspired by the study of biological neural systems.

Due to their disruptive potential, both families of cognitive technology raise security concerns, especially due to potential military implications. These implications are associated with three shared characteristics of both types of cognitive technology, namely: (a) proliferation outside supervised research domains, (b) re-purposing for military aims and (c) highly transformative, even disruptive, potential. Physicist Stephen Hawking famously said a few months before he passed away that,

> success in creating effective AI could be the biggest event in the history of our civilization, or the worst. We just don't know. So we cannot know if we will be infinitely helped by AI, or ignored by it and sidelined or conceivably destroyed by it (quoted in [3]).

In recent years, similar predictions have been made by other prominent figures such as philosopher Nick Bostrom and entrepreneur Elon Musk, who both raised the prospects that technologies related to AI might turn bad. Similarly, lieutenant colonel of the United States Air Force Brian E. Moore has predicted that neurotechnology, especially brain-computer interfacing, "has the potential to revolutionize military dominance much the same way nuclear weapons have done" [4]. A fierce debate pitting proponents and adversaries of cognitive technology ensued. Though this debate is very often characterized by exaggerations, hyperboles or even fear-mongering statements about the either utopian or dystopian consequences of AI and neurotechnology, it has the merit to raise public awareness about the security implications of these emerging technologies.

This chapter aims at taking stock of technological advances in artificial intelligence and neurotechnology and looks at the security and military implications of these technologies in light of their current capabilities. For each technological category, five security-relevant issues are discussed: data bias and accountability, manipulation, social control, weaponization and democratization of access.

15.2 The Security Implications of Artificial Intelligence

There is no universally agreed upon definition of artificial intelligence. As noted by the group of researchers at the University of Helsinki, AI is a scientific discipline, meaning that AI is a "collection of concepts, problems and methods for solving them" [5]. Nonetheless, the definition provided by the independent high-level expert group on artificial intelligence of the European Commission is a good starting point. Thus, artificial intelligence systems are

> "software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn

> a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions" [6].

Artificial intelligence methods enable "machines to learn from experience, adjust to new inputs and perform human-like tasks" [7]. Nowadays, it is hard to come across artificial intelligence without encountering the words "machine learning" (ML) which is a subset of AI. Essentially, ML refers to the development of algorithms which progressively improve performance on a specific task by making and testing predictions on data without being explicitly programmed. ML provides computers with the ability to use data to teach themselves, instead of via humans who program the machine.

There are two different "types" or categories of AI, known as "narrow" or "weak" AI and "general" or "strong" AI [8]. The distinction here is between machines that can perform and outperform humans in one specific task, and machines that might be able to adapt to any tasks. Today we are good, and getting better, at "narrow" AI, but are still decades away from creating machines which can perform the wide array of human-like tasks of "general" AI. In a recent survey of AI experts, the median timeframe predicted for the achievement of artificial general intelligence (AGI) is 45 years from now [9]. If, and once AGI is reached some posit that then AI shall be very rapidly developed into superintelligence surpassing any human intelligence [10]. In this chapter, we are considering current state of AI, that is narrow AI. Narrow AI pulls information from one specific dataset; it is programmed to perform a single task and does not perform outside of that single task which it was designed to perform. Algorithms relying on narrow AI include those used by Google Translate, spam-filtering systems, facial recognition technology and algorithms designed to learn and play video games, for instance.

The past few years have seen major progress in the development of AI. This is not only due to the vast improvement of the algorithms and techniques used, such as deep learning or machine learning, but also due to the incredible computer capacity that is now available and the vast amount of data that can be used to train the algorithms better than ever before. Thus, Google Deepmind, through its Alpha-class algorithms, achieved superhuman capabilities at the games of chess, shogi and Go, competed at the same level as the best players of the video game Starcraft II and won a global competition based on folding proteins [11–13]. Another algorithm, Libratus, developed by Carnegie Mellon University, defeated the best Texas Hold 'Em Poker players in January 2017 [14]. In February 2019, the most advanced Natural Language processing (NLP) algorithm was developed by Open AI, a leading AI research organization based in San Francisco. NLP is a subcategory of artificial intelligence which focuses on training computers to understand and process human language. This is a particularly difficult strand of AI, as computers do not have the same intuitive understanding of human languages; computers cannot "read between the lines" and understand implied meaning. The Open AI algorithm was trained to predict the next word, given all the previous words within a text [15]. The result has been the ability to generate lengthy text samples of unprecedented quality based on an input [16].

The transformative nature of AI offers fantastic prospects for improving human life in every domain. Thus, algorithms have surpassed humans at image recognition, which has had positive implications in the medical imagery domain, for instance. Algorithms are now much better at reading MRI, scans or X-rays than doctors are, therefore also reducing the risk of mistakes [17]. However, this technology also entails potential risks related to their misuse or malevolent use. The following sections will deal with the issues of data bias and accountability, manipulation especially for political purposes, social control, military applications and democratization of access.

15.3 Data Bias and Accountability

As AI is highly dependent on the data that it is fed with, biased data will lead to biased results. An experiment at Massachusetts Institute of Technology (MIT) which fed an algorithm with data only depicting crime scenes and death, led the algorithm to interpret any picture as death-related. The researchers called this algorithm "Norman", the world's first psychopath AI [18]. This experiment convincingly demonstrates the crucial importance of the quality of data required to train algorithms and the consequent inherent problem of biases in artificial intelligence.

When applied to real-world problems, the use of such technology, not being entirely aware of real-world subtleties, can entail moral and ethical problems. This is true, for example, for the criminal justice system. In 2016, ProPublica released an investigation into a machine learning system used by some courts in the USA [19]. The system was used to predict which individuals would be more susceptible to commit another crime after their release. It was observed that a system originally intended to operate free of human bias, only perpetuated this bias on a wider scale [20]. Indeed, as it was fed with historical criminal data from a criminal justice system that is historically biased against African American individuals, the system rated black individuals more negatively than white individuals to the point that the predictive algorithm was twice as likely to incorrectly classify black defendants as being at a higher risk than whites. In this sense, the results represented an automation of bias. We can see here a clear ethical conundrum. This led the major American tech companies to regroup in the consortium "Partnership on AI" to speak out against the use of algorithms for jailing people [21].

Furthermore, the results yielded by an AI powered algorithm are by definition not transparent and explainable. This is called the black-box problem of AI [22]. Because of its complex mathematical and probabilistic operations, the accountability of the machine learning process is very difficult to guarantee. Indeed, once fed with certain inputs, it is very complex to understand how the algorithm goes about producing the outputs. This impedes the understanding of "why" an algorithm has come to a certain conclusion. If AI is to have an increasingly influential role in the world and control greater parts of our lives, it is essential that they are accountable because people and society will want to know "why" algorithms make certain decisions that determine access to loans, recommend a medical treatment, or identify

national security threats. Being unable to answer these questions might reduce the overall trust in these systems and therefore hinder their adoption [23].

15.4 Manipulations

The 2016 US presidential election can be seen as a turning point in the history of political manipulation. A private company named "Cambridge Analytic" was involved in a disinformation campaign to sway political vote in favour of Republican candidate Donald Trump. Cambridge Analytica did this by targeting voters based on their personal data generated on social media and other digital platforms [24].

Now picture the same process with an incredibly accurate AI, capable of automating the creation of fake and targeted content and flooding the web so that everybody could potentially receive personalized advertising and information that only reinforce held beliefs. This would raise enormous ethical and political concerns as it would undermine democratic processes by enabling malicious actors to stir political debate and dilute the truth.

The development of generative adversarial networks (GANs)—which are algorithms pitting neural networks against each other—has made it possible to manipulate data to a level unseen before, notably through deepfake which is a technique that superimposes images and videos onto other source images or videos. Deepfake pornography surfaced on the internet in 2017 and in January 2018, a desktop application called FakeApp was launched. Similarly, voice mimicking software such as Lyrebird or Baidu's Deep Voice can "clone" anyone's voice. The Chinese tech giant application only needs 3.7 s of audio of a voice to reproduce it [25]. The combination of voice and image forgery will make any piece of media on the internet suspicious. Such applications have democratized the ability to create perfect visual and audio manipulations [26]. This is often referred to as the "end of truth" or the end of "seeing is believing", which Henry Kissinger has identified as leading to the "end of the Enlightenment era" [27]. Building such algorithms without security in mind, and without thinking about the possible repercussions on society carries enormous risk.

15.5 Social Control and Discrimination

As mentioned earlier, algorithms have surpassed humans at image recognition, which means that AI is much better at identifying visual patterns, including for facial recognition. Some governments have seen the benefits of such technologies and use it to increase the surveillance of their citizens. China has gone the furthest in this field. AI-enabled technologies have allowed Beijing to create an advanced surveillance system by awarding Chinese citizens a social score based on their online and offline behaviour. As Rickli stated previously,

> "the Chinese government has implemented a surveillance system based on the gamification of obedience through big data and artificial intelligence. It relies on punitive and reward measures that influence the way its citizen should behave (quoted in [28])".

Beyond this, the Chinese government is also using facial recognition algorithms to identify one specific ethnic group, the Uighurs, for law enforcement purposes. The Uighurs are a minority of 11 million, mostly located in the western region of Xinjiang. China is mainly populated by the Han ethnic group. The Chinese police has used "facial recognition technologies to target Uighurs in wealthy eastern cities like Hangzhou and Wenzhou and across the coastal province of Fujian" and it is spreading to more than 16 different provinces and regions across China [29]. In one city, law enforcement authorities ran such a system more than 500,000 times within the course of a month in 2019 to screen whether residents were Uighurs. The purpose of this technology is to monitor and track this ethnic group, which the Chinese government accuses of ethnic violence and terrorist attacks.

Ethnic profiling is a dangerous development in facial recognition technologies and AI more generally that is very appealing to authoritarian regimes. Chinese AI surveillance technologies are now also being exported to other states such as Zimbabwe, Singapore, Malaysia or Mongolia [29].

15.6 Military Applications of AI

The military domain is not immune to developments in AI. With artificial intelligence, the new tactic of swarming will become possible in the physical domain. Swarming relies on overwhelming and saturating the adversary's defence system by synchronizing a series of simultaneous and concentrated attacks [30]. In October 2016, the US Department of Defense conducted an experiment that saw 103 Perdrix micro drones autonomously deal with four different objectives. Meanwhile, the world record for swarming drones was broken by a Chinese company, EHang, in May 2018 with an AI-assisted swarm of 1374 drones flying over the City wall of Xi'An and then by the US company Intel in July 2018 with 2018 drones [31, 32]. Swarming tactics are potentially disruptive because they combine firepower, mass and speed.

These factors combined with the specific capabilities of artificial narrow intelligence systems means that defence is rapidly becoming costlier and less effective than offence, shifting the dynamics of security towards pre-emption [33].

The development of lethal autonomous weapons systems (LAWS) in particular will likely have a destabilizing impact on strategic stability in the future [34]. Since WWII, strategic stability has been guaranteed by the supremacy of the defensive, especially due to the sheer destructive power of the second-strike retaliatory capabilities of nuclear weapons. If the applications of swarming tactics make second-strike retaliatory capabilities an illusion because of the offensive advantage provided by swarming, it will follow that deterrence will be replaced by pre-emption. These

changes in strategy are very likely to create an unstable international configuration that encourages escalation and arms races [35].

So far, international law prohibits the use of military force except in cases of self-defence and if the UN Security Council allows it under Chap. 7 of the UN Charter. If the offensive has the advantage, the only way to protect yourself is by attacking first. Pre-emption is therefore in direct contradiction to the spirit of the UN Charter and its application is a violation of Art 2(4) of the Charter. As Rickli argues, this new international system that stems from the militarisation of AI will be much more unstable and prone to conflicts and will make pre-emption the strategy of choice to deal with adversaries [36].

Moreover, the growing use of autonomy in weapon systems allows the potential development of weapons that will be fully autonomous. These weapons will be able to move independently through their environment to arbitrary locations, select and fire upon targets in their environment and create and/or modify its goals, incorporating observation of its environment and communication with other agents [37]. Such weapons will accelerate a trend in the development of warfare in the twenty-first century, which entails that state and non-state actors increasingly rely on both human and technological surrogates to fight on their behalf [38]. Such developments favour international instability because it reduces the threshold to use force as well as a drastic reduction in the accountability of the use of force.

15.7 Security Implications of Democratization of Access

A key characteristic of emerging technologies is the rapid decrease in the cost of access [39]. In the case of AI, the drop in the cost of the technology is due to the growth of the processing power of CPUs and the creation of larger data sets. Furthermore, the digital nature of AI systems—and the fact that AI algorithms are often public or even open-source—allows them to be distributed and scaled rapidly [40].

As a result of these cost shifts, lower barriers to entry incentivize new actors to use this technology. From a security perspective, the automation of tasks mean that individuals will potentially become more dangerous as they may have access to technologies with disruptive impacts. As greater numbers of actors invest in AI-driven tactics, higher rates of experimentation and innovation will result in the emergence and proliferation of new threats and tactics [41].

The falling costs and the accessibility to AI particularly empowers individuals, small groups, criminal enterprises and other non-state actors [42]. This is very visible in the cyber domain, where the acquisition of new cyber capabilities is cheap and the marginal cost of additional production—adding a target—is close to zero [43]. Equally, in the physical domain, AI-enabled commercial products can easily be repurposed for surveillance purposes or to attack targets [40]. Although not AI, ISIS mounted high-definition cameras under drones to improve intelligence and acquire situational awareness during their combat operations. They also used drones

to drop 40 mm grenades on Iraqi positions, allegedly killing up to 30 Iraqi soldiers per week during the battle of Mosul in 2017 [44]. This demonstrates how agile terrorist organizations are in using commercial technologies to support their goals. AI will probably not be an exception to the rule in that once algorithms have been developed they are either easily accessible once they are released into databases (e.g. Tensorflow) or can be deducted from adversarial black-box attacks. The next section looks at the security implications of neurotechnology.

15.8 The Security Implications of Neurotechnology

Neurotechnology can be defined as "devices and procedures that are used to access, monitor, investigate, assess, manipulate and emulate the structure and function of neural systems" [45, 46]. While AI systems emulate or simulate functional aspects of the (human) brain, neurotechnologies are designed to record, monitor, functionally understand and modulate processes in the (human) brain. Neurotechnologies stricto-sensu include non-invasive medical imaging technologies such as magnetic resonance imaging (MRI) and near-infrared spectroscopy (NIRS), electrode-based electrophysiological monitoring (EEG), non-invasive neuromodulation techniques such as transcranial magnetic stimulation (TMS) or transcranial electric stimulation (tES), sensory neuroprosthetics such as visual or auditory prostheses as well as invasive neurostimulation techniques involving implant neurosurgery such as deep brain stimulation (DBS). Broader definitions of neurotechnology also encompass computational simulations of neural functions and neuromorphic engineering.

Neurotechnology originated in the clinical domain as an array of tools and techniques aimed at monitoring, modulating, restoring or enhancing neural structures or functions. Furthermore, neurotechnology plays a critical role in research and is a major enabler of discovery and translational neuroscience. Advances in neurotechnology are necessary requirements for achieving the grand challenges of contemporary neuroscience, namely: (a) reliably measuring neuronal activity, (b) mapping neuronal activity onto a reliable and highly detailed anatomical and functional atlas of the brain and (c) making sense of the brain by mining large volumes of brain data through reliable and high-velocity analytic techniques [47]. Meeting these three scientific challenges, in turn, is essential to the development of preventative, diagnostic, therapeutic or assistive solutions that might reduce the burden of neurological disorders and improve the lives of millions of patients.

In recent years, advances in neuroengineering and pervasive computing, combined with increased extra-clinical interest in the potential of neurotechnology, have propelled neurotechnologies from the exclusive clinical and biomedical domains onto a broad variety of commercial [48], educational [49] and military applications. Consumer-grade neurotechnologies include several non-invasive neurodevices such as neuromodulatory devices based on transcranial direct current stimulation (tDCS) or transcranial magnetic stimulation (TMS), brain–computer interfaces (BCIs) for self-neuromonitoring and device control, and an associated ecosystem of both proprietary and open-source software (including mobile applications).

The proliferation of neurotechnology outside clinics and research domains raises security implications. The reason is threefold. First, the domain of consumer neurotechnologies is, to date, largely unregulated. While clinical neurotechnologies are subject to medical device regulation and stricter privacy rules for the processing of health data, consumer neurotechnologies are currently being developed in an undefined legal territory, and existing regulatory oversight has been deemed "insufficient" by experts [48, 50]. The absence of adequate oversight mechanisms and unambiguous regulation increases the chances that security breaches might emerge [51], some of which reportedly already have [52]. Furthermore, unlike clinical and research applications, consumer neurotechnologies are not typically used in a medically supervised environment and are not subject to continuous safety monitoring by researchers. This increases the chances that the technology might be misused either by the users themselves or by third parties. Finally, the proliferation of unsupervised neurotechnology applications causes a proliferation of actors involved in the handling of neurodevices and derived brain data. Today, the categories of actors involved in the development and use of neurotechnology do not exclusively comprise neuroscientists, neuroengineers and neurological patients. Consumer neurotechnology applications have opened the gates of neurotechnology use to the general population, including healthy individuals. Furthermore, following sociotechnical trends such as do-it-yourself (DIY) neurotechnology and biohacking, neurotechnologies are increasingly being developed and experimented with by non-professional scientists. These trends are causing both a proliferation of actors and a fragmentation of oversight measures, with a consequent increase in security risks.

To comprehensively map the dual-use landscape of neurotechnology, it should be noted that the extra-clinical proliferation of neurotechnology is not limited to the civilian domain, but also extends to the military sector. In the last decade, several neurotechnologies have gained ground as experimental applications among governmental national security agencies such as the Defense Advanced Research Projects Agency (DARPA), a research agency of the United States Department of Defense. Military uses of neurotechnology include experimental applications for brain-based intercept-proof communication, remote device control (e.g. brain-controlled unmanned aerial vehicles), warfighter enhancement and post-traumatic treatment of veterans. This process of permeation of neurotechnology in the state military sector has been termed "weaponization of neuroscience" [53], even though authors have argued that neurotechnology has been "a toll of war from the start" [54]. For example, Howell has observed that the origin of clinical neurology is intertwined with the American civil war and that the birth of modern neuroscience was highly dependent on research conducted with military research institutions such as the Walter Reed Army Institute of Research (ivi). Finally, misuse of neurotechnology by malign non-state actors has also been indicated as a primary source of risk for international security [52, 55].

15.9 Data Bias, Agency and Accountability

As the functioning of neurotechnology, especially BCI, is highly dependent on data, data quality and data protection measures are paramount to ensure safety and security. Biases in datasets, poor data quality and corrupted data can all negatively affect the functioning of neurotechnologies and lead to suboptimal or even harmful outcomes. Furthermore, experts have argued that algorithmic biases, such as those affecting datasets used to feed AI applications, could become embedded in neural devices [56]. The reason for this stems from the fact that most neurotechnologies rely on machine learning and other AI techniques to decode brain signals and translate them into utilizable output. Consequently, biases contained in the datasets used to train those algorithms are likely to be transferred or even amplified during the process. This risk is exacerbated in the context of neurotechnologies used by vulnerable user groups such as children, patients with neurological disorders or socially marginalized individuals.

The increasing use of machine learning and, more generally, of artificial intelligence to optimize BCI functions also has implications for the notion of action and responsibility. For example, Klaming and Haselager [57] have hypothesized that when BCI control is partly dependent on intelligent algorithmic components, it may become difficult to discern whether the resulting behavioural output was actually executed by the user. This difficulty introduces a principle of indetermination into the cognitive process that starts from the conception of an action (or intention) to its execution, with consequent uncertainty in attributing responsibility to the author of such action. This principle of indetermination could call into question the notion of individual responsibility, with obvious consequences of a criminal and insurance nature. More broadly, it could also jeopardize the entire concept of legal liability because liability is predicated upon the state of a legal person of being legally responsible. If the intelligent components embedded in the BCI override the human user's volition or simply make any discrete attribution of responsibility indeterminable, this would represent a fundamental transformation of both the civil and criminal law systems as they both rely on the establishment of liability to make actors responsible or answerable in law. Moreover, the principle of indetermination could generate a sense of estrangement in the user, whose ethical relevance is all the greater if he/she is a vulnerable individual such as a neurological patient. In addition, there is a possibility that the centrality of these intelligent components in the functioning of the BCI may affect the user's subjective experience, and thus their personal identity [58]. This hypothesis has recently obtained a preliminary empirical confirmation in a qualitative study about the personal experience of DBS patients [59].

15.10 Manipulations

Unlike disembodied AIs, manipulation risks associated with neurotechnology involve the modification of underlying neurobiological functioning for the obtainment of emotional, cognitive or behavioural aims. An example is research on

neurotechnology for selective memory manipulation. Nabavi and colleagues used an optogenetics technique to erase and subsequently restore selected memories by applying a stimulus via optical laser that selectively strengthens or weakens synaptic connections [60]. As noted by Ienca and Andorno [61], the future sophistication and misuse of these techniques by malevolent actors may generate unprecedented opportunities for mental manipulation and brain-washing [61]. In particular, it has been observed that neurostimulation may have an impact on the psychological continuity of the person, i.e. the crucial requirement of personal identity consisting in experiencing oneself as persisting through time as the same person [57]. Consequently, by using neurostimulation it is possible, in principle, to manipulate the psychology of a person in manners that might affect that person's identity. It has been reported, for example, that invasive BCIs, such as DBS, may lead to behavioural changes such as increased impulsivity and aggressiveness [62], different taste in music [63] or changes in sexual behaviour [64]. Such induced behavioural changes might be of potential interest for state and non-state actors.

More subtle forms of manipulation based on non-invasive neurotechnology have also been discussed in the literature. An example is unconscious neural advertising via neuromarketing. Neuromarketing allows the use of techniques such as embedding subliminal stimuli with the purpose of eliciting responses (e.g. preferring item A instead of B) that people cannot consciously register. This has raised criticism among consumer advocate organizations, such as the Center for Digital Democracy, which have warned against neuromarketing's potentially manipulative application. Jeff Chester, the executive director of the organization, has claimed that "though there has not historically been regulation on adult advertising due to adults having defense mechanisms to discern what is true and untrue", it should be regulated "if the advertising is now purposely designed to bypass those rational defenses" [65].

15.11 Social Control and Discrimination

Neuromonitoring technology is vulnerable to the risk of being co-opted for surveillance and social control. The South China Morning Post has reported, for example, that in China state-backed neuroheadsets for EEG-based neuromonitoring are being deployed to detect changes in emotional states in three categories of individuals: public employees on the production line, the military and conductors of high-speed trains on the Beijing-Shanghai rail line [66]. Compelled use of neuromonitoring technology has raised concerns in terms of cognitive liberty and mental privacy. Authors have argued that every individual should be free to decide whether to use a certain neurotechnology application or refuse to do so, hence that coercive use should be prohibited [61]. Furthermore, the informational richness of brain data and their localization under the threshold of conscious control make it difficult for neurotechnology users to consciously segregate the information they want to seclude from what they want to share. Therefore, there is a risk that neuromonitoring activities might cause privacy breaches into a person's psychological life, hence resulting in violations of mental privacy. Mental privacy breaches can lead to discrimination in a twofold manner: either as a result of bias contained in the datasets or as the

purposive extraction from brain recordings of predictive information about health status and behaviour. For example, neural signatures of Alzheimer's disease or risk-taking behaviour can be used to discriminate individuals in manners that range from job termination to increased insurance premiums.

15.12 Military Applications of Neurotechnology

According to Tennison and Moreno, military applications of neurotechnology fall into three main categories: brain-computer interfaces (BCIs), neurotechnologies for warfighter enhancement, and neurotechnological systems for deception detection and interrogation [67]. The first category encompasses systems that establish a direct connection channel between the human brain and an external computer device, bypassing the peripheral nervous and muscular system. Military uses of BCIs include the acquisition of neural information gathered from warfighters' brains to adaptively modify their equipment and the development of threat warning systems that convert subconscious, neurological responses to danger into consciously available information [68]. Some authors refer to "disruptive BCIs" when they are planned to be used in an offensive manner, especially in a military setting such as the degradation and/or reading of enemy cognitive, sensory, motor neural activity [69]. These BCIs could be used, in the future, for torture or interrogation purposes, raising particular ethical questions.

Warfighter enhancement applications include various forms of transcranial electric stimulation technology such as transcranial direct current stimulation (tDCS) for selective cognitive enhancement in targeted brain areas. Finally, the deception detection domain encompasses devices such as the so-called "brain-fingerprints" capable of accessing concealed information in response to a stimulus. While these applications, especially those based on functional magnetic resonance (fMRI) and electroencephalography (EEG), hold great potential for medical diagnostics, they can be used as surveillance and interrogation tools for national security purposes. Unlike more rudimental interrogation technologies such as polygraph-based lie detection (based on the recording of extra-cranial physiological indices such as pulse and skin conductivity), brain-based lie detection technologies associate the truth-values of an uttered sentence or a mental state with specific patterns of brain activity.

The rise of network-centric warfare, a networked form of warfare relying on digital technologies, has increased the prevalence of hacking as a real threat to the capacity of armed forces to conduct operations. This concern can be extended to BCIs in ways which can be even more unsettling as we are speaking of hacking the cognitive, emotional and life-support functions of humans. This risk opens the prospect of "malicious brain-hacking", namely the "possibility of co-opting brain-computer interfaces and other neural engineering devices with the purpose of accessing or manipulating neural information from the brain of users" [52]. The ability to penetrate human brains through BCI will in fact add a new dimension to physical and cyber security and warfare in the future. This could, in the distant future, potentially lead to weapons that could "capture minds" (for example, via selective memory

manipulation, coercive neurostimulation or brain-to-brain control) with consequent implications not only for biosecurity [70], but also for human rights [61]. Artificial intelligence approaches such as deep learning have already been successfully used for neural control purposes in animal models involving monkeys [71].

15.13 Security Implications of Democratization of Access

As a consequence of decreasing hardware costs, improvements in sensorics and the increasing feasibility of developing portable EEG, functional near-infrared spectroscopy (fNIRS), transcranial electrical stimulation (tES) and transcranial magnetic stimulation (TMS) based neurotechnologies, the neurotechnology spectrum is not restricted to clinical and research applications, but includes a wide variety of direct-to-consumer systems [48]. This consumer neurotechnology trend is determining a proliferation and democratization of actors involved in the utilization of neurotechnologies. For instance, commercially available EEG-based consumer neurotechnologies start at about €120, hence making them affordable for many individuals globally [72]. Another sociotechnical trend known as do-it-yourself (DIY) neurotechnology has empowered non-professional individuals (often self-proclaimed biohackers) to self-assemble neurotechnology devices for personal use, most frequently via transcranial electrical stimulation for self-improvement purposes. Furthermore, as DTC neurotechnologies are typically utilized in absence of medical or other professional supervision, this proliferation also implies a reduced ability of authorities to monitor who is using neurotechnologies, how they are being used and for which purposes. Democratizing cognitive technology, neurotechnology in particular, is a laudable and to-be-pursued ethical goal [2] because it favours fair access and the just distribution of the societal benefits of this technology. Furthermore, it minimizes the risk that advantaged individuals, organized groups or states could achieve disproportionate control over the technology and use it for personal gain, surveillance or social control purposes at the expense of the majority of the population. At the same time, however, the proliferation of actors and the increased opacity of neurotechnology uses increase the statistical probability that these technologies might be used by malevolent actors for non-benign purposes. In light of these trends, authors have highlighted the urgent need for more agile, adaptive and systemic oversight mechanisms, neurosecurity standards, global governance frameworks and ethically aligned design via responsible innovation [2, 48, 50, 55, 70, 73].

15.14 Conclusion

This article has illustrated that the two families of cognitive technology, namely artificial intelligence and neurotechnology, are not only converging in terms of development and applicability, but also raising parallel security implications. In fact, both technologies hold great transformative potential due to their ability to

read, modify and amplify human cognition in a variety of domains and in response to a variety of cognitive and analytical tasks. Furthermore, both neurotechnology and artificial intelligence are rapidly proliferating outside of traditional supervised settings (e.g. clinics and academic research) onto multiple and unsupervised domains, a phenomenon that can be labelled "horizontal proliferation". Among these domains, their co-optation into the military sector and subsequent weaponization are of particular concern from an international security perspective. Similarly, as it is the case with artificial intelligence, proliferation also happens vertically, from state to non-state actors and individuals and vice versa. This is because of the dual-use nature of these technologies. Thus, it is extremely difficult to monitor and control the way they are used and, more importantly, misused. Indeed, with the ease of proliferation, one cannot exclude that these technologies will be used for malevolent purposes. This can already be observed with AI and deepfakes used to purposely modify satellite pictures, for instance [74].

In light of their disruptive potential and rapid proliferation, both neurotechnology and artificial intelligence urge global governance responses that deal with their accessibility, their proliferation, their dual-use nature including how easily these technologies can be repurposed and obviously, the ethics and values that should accompany the development and use of these technologies. These responses should be inclusive and comprise all the different stakeholders (governments, private sector, scientific community, civil society and tech companies) and be very versatile in that these technologies and applications evolve rapidly.

Acknowledgments Jean-Marc Rickli would like to thank Federico Mantellassi and Alexander Jahns for the background research conducted. Marcello Ienca would like to thank Fabrice Jotterand and Ralf Jox for their insightful comments to the research presented in this chapter.

Author contributions: JMR & MI conceived of the study and wrote the chapter. The two authors contributed equally.

References

1. Dascal M, Dror IE. The impact of cognitive technologies: towards a pragmatic approach. Pragmat Cogn. 2005;13(3):451–7.
2. Ienca M. Democratizing cognitive technology: a proactive approach. Ethics Inf Technol. 2019;21(4):267–80.
3. Ingham L. Stephen Hawking: the rise of powerful AI will be either the best or the worst thing ever to happen to humanity. A factor. 2018. https://www.factor-tech.com/feature/stephen-hawking-the-rise-of-powerful-ai-will-be-either-the-best-or-the-worst-thing-ever-to-happen-to-humanity/.
4. Moore BE. The brain computer interface future: time for a strategy. A research report submitted to the faculty. Air War College: Air War College Air University Maxwell AFB United States; 2013. https://apps.dtic.mil/dtic/tr/fulltext/u2/1018886.pdf.
5. Elements of AI. How should we define AI. 2019. https://course.elementsofai.com/1/1.
6. Independent High Level Expert Group on Artificial Intelligence. A definition of AI: main capabilities and disciplines: Brussels, European Commission; 2018.
7. SAS. Artificial intelligence, what it is and why it matters. 2019. https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html.

8. Jajal TD. Distinguishing between narrow AI, general AI and super AI. Medium; 2018.
9. Grace K, Salvatier J, Dafoe A, Zhang B, Evans O. When will AI exceed human performance? Evidence from AI experts. J Artif Intell Res. 2018;62:729–54.
10. Bostrom N. Superintelligence. Paths, dangers, strategies. Oxford: Oxford University Press; 2014.
11. Deepmind. AlphaStar: mastering the real-time strategy game StarCraft II. 2019. https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/.
12. Service RF. Google's deepmind aces protein folding. Science. 2018. https://www.sciencemag.org/news/2018/12/google-s-deepmind-aces-protein-folding.
13. Metz C. How Google's AI viewed the move no human could understand. Wired. 2016.
14. Brown N, Sandholm T, editors. Libratus: the superhuman AI for no-limit poker. In: Twenty-sixth international joint conference on artificial intelligence (IJCAI-2017); 2017.
15. Open AI. AI and compute. San Francisco: OpenAI; 2018.
16. Open AI. Better language models and their implications. San Francisco: OpenAI; 2019.
17. Agence France Press. Computer learns to detect skin cancer more accurately than doctors. The Guardian. 2018.
18. Yarnardag P. Normann: world's first psychopath AI. Cambridge: MIT; 2018. http://norman-ai.mit.edu.
19. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. Pro Publica. 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
20. Resnick B. Yes, artificial intelligence can be racist. VOX. 2019.
21. Kahn J. Major tech firms come out against police use of algorithms. Bloomberg. 2019. https://www.bloomberg.com/news/articles/2019-04-26/major-tech-firms-come-out-against-police-use-of-ai-algorithms.
22. Sentient. Understanding the "blackbox" of artificial intelligence. San Francisco: Sentient Technologies Holdings Limited; 2018. https://www.sentient.ai/blog/understanding-black-box-artificial-intelligence/.
23. Henschen D. How ML and AI will transform business intelligence analytics. ZDNet. 2018. https://www.zdnet.com/article/how-machine-learning-and-artificial-intelligence-will-transform-business-intelligence-and-analytics/.
24. Hern A. Cambridge analytica: how did it turn clicks into votes. The Guardian. 2018. https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie.
25. Cole S. Deep voice software can clone anyone's voice with just 3.7 seconds of audio. Motherboard. 2018. https://motherboard.vice.com/en_us/article/3k7mgn/baidu-deep-voice-software-can-clone-anyones-voice-with-just-37-seconds-of-audio.
26. Cauduro A. Live deep fakes—you can now change your face to someone else's in real time video applications. Medium. 2018. https://medium.com/huia/live-deep-fakes-you-can-now-change-your-face-to-someone-elses-in-real-time-video-applications-a4727e06612f.
27. Kissinger H. How the enlightenment ends. Atlantica. 2018. https://www.theatlantic.com/magazine/toc/2018/06/.
28. Joplin T. Long form: China's global surveillance-industrial complex. Albawaba News. 2018. https://www.albawaba.com/news/long-form-china's-global-surveillance-industrial-complex-1141152.
29. Mozur P. One month, 500,000 face scans: how China is using A.I. to profile a minority. Ney York Times. 2019. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificialintelligence-racial-profiling.html.
30. Scharre P. Robotics on the battlefield part II. Washington, DC: Center for a New American Security; 2014.
31. EHang. EHang Egret's 1374 drones dancing over the city wall of Xi'an, achieving a Guinness World Title. 2018. https://www.ehang.com/news/365.html.
32. Weaver D, Black E. Behind the scenes as Intel sets the world record for flying over 2000 drones at once. CNBC. 2018. https://www.cnbc.com/2018/07/17/intel-breaks-world-record-2018-drones.html.

33. Rickli J-M. The destabilizing prospects of artificial intelligence for nuclear strategy, deterrence and stability. In: Boulanin V, editor. The impact of artificial intelligence on strategic stability and nuclear risk: European perspectives. I. Stockholm: Stockholm International Peace Research Institute; 2019. p. 91–8. https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf.
34. Altmann J, Sauer F. Autonomous weapon systems and strategic stability. Survival. 2017;59(5):117–42.
35. Rickli J-M. The impact of autonomous weapons systems on international security and strategic stability. In: Ladetto Q, editor. Defence future technologies: what we see on the horizon. Thun: Armasuisse; 2017. p. 61–4. https://deftech.ch/What-We-See-On-The-Horizon/armasuisseW%2BT_Defence-Future-Technologies-What-We-See-On-The-Horizon-2017_HD.pdf.
36. Rickli J-M. The impact of autonomy and artificial intelligence on strategic stability. UN Special. 2018. p. 32–3. https://www.unspecial.org/2018/07/the-impact-of-autonomy-and-artificial-intelligence-on-strategic-stability/.
37. Roff H, Moyes R. Autonomy, robotics and collective systems. Tempe: Global Security Initiative, Arizona State University; 2016. https://globalsecurity.asu.edu/robotics-autonomy.
38. Krieg A, Rickli J-M. Surrogate warfare: the transformation of war in the twenty-first century. Georgetown: Georgetown University Press; 2019.
39. Rickli J-M. Education key to managing risk of emerging technology. European CEO. 2019. https://www.europeanceo.com/industry-outlook/education-key-to-managing-the-threats-posed-by-new-technology/.
40. Davis N, Rickli J-M. Submission to The Australian Council of Learned Academies and the Commonwealth Science Council on the opportunities and challenges presented by deployment of artificial intelligence. ACLO, Melbourne 2018.
41. Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, et al. The malicious use of artificial intelligence: forecasting, prevention, and mitigation. arXiv preprint arXiv:180207228. 2018. https://arxiv.org/abs/1802.07228.
42. Rickli J-M. The economic, security and military implications of artificial intelligence for the Arab Gulf Countries. Emirates Diplomatic Academy Policy Paper. 2018. https://www.gcsp.ch/News-Knowledge/Global-insight/The-Economic-Security-and-Military-Implications-of-Artificial-Intelligence-for-the-Arab-Gulf-Countries.
43. Allen G, Chan T. Artificial intelligence and national security. Cambridge: Belfer Center for Science and International Affairs; 2017. https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf.
44. Chovil P. Air superiority under 2000 feet: lessons from waging drone warfare against ISIL. War on the Rocks. 2018. https://warontherocks.com/2018/05/air-superiority-under-2000-feet-lessons-from-waging-drone-warfare-against-isil/.
45. Garden H, Bowman DM, Haesler S, Winickoff DE. Neurotechnology and society: strengthening responsible innovation in brain science. Neuron. 2016;92(3):642–6.
46. Giordano J. Neurotechnology: premises, potential, and problems. Boca Raton: CRC Press; 2012.
47. Abbott A. Neuroscience: solving the brain. Nature. 2013;499(7458):272.
48. Ienca M, Haselager P, Emanuel EJ. Brain leaks and consumer neurotechnology. Nat Biotechnol. 2018;36:805.
49. Behneman A, Berka C, Stevens R, Vila B, Tan V, Galloway T, et al. Neurotechnology to accelerate learning: during marksmanship training. IEEE Pulse. 2012;3(1):60–3.
50. Wexler A, Reiner PB. Oversight of direct-to-consumer neurotechnologies. Science. 2019;363(6424):234–5.
51. Dupont B. Cybersecurity futures: how can we regulate emergent risks? Technol Innov Manag Rev. 2013;3(7):6–11.
52. Ienca M, Haselager P. Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. Ethics Inf Technol. 2016;18(2):117–29.

53. Walther G. Weaponization of neuroscience. In: Clausen J, Levy N, editors. Handbook of neuroethics. Dordrecht: Springer; 2015. p. 1767–71.
54. Howell A. Neuroscience and war: human enhancement, soldier rehabilitation, and the ethical limits of dual-use frameworks. Millennium. 2017;45(2):133–50.
55. Ienca M, Vayena E. Dual use in the 21st century: emerging risks and global governance. Swiss Med Wkly. 2018;148:w14688.
56. Yuste R, Goering S, Agüera y Arcas B, Bi G, Carmena JM, Carter A, et al. Four ethical priorities for neurotechnologies and AI. Nature. 2017;551(7679):159–63.
57. Klaming L, Haselager P. Did my brain implant make me do it? Questions raised by DBS regarding psychological continuity, responsibility for action and mental competence. Neuroethics. 2013;6(3):527–39.
58. Ferretti A, Ienca M. Enhanced cognition, enhanced self? On neuroenhancement and subjectivity. J Cogn Enhancement. 2018;2(4):348–55.
59. Gilbert F. Deep brain stimulation: inducing self-estrangement. Neuroethics. 2018;11(2):157–65.
60. Nabavi S, Fox R, Proulx CD, Lin JY, Tsien RY, Malinow R. Engineering a memory with LTD and LTP. Nature. 2014;511(7509):348–52.
61. Ienca M, Andorno R. Towards new human rights in the age of neuroscience and neurotechnology. Life Sci Soc Policy. 2017;13(1):1–27.
62. Frank MJ, Samanta J, Moustafa AA, Sherman SJ. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science. 2007;318(5854):1309–12.
63. Mantione M, Figee M, Denys D. A case of musical preference for Johnny Cash following deep brain stimulation of the nucleus accumbens. Front Behav Neurosci. 2014;8:152.
64. Houeto JL, Mesnage V, Mallet L, Pillon B, Gargiulo M, du Moncel ST, et al. Behavioural disorders, Parkinson's disease and subthalamic stimulation. J Neurol Neurosurg Psychiatry. 2002;72(6):701–7.
65. Singer N. Making ads that whisper to the brain. New York Times. 2010.
66. Chen S. Forget the Facebook leak: China is mining data directly from workers' brains on an industrial scale. South China Morning Post. 2018.
67. Tennison MN, Moreno JD. Neuroscience, ethics, and national security: the state of the art. PLoS Biol. 2012;10(3):e1001289.
68. Miranda RA, Casebeer WD, Hein AM, Judy JW, Krotkov EP, Laabs TL, et al. DARPA-funded efforts in the development of novel brain–computer interface technologies. J Neurosci Methods. 2015;244:52–67.
69. Munyon CN. Neuroethics of non-primary brain computer interface: focus on potential military applications. Front Neurosci. 2018;12:696.
70. Ienca M, Jotterand F, Elger BS. From healthcare to warfare and reverse: how should we regulate dual-use neurotechnology? Neuron. 2018;97(2):269–74.
71. Bashivan P, Kar K, DiCarlo JJ. Neural population control via deep image synthesis. Science. 2019;364(6439):eaav9436.
72. Wexler A. The social context of "do-it-yourself" brain stimulation: neurohackers, biohackers, and lifehackers. Front Hum Neurosci. 2017;11:224.
73. Goering S, Yuste R. On the necessity of ethical guidelines for novel neurotechnologies. Cell. 2016;167(4):882–5.
74. Tucker P. The newest AI-enabled weapon: deep-faking photos of the earth. Defense One. 2019. https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/.