




Democratizing cognitive technology: a proactive approach

Marcello Ienca^{1,2} 

Published online: 19 June 2018

© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

Cognitive technology is an umbrella term sometimes used to designate the realm of technologies that assist, augment or simulate cognitive processes or that can be used for the achievement of cognitive aims. This technological macro-domain encompasses both devices that directly interface the human brain as well as external systems that use artificial intelligence to simulate or assist (aspects of) human cognition. As they hold the promise of assisting and augmenting human cognitive capabilities both individually and collectively, cognitive technologies could produce, in the next decades, a significant effect on human cultural evolution. At the same time, due to their dual-use potential, they are vulnerable to being coopted by State and non-State actors for non-benign purposes (e.g. cyberterrorism, cyberwarfare and mass surveillance) or in manners that violate democratic values and principles. Therefore, it is the responsibility of technology governance bodies to align the future of cognitive technology with democratic principles such as individual freedom, avoidance of centralized, equality of opportunity and open development. This paper provides a preliminary description of an approach to the democratization of cognitive technologies based on six normative ethical principles: avoidance of centralized control, openness, transparency, inclusiveness, user-centeredness and convergence. This approach is designed to universalize and evenly distribute the potential benefits of cognitive technology and mitigate the risk that such emerging technological trend could be coopted by State or non-State actors in ways that are inconsistent with the principles of liberal democracy or detrimental to individuals and groups.

Keywords Cognitive technology · Democratization · Open source · Open access · Neurotechnology · Artificial intelligence · Ethics · Governance

Cognitive technology

Cognitive technology (CT), also referred to as *cognition-related technology*, is an umbrella term used to designate the realm of technologies that assist, enhance or simulate cognitive processes or that can be used by humans “for the achievement of cognitive aims” (Dascal and Dror 2005).

The notion of CT was originally coined in the context of educational psychology to describe strategies and tools that could facilitate cognitive processes such as learning and problem solving (Sweller 1989). With advances in personal

computing, the notion of CT has been increasingly used to refer to “virtual environments, new computer devices and software tools” (Beynon et al. 2003) or other “*informational artifacts*” (Gorayska and Mey 1996) that can support or expand human cognition. An important step towards the establishment of CT as an area of scientific investigation was the creation in the late 1990s of a *Cognitive Technology Society* (Walker and Herrmann 2004) and the subsequent organization, during the early 2000s, of various CT-focused international conferences where experts from various fields of the cognitive sciences gathered to discuss “the impacts these technologies will have on human cognitive and social capacities” (Beynon et al. 2003).¹

In the last decade, in parallel with advances in Artificial Intelligence (AI), the label of CT has gained momentum in computer science and in the ICT industry to describe information technologies capable of performing cognitive tasks

¹ In 2001, the Coventry University organized a conference called “Cognitive Technology: Instruments of Mind” which marked an important milestone in the study of CT (Beynon et al. 2003).

✉ Marcello Ienca
Marcello.ienca@hest.ethz.ch

¹ Health Ethics & Policy Lab, Department of Health Sciences & Technology, ETH Zurich, Auf der Mauer 17, 8092 Zurich, Switzerland

² Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

traditionally performed by humans (Manuti and de Palma 2018; Schatsky et al. 2015), in particular when they are used to “assist and influence humans’ mental activities” (Kiger 2017). Among other companies, IBM has put CT at the center of their business transformation in what they called the “cognitive era”.²

CT is a macro-domain encompassing, at least, two major sub-domains:

- a. Neurotechnologies: Systems or devices that interface human nervous systems to assist, enhance or monitor natural cognitive processes.
- b. Artificial Intelligent Systems: Artificial systems that simulate (aspects of) intelligence and exhibit it across a wide range of processes including reasoning, planning, learning, natural language processing, perception and the ability to move and manipulate objects in the physical space.

Neurotechnologies include brain-computer interfaces (BCIs), electrical and magnetic brain stimulation, neurosensor-based vehicle operator systems, real-time neuromonitoring, neural prosthetics and others. These technologies are capable of establishing either invasive or non-invasive connection pathways between (human) nervous systems and computing devices for a variety of purposes. For example, medical applications of BCI technology have shown clinical effectiveness in monitoring, repairing, assisting or augmenting cognitive or sensory-motor functions in patients experiencing cognitive or sensory-motor impairments including spinal cord injury (Ikegami et al. 2011), stroke (Buch et al. 2008), motor neuron disease such as amyotrophic lateral sclerosis (ALS) and muscular dystrophy (Kübler et al. 2005; McCane et al. 2015), and, more recently, age-related cognitive decline (Lee et al. 2013), and dementia (Liberati et al. 2012). In parallel, direct-to-consumer applications of electroencephalography-based neuromonitoring are gaining increasing commercial interest as tools for self-monitoring, self-quantification as well as tools for physical and mental training.

Artificial intelligent systems include virtual personal assistants, question answering computer systems (such as IBM Watson), intelligent robots, self-repairing hardware and others. These systems mimic (components of) functions that humans usually associate with cognitive agents such as flexibility, automatic self-improvement through experience (as in the case of machine learning algorithms), perception of the external environment (e.g. speech recognition, facial

recognition, object recognition etc.), motion and manipulation (e.g. mapping, motion planning, path planning and localization) and knowledge representation.

Both neurotechnologies and artificial intelligent systems fall into the category of CT when they are utilized with the purpose of influencing, assisting, or augmenting human cognitive capacities. However, these two subdomains tend to differ with regard to how such an influence on cognition is realized. In most cases, neurotechnologies mostly affect cognition by intervening on “internal information processing systems”, i.e. by mapping or electrically modifying its underlying neurobiology. In contrast, artificial intelligent systems mostly intervene at the level of “external processing systems” (Bostrom and Sandberg 2009), that is they emulate (aspects of) human intelligence and provide external cognitive resources to support human cognition without any direct interface with the nervous system, a phenomenon known as *environmental enrichment* (Halperin and Healey 2011). Since the external processes enabled by artificial intelligent systems are, under some circumstances, functionally similar—according to some researchers, even equivalent—to internal processing, authors have argued that these technologies might be considered, under such circumstances, *extensions* of the human mind (Clark 2001; Clark and Chalmers 1998; Fitz and Reiner 2016). For example, Clark (p. 4) has argued that CTs “do far more than merely allow for the external storage and transmission of ideas” and rather “constitute [...] a cascade of mindware upgrades: cognitive upheavals in which the effective architecture of the human mind is altered and transformed” (Clark 2003).

In recent years, these two domains have experienced a strong convergence. In fact, AI features have been increasingly embedded in most advanced neurotechnologies. For example, most current BCIs use components of artificial intelligence, especially classifiers based on machine learning (ML) algorithms, to extract, classify and decode brain signals (Müller et al. 2008). At the same time, several artificial intelligent systems are provided with the capacity of being controlled via direct brain-machines interfaces or are designed to mimic the functioning of the human brain. These include smartphones and wearables (Powell et al. 2013), semi-autonomous cars (Göhring et al. 2013), unmanned aerial vehicles (Kosmyrna et al. 2015), and assistive robots (Tonin et al. 2011). This convergence is also occurring at market level with the increasing involvement in the neurotechnology sector of major players in artificial intelligence. For example, IBM, a major producer of artificial intelligent systems and developer of the famous intelligent digital assistant *Watson*, has entered the neurotechnology market and is among the top-15 patent holders in pervasive neurotechnology (Fernandez and Nikhil 2015). This market integration has even led to the creation of entire new research and business ventures precisely designed with the mission of

² See IBM’s best practices for cognitive technology: <https://www.ibm.com/watson/advantage-reports/getting-started-cognitive-technology.html>

accelerating the convergence of neurotechnology and artificial intelligent systems. An example of this trend is a newly launched venture called Neuralink. During the Code Conference 2016, entrepreneur Elon Musk announced a plan to accelerate the convergence between neurotechnology and artificial intelligence systems, followed by great media coverage. Although Musk himself remained cryptic about this project, he initially dubbed it “neural-lace” to emphasize the element of entwining brains and artificial systems together. In March 2017, Musk unveiled his project and launched Neuralink, a company whose stated mission is to “merge the human brain with AI” (Statt 2017).

It is worth to point out that CT is a functional characterization; hence it is not based on the type of hardware or software but on the type of function that a certain technology executes, namely assisting, supporting or expanding human cognitive capacities. Therefore, CT is creating an increasing need for addressing the ethical and social implications of CTs regardless of their hardware/software realization, but based on how these technologies influence human cognition.

This consideration has generated more interaction and dialogue among two main research communities: the Neuroethics community—primarily concerned with the ethics of neurotechnology (Goering and Yuste 2016; Illes and Bird 2006; Jotterand and Ienca 2017)—and the Computer Ethics community—primarily concerned with the ethics of computer systems and AI (Floridi 2010). The more technology is capable of interfacing, assisting and, possibly, expanding human cognition, there higher the need for comprehensive conceptual and normative approaches that study (the ethics of) cognition across the entire bio-digital continuum. Some early signs of convergence at the level of ethical and social assessment are already observable. For example, the 2016 Annual Meeting of the International Neuroethics Society in San Diego featured a public event on future and emerging technologies where a panel of experts discussed the ethical and social implications of both neurotechnologies and artificial intelligent systems such as care robots and intelligent digital assistants (Ienca 2016). Similarly, the 2017 IEEE TechEthics Conference in Washington D.C. (<https://techeethics.ieee.org/events/dc-2017>) featured one keynote talk and one panel on neurotechnology.

In light of the increasing convergence between these two main sub-domains, this paper will address the ethics and governance of cognitive technologies in a unitary manner.

Ethics, security and the dual-use dilemma

Some implications of cognitive technology have sparked ethical controversy. These include issues of cognitive enhancement and augmentation (Farah et al. 2004; Sententia 2004; Yuste et al. 2017), superhuman intelligence (Russell et al.

2015), agency and identity (Gilbert 2017; Yuste et al. 2017), human–machine hybridization (Ienca 2018), algorithmic bias (Kirkpatrick 2016; Yuste et al. 2017) and others. More recently, the application of CTs for purposes such as military dominance, surveillance, and cybercriminality has also associated CT to the ethical problem of dual-use (Ienca et al. 2018; Taddeo and Floridi 2018).

Dual-use technologies are artefacts that can be coopted “for making things quite unrelated to their primary purposes” (Forge 2010), in particular when these secondary purposes involve activities that are ethically questionable or potentially detrimental to individuals and groups such as military operations, terrorism, general criminality etc. In ethical terms, dual-use potentials inherent in technological artefacts are often presented as ethical conflicts between opposing ethical duties (Selgelid 2009); for example, between the promotion of good through free technological development vs. the prevention of possible collateral harm resulting from the cooptation of such technological potential for new purposes. A common example is the conflict between health promotion through effective clinical applications of a civil technology X vs. the provision of resources for the harming of innocents through military operations involving X.

Information technologies have instantiated a dual-use potential since their very first applications (Floridi 2014b). During the Second World War, Alan Turing’s early work on computability was coopted for military purposes, especially for the cryptanalysis of Morse-coded radio communications of the Axis powers enciphered using Enigma machines (Hodges 2012). The first contracts for packet network systems, including the development of the ARPANET, were awarded by the US Department of Defense as early as the 1960s and the first rogue program to spread through a network was created as early as in 1971.³ Today, several sub-components of the digital revolution—sometimes referred to as the “4th revolution” (Floridi 2014a), including networks, mobile communications technologies and robotics, demonstrably raise dual-use concerns.

Reports show that cyber-attacks have been growing in frequency and size in recent years. According to the Europol’s 2016 Internet Organised Crime Threat Assessment (IOCTA), cybercrime offences “remain on an upward trend and have reached very high levels” (Europol 2016). In October 2016, a massive cyber-attack targeted one of the central nodes of Internet traffic in the US, striking Twitter, Paypal, Spotify and sites of an infrastructure company in New Hampshire. Such increase in volume, scope and material cost of cybercrime has dramatically affected public perceptions on

³ The program was called the Creeper and spread through the early Bulletin Board networks (Ferbrache 1992).

information security. Survey data of the World Economic Forum's Global Risk Report 2016, show that cyber-attacks are perceived among the top five risks globally (WEF 2016). Increasingly, cyber-attacks have become a critical problem not only for private businesses, but also for public entities such as democratic governments (Mitterlehner 2014), healthcare institutions (Ehrenfeld 2017), and national security organizations (Nissenbaum 2005). Cyberterrorist acts have increased in number, magnitude and variety causing destruction and harm to personal computers, networks and the public Internet—including large-scale disruption of government systems, hospital records, and national security programs—for personal or ideological objectives (Matusitz 2005). When occurring between State actors, cyberoffences have shown the potential to influence geopolitical scenarios and strategic equilibria (Deibert 2015; Lacy and Prince 2018). A widely media-covered example is the role of cyber-attacks during the 2016 US presidential election culminated in the unprecedented hacking of a presidential candidate's email server and the following diplomatic crisis between the US and Russia (Stewart III, 2017). Concurrently, cyberwarfare concerns have emerged as a consequence of using CTs like artificial neural networks, gun data computers, secure cryptoprocessors, and robotics for military purposes (Gershgorin 2016; Sapaty 2015). The large-scale deployment of AI has been associated by experts with an increased risk to trigger a cyber arms race, which could ultimately escalate into conventional warfare (Taddeo and Floridi 2018). As Taddeo has observed, these emerging trends in cybercrime, cyberterrorism, and cyberwarfare “remark on the extent to which our societies depend on ICTs” and show how information technology has changed “the very infrastructure on which our societies rely” (Taddeo 2017). Following a socio-technological trend known as the Internet of Things (IoT), a large number of physical devices are becoming increasingly embedded with computing technology for a variety of purposes. Internetworked technologies embedded with electronics, software, sensors, actuators, and network connectivity are being tested or preliminary deployed by armed forces and governmental agencies (Callam 2015).

One common feature of these diverse cybercrime and cyberwarfare trends is that they often involve the use of computing systems with the deliberate purpose or unintended consequence of eroding basic democratic principles like individual freedoms, civil liberties, rule of law and democratic elections. This has raised the question of whether democratic principles and values will survive the digital era (Helbing et al. 2017).

This technology-mediated erosion of democratic principles is not exclusively caused by cyberterrorism and cyberwarfare. Global surveillance programs reportedly run by national security agencies and other governmental actors are also fueling controversies over the violation of

civil liberties and other democratic principles. Government agencies in various countries have proven able to deploy technology infrastructures for mass surveillance, enabling the collection of digital detritus—e-mails, calls, text messages, cellphone location data and a catalog of computer viruses, from individual citizens and groups. The government of China, for example, has reportedly installed over 20 million surveillance cameras across the country over the last few years and merged state surveillance with big data analytics to curb social unrest (Langfitt 2013). In 2014, the Chinese Ministry of Industry and Information Technology ordered a major mobile telephone company, to put a real name registration scheme into effect and to “regulate the dissemination of objectionable information over the network” (Limited 2014). In Russia, the Federal Security Service is legally allowed to use a system for Internet-based search and surveillance called *System for Operative Investigative Activities* (SORM). Since 2000, FSB is no longer required to provide telecommunications and Internet companies documentation on targets of interest prior to accessing information and in 2014 SORM-usage was extended to monitoring of social networks, chats and online forums (Paganini 2014). In response to these attempts of invasive governmental control, unauthorized disclosures of national security documents—as in the famous case of Edward J. Snowden vs. the United States' National Security Agency (NSA)—have been advocated by some authors as a proportionate response to preserve personal privacy and set the limits of invasive State-based surveillance (Lyon 2014).

This paper will argue that cognitive technologies can further “jeopardize democracy” (Vincent 2018) if they are not adequately aligned with fundamental democratic values and principles. I will proceed as follows. First, I will review dual-use issues associated with CT. Second, I will argue that the preferable approach to the governance of CT in light of dual-use risk is neither strict regulation nor *laissez-faire* but rather proactive democratization. In particular, I will argue that the potential held by CT for influencing human cognition urges the development of inclusive strategies that can direct cognitive technology for the benefit of people and the whole democratic society, not just restricted groups. Based on these considerations, I will outline six possible steps towards the proactive democratization of cognitive technology in the upcoming decade.

Dual-use cognitive technology

Cognitive technologies hold a promising potential for improving the life of human beings through a wide spectrum of non-hostile civil applications. For example, intelligent cognitive assistants are opening new possibilities for supporting people suffering from cognitive deficits such as

older people and people with dementia (Ienca et al. 2017; Jamieson et al. 2014). Similarly, BCIs are becoming increasingly effective in enabling novel opportunities for communication in patients suffering from stroke, spinal cord injury or amyotrophic lateral sclerosis (ALS) (Buch et al. 2008; Ikegami et al. 2011; Kübler et al. 2005).

At the same time, however, these technologies have recently shown some malleability to dual-use, especially in the context of military applications. In recent years, several global players including USA, EU, Russia, Iran, India, China and Japan have been actively working on military applications of neurotechnology, especially BCI (Moore 2013). Tennison and Moreno (2012) have comprehensively reviewed the spectrum of neurotechnologies with applications in military and national security contexts with special focus on projects funded via the United States' Defense Advanced Research Projects Agency (DARPA). Their review identified three main categories of dual-use neurotechnology: brain-computer interfaces (BCIs), neurotechnologies for warfighter enhancement, and neurotechnological systems for deception detection and interrogation (Tennison and Moreno 2012). In a similar fashion, Miranda et al. have assessed DARPA-funded BCI-applications for military purposes. Their review identifies two major avenues of ongoing research: (1) restoring neural and/or behavioral function in warfighters, and (2) enhancing training and performance in warfighters and intelligence agents (Miranda et al. 2015). For example, the *Neurotechnology for Intelligence Analysts (NIA)* program was designed to develop BCI systems utilizing non-invasively recorded EEG signals to significantly increase the efficiency and throughput of imagery analysis (Miranda et al. 2015). Using the same technological paradigm, national security uses of BCI include the acquisition of neural information gathered from warfighters' brains to modify their equipment accordingly and the development of a Cognitive Technology Threat Warning System (CT2WS) that convert subconscious, neurological responses to danger into consciously available information (Kirkpatrick 2007).

Current military applications of artificial intelligent systems mostly focus on non-cognitive applications such as unmanned aerial vehicles (UAVs—commonly known as a drones), unmanned ground vehicles (UGVs) such as the MIDARS, a four-wheeled robot that automatically performs random or preprogrammed patrols, and other autonomous or semi-autonomous robots such as Atlas, a bipedal humanoid robot designed for search and rescue tasks. In the near future, however, artificial intelligent systems will likely be used to augment physical and cognitive capacities of combatants. For example, by the end of 2017 the US Department of Defense is announced to launch the Tactical Assault Light Operator Suit (Talos), a military hardware that encloses soldiers within a computerized exoskeleton (White 2014). In parallel, augmented reality (AR) systems are being tested

with the purpose of enhancing attention, learning (Mao et al. 2017) and situational awareness (Gans et al. 2015). Particular ethical concern was raised by a special type of robotic applications, the so-called lethal autonomous weapons (LAWs). Unlike vehicles that are remote-controlled by a pilot or designed for non-combatting tasks such as reconnaissance, surveillance, and sniper detection, LAWs are designed to replace an important component of human cognition, namely decision-making.

Besides State-funded military applications, cognitive technologies have proven to hold dual-use potentials also in relation to non-State cyberterrorism and general cybercrime. Pycroft et al. (2016) have illustrated the possibility of targeting attacks against users of invasive neuromodulation technologies—especially deep brain stimulation (DBS), where the attackers may take control of the user's motor function, emotional dimension or simply disrupts the device's functionality (Pycroft et al. 2016). In experimental settings, Martinovic et al. (2012) have demonstrated the actual feasibility of performing side-channel attacks against users of currently marketed BCIs to reveal private and sensitive information about the users such as their pin-codes, bank membership, months of birth, debit card numbers, home location and faces of known persons (Martinovic et al. 2012). Hacking attacks have been proven feasible also against artificial intelligent systems, especially autonomous cars. The findings presented to the 2011 *National Academies Committee on Electronic Vehicle Controls and Unintended Acceleration* demonstrated the possibility of taking control of a car's computer system without direct physical access exploiting the car's Bluetooth connection (National Academies of Sciences 2012).

Finally, several cognitive technologies can be used as powerful surveillance tools for national security, judicial and military purposes due to their dual-use character. While no deception detection technology is being currently used in official security operations, several devices currently in-development either are directly DARPA-commissioned (Langleben et al. 2005, 2016) or market their services to national security agencies including the Department of Homeland Security such as the No Lie MRI device (Hughes 2010). This evidence shows that CTs can be potentially coopted for a number of purposes that involve the possible diminishment or even violation of democratic principles and values.

In this scenario, it is important for the future of democratic societies to anticipate possible challenges associated with the governance of cognitive technology and prevent that these systems can be coopted by malevolent governmental or non-governmental actors for anti-democratic aims including the triggering of a cyber arms race, the limitation of individual liberties, disproportionate mass-surveillance, the exacerbation of intra- and intergroup differences in social

dominance, or direct harm to individuals and groups. This risk is believed to be particularly cogent in light of the ongoing “shrinking” of Western democracy as a consequence of the recent rise of nationalism and authoritarian populism (Chacko and Jayasuriya 2017; Inglehart and Norris 2016). In such a rapidly changing global scenario, it is vital for democratic societies to prevent that cognitive technologies can be used to accelerate the crisis of democracy or to empower actors pursuing anti-democratic goals. In contrast, coordinated and proactive approaches are required to make sure that future developments of CT will be compatible with the principles of liberal democracy or even expand those principles through the human-centered permeation of such technologies in human societies. This paper proposes a preliminary characterization of the basic principles and safeguards to democratize cognitive technology in the upcoming decades.

Democratizing cognitive technology

Given their high dual-use potential, cognitive technologies have raised ethical concerns and elicited several proposals for policy response. Back in 2006, delegates of a workshop organized at Arizona State University addressed the issue of sociocultural risk in relation to cognitive technology.⁴ Their analysis identified in cognition-related technology a “capacity for sociocultural change” due to its potential to change human intelligence and performance capabilities, and anticipated that such potential could have destabilizing effects on individuals and groups (Sarewitz and Karas 2007). In the resulting white paper, experts delineated an entire spectrum of possible approaches to the governance and regulation of cognitive technologies that could prevent misuse and unintended risks. The two extremes of this spectrum were represented by the following options:

- a. Lassaiz-faire approaches—which emphasize the individual freedom of technology producers and end-users as well as the alleged capacity of financial markets to filter out potentially detrimental applications
- b. Strict regulatory approaches—which emphasize the need for State-led regulatory interventions (often based on essentialist views on human cognition according to which the natural cognitive boundaries should not be trespassed through technology)

⁴ The workshop and the resulting white paper adopted the label “technologies for cognitive enhancement” to describe a large variety of technological applications holding “capabilities to enhance human cognition” (Sarewitz and Karas 2007).

Lassaiz-faire approaches are being often advocated by producers of commercial neurotechnologies with the purpose of reducing FDA oversight on novel commercial products, especially limiting the applicability of FDA regulations on mobile medical applications to neurodevices for mental wellbeing.⁵ In contrast, particularly restrictive approaches were recently advocated by critics of dual-use artificial cognitive systems. The most restrictive of these approaches is the call for a collective ban or moratorium. While a collective ban is usually considered “much to extreme a response” in the context of dual-use neurotechnology (Giordano 2014), it has been advocated by a large number of experts in relation to LAW. Through the group *Campaign to Stop Killer Robots* (<https://www.stopkillerrobots.org/>) over 1000 experts in artificial intelligence signed an open letter calling for a global ban on LAWs arguing that it could trigger an arms race in military artificial intelligence and robotics.

This paper attempts to find a third way between these extreme approaches and argues that the best response to dual-use cognitive technology in a free society is a calibrated combination of technological freedom and risk-management strategies based on the principles of open development, responsible innovation and liberal democracy. I call this approach *democratization of cognitive technology*. In the following, I will describe this approach by delineating its core ethical principles and make a case for its implementation as a proactive strategy for the governance of CT and its accelerating impacts on human capabilities in a free society.

By *democratization* of a technological domain, I mean, very generally, a process of group decision-making about a certain technology characterized by the possibility of fair access to the technology by all participants and a principle of equality among the participants across various stages of the collective decision-making process.⁶ Consequently, democratizing cognitive technology implies a process of decision making about CT that will guarantee a possibility of fair access to CT for all users and a principle of equality among users during various stages of decision-making (including design, development and application).

In its general definition, this *democratizing* approach has elements of analogy with both strict-regulatory and lassaiz-faire approaches. With the strict-regulatory approaches it

⁵ During the 2012 Neurotech Leaders Forum, leaders of the neurotechnology industry and venture capital professionals discussed the impact of FDA approval cycles on commercialization of neurotechnology devices and investment in neurotechnology startups. They stated that “it was very difficult for them to invest in devices that require a premarket approval path through the FDA” due to “FDA tardiness in approving new devices” (Cavuoto 2012).

⁶ This definition of *democratization* is built upon the broad definition of *democracy* developed by T. Christiano. See Christiano (1993, 2004).

shares the observation that (i) cognitive technology requires urgent ethical assessment and policy interventions to minimize the risks associated with its dual-use potential, and (ii) that markets alone may not be conceptually and practically equipped to provide such assessment and intervention. This observation is based on a threefold factor.

First, novelty: cognitive technology is a relatively recent field of technological development. Consequently, it is still characterized by *conceptual muddles* and *policy vacuums* (Moor 2005) that prevent the maximization of benefits of these technologies while minimizing the risks. Many of these muddles and vacuums facilitate new opportunities for malicious exploitation generated by rapid changes in the technological or social environment, unprepared technological infrastructures, defective legal coverage, and the increase in quantity, variety and velocity of data flows (Dupont 2013).

Second, magnitude: CTs hold the potential of influencing human cognitive capabilities, hence determining a non-negligible effect on human cultural evolution and global equilibria. As observed by Moor (2005), neurotechnologies “could be the most revolutionary of all of the technologies” (Moor 2005) given their capacity to reconstruct, manipulate or augment cognitive processes, and impact human societies in manners that are currently difficult to predict. In the military context, B.E. Moore, lieutenant colonel of the United States Air Force, has predicted that BCI technology «has the potential to revolutionize military dominance much the same way nuclear weapons have done» (Moore 2013). Similar predictions have been made also in relation to artificial intelligent systems (Bostrom 2014). Due to its novelty, this alleged revolutionary potential of CT is still largely unexpressed. To date, for example, artificial intelligent systems are still distinguishable (from the Turing’s test perspective) from human intelligence across many cognitive tasks, while current neurotechnologies enable only a small degree of access to and modification of human neural processing. However, on the long term, the dual-use potential of CTs could enable unprecedented levels of intrusion into personal privacy or modification of personal autonomy (Ienca and Haselager 2016), concentration of economic power, and possibilities for offending individuals and groups (Dupont 2013; Yudkowsky 2008). As such, CTs could affect the fundamental mediators of human social interaction in the information era. Special oversight may be required to guarantee that these potentially revolutionary changes occur in accordance with the mechanisms and values of democratic societies.

The third factor is timing: given their historical novelty, cognitive technologies are still at an initial stage of market maturity and societal adoption. During this introduction phase, a technological trend shows a higher degree of malleability (Moor 2005). Therefore, control or change is less difficult to achieve compared to when the technology has become entrenched. Assumed that CT will be a critical

component of our future, human societies are now at a historic juncture in which they can make proactive decisions on the type of co-existence they want to establish with these technologies. Privileging *lassaiz-faire* approaches at this stage of development would defer risk-management interventions to a time when cognitive technology is extensively developed and widely used, hence refractory to modification.

At the same time, the democratizing approach shares with *lassaiz-faire* approaches the observation that over-regulation can (a) obliterate the benefits of cognitive technology for society at large, and, if managed by non-democratic or flawed democratic governments, (b) produce an undesirable concentration of power and control. In fact, if adequately implemented, CTs open the prospects of unparalleled improvement in the quality of life of human societies across a wide range of domains: medical, economic, infrastructural, communicational etc. For example, Russel et al. project that, thank to AI, “the eradication of disease and poverty is not unfathomable” (Russell et al. 2015). Therefore, over-regulatory strategies that limit technological freedom and open development could constrain technological progress and the resulting benefits for individuals and society at large. Second, top-down approaches to regulation could concentrate the power generated by CT among restricted political or economic groups, hence exacerbate existing political and economic inequalities. This risk has accompanied many breakthroughs in the history of technology. For example, during the introduction stage of the computer revolution, US authorities debated “whether a central government database for all United States citizens should be created” (Moor 2005). The creation of such government database would have produced a very different type of World Wide Web than the current one, with services distributed top-down, more concentration of power and control, and increased intrusion into individual privacy. The resulting decision not to create the data base contributed to the current informational landscape.

In the next section, I will describe a proactive democratizing approach to cognitive technology by delineating its core ethical principles. In addition, I will list, as an ostensive description, examples of currently ongoing projects and cooperative efforts that go into the direction of democratizing cognitive technology. It is worth noting that this description should not be seen as an exhaustive characterization of the ethics of CT or as a complete solution to the problems posed by dual-use dilemmas in cognitive technology. Of course, the answer to specific ethical dilemmas rising within this technological domain (e.g. trolley dilemmas for artificial intelligent agents, the personal autonomy of BCI users or the moral desirability of artificial superintelligence) may not necessarily depend on the level of democratic openness of the domain itself. Rather, this description is aimed at providing a preliminary conceptual and normative clarification of

the democratizing approach and opening a public debate on its realization.

Paths to democratization: the six principles

This proposal for democratizing cognitive technology consists of the combination of six normative principles:

- I. Avoidance of centralized control
- II. Openness
- III. Transparency
- IV. Inclusiveness
- V. User-centeredness
- VI. Convergence

These six principles condense and accentuate recurrent normative stances in the literature on the link between computing technology and democracy (Gil de Zúñiga et al. 2010; Helbing et al. 2017; Helbing and Pournaras 2015), and set out a way forward towards responsible and democratic development in cognitive technology. These principles can be used to guide the discussion on responsible innovation in CT at various levels of technology governance including individual researchers, funding agencies, as well as national and international regulatory bodies.

Avoidance of centralized control is the principle according to which it is morally preferable to avoid centralized control on CT to prevent risks associated with unrestricted accumulation of capital, power, and control over the technology among organized groups such as large corporations or governments. This preventive measure is designed to mitigate two critical types of technological risk. Type one risk: reduction in number of actors within the technological domain. To appreciate this type of risk, consider by analogy the transformation of the Internet over time, especially the transition from Web 1.0 to 2.0. While Web 1.0 was characterized by a coexistence of many service generators, the increase in data volumes and users typical of Web 2.0 is counterbalanced by a contraction of the number of actors, with most online traffic being driven by a limited number of powerful actors such as Google, Facebook, or YouTube. In the context of CT, the centralization of technological power among certain State or non-State actors could result in monopolistic operations or even destabilize economic, geopolitical and military dominance. In parallel, at the intra-state level, it could centralize power among restricted groups or elites hence potentially enable disproportionate control over the rest of the population and their civil liberties. I call this second scenario *type two risk* and can be conceptualized as an asymmetry between the level of governmental surveillance of individual citizens and the level of surveillance of governments by individual citizens.

Normative interventions aimed at limiting this risk of centralization may be conceptualized as cyberethical counterparts of anti-trust laws. Just like anti-trust laws are required to prevent monopolies and eliminate anti-competitive practices, proactive regulatory interventions may be required to prevent practices that restrain access to or development of CT, or cause the accumulation of power and control among restricted entities (Posner 2009). Such safeguards should apply to all societal actors and levels (including design, coding, and physical manufacturing), and are intended to allow smaller actors such as small groups or single individuals to enter the domain of cognitive technology and take advantage of its benefits. According the principle of avoiding centralized control, decentralized development models should be privileged over centralized models. Successful examples of decentralized development are open and participatory platforms such as the free encyclopedia Wikipedia, the open-source software operating system Linux (Helbing and Pournaras 2015) and the use of distributed ledger technology in trading and governance (Collomb and Sok 2016; Ølnes et al. 2017). An interesting attempt to implement the principle of decentralized control in the context of CT is *Nervousnet*, a large-scale distributed platform using sensor networks “to measure the world around us and to build a collective *data commons*”, which is often presented as a “digital nervous system” (Helbing and Pournaras 2015).

Openness is the principle of promoting universal access to (components of) the design or blueprint of cognitive technologies, and the universal redistribution of that design or blueprint, through an open and collaborative process of peer production. This principle also entails that the outputs of research in cognitive technology should be free of restrictions on access and use. Openness and the avoidance of control are critical requirements to make these same capabilities that will be recorded through or infused in cognitive technology—the cognitive capabilities—available to everyone. A good example in this direction is Microsoft’s effort to take those same capabilities infused in intelligent apps and made them available as a set of application programming interfaces (APIs) to every developer.⁷ This attempt is a form of democratization because it enables everyone to use the same *building blocks* that Microsoft uses to build intelligent devices or to make existing applications more intelligent. Another important step towards the democratization of cognitive technology through openness is Microsoft-sponsored research company *Open AI*. Open AI is a nonprofit company dedicated to precluding malicious AI,

⁷ For more detailed information on Microsoft’s approach see Microsoft Cognitive Services’ Documentation: <https://www.microsoft.com/cognitive-services/en-us/documentation>. Last accessed: 30 January 2017.

producing benevolent and safe AI, and ensuring that “AI’s benefits are as widely and evenly distributed as possible” (“Open AI” 2016). Examples of successful application of the openness principle have also emerged within the domain of neurotechnology, especially brain-computer interfacing. A positive example is *OpenBCI*, an open source brain-computer interface platform created by Joel Murphy and Conor Russomanno in 2013. Open BCI’s mission is to “provide anyone with a computer, the tools necessary to sample the electrical activity of their brains” and “harness the power of the open source movement to accelerate ethical innovation of human–computer interface technologies.”⁸ Today, Open BCI already offers an assortment of open source, versatile and affordable bio-sensing systems to sample electrical brain activity (EEG), some of which can be 3D printed. Development is open and new discoveries are made and shared through “an open forum of shared knowledge and concerted effort, by people from a variety of backgrounds.” Openness of cognitive technology has been seen as a critical strategy for harnessing *collective intelligence* (Helbing et al. 2017). In fact, a pervasive distribution of CT across all socioeconomic strata of society could empower people and enable a more informed and participative deliberation.

In a more abstract sense, openness in CT involves the principle of infusing every application that we interact with, on any device, at any point in time, with (components of) cognitive technology. This process is currently ongoing. For example, an increasing number of routinely used applications incorporate (components of) artificial intelligence. These include search engines, social media, e-commerce services, video-games, medical devices and many others. At the same time, an increasing number of applications are designed to interface human cognition through neurotechnology. For example, several mobile communication companies including Samsung and Apple are testing brain-controlled handheld devices (Powell et al. 2013). In this more general sense, openness is strictly linked to the avoidance of centralized control. In fact, the more cognitive capabilities are pervasively embedded and disseminated across the entire digital ecosystem, the harder it is for actors to centralize power and exert control over those systems. In the words of engineer and entrepreneur Elon Musk: “if everyone has AI powers, then there’s not any one person or a small set of individuals who can have AI superpower” (Mascarenhas 2016). Therefore, the principle of openness incentivizes the infusion of cognitive capabilities into an increasing number and variety of technologies in order to prevent their uneven accumulation among restricted applications or tools.

It is worth considering, however, that while “openness may reduce the probability of AI benefits being monopolized

by a small group” (Bostrom 2017) it could also cause unintended detrimental consequences. For example, Bostrom (2017) has argued that a high degree of openness could exacerbate a racing dynamic in which competitors trying to be the first to develop advanced AI may accept higher levels of existential risk in order to accelerate progress (Bostrom 2017). Further research is required to assess which degree of openness would ensure the optimal balance between benefits sharing and individual, national or international security.

Transparency is the principle of enabling a general public understanding of the internal processes of cognitive technologies. This is particularly challenging for approaches such as artificial neural networks, which learn or evolve to carry out a task in absence of clear mappings to chains of inference that are easy for humans to understand. This path to democratization through transparency is critical for artificial cognitive systems. For example, the principle of transparency is at core of IBM’s “Guiding Ethics Principles for the Cognitive Era”, a recently released ethics framework characterizing IBM’s digital transformation. According to this framework, “for cognitive systems to fulfil their world-changing potential”, it is vital to ensure the trust of end-users in the systems through transparency enhancing strategies (IBM-THINK 2017). In particular, there is a need for transparency in relation to (a) when and for what purposes AI is being applied in cognitive solutions, (b) the major sources of data and expertise “that inform the insights of cognitive solutions, as well as the methods used to train those systems and solutions”; (c) data protection and ownership. It is worth noting that the transparency principle has also educational relevance, since it allows making the necessary informational tools to learn and use cognitive technologies available for everyone, including students, workers and general citizens. Ideally, with advancing CT, such educational function will be institutionalized by the school system with the purpose of helping future citizens acquire the skills, knowledge and norms to engage successfully and securely with cognitive systems and use those skills and knowledge for achieving their life objectives.

An example of practical realization of algorithmic transparency is *Automatic Statistician*, an intelligent software capable of spotting trends and anomalies in data sets and presenting its conclusion, including a detailed explanation of its reasoning (Ghahramani 2015). According to the researcher who created this software, such *transparency* is “absolutely critical” not only for applications in science but also for many commercial applications (ibid). At the policy level, authors have linked the principle of transparency to public trust and proposed that “in order to create sufficient transparency and trust, leading scientific institutions should act as trustees of the data and algorithms that currently evade democratic control” (Helbing et al. 2017). This proposal would be particularly relevant in the context of data

⁸ See <http://openbci.com/>

and algorithms related to reasoning and decision-making as these could have a profound impact on individual deliberation and social cohesion.

Inclusiveness is the principle of ensuring that no group of individuals or minority is marginalized or left behind during the process of permeation of cognitive technology in our society. For example, a 2012 study co-authored by a senior FBI technologist, found that face recognition algorithms of commercial vendors consistently performed 5–10% worse on African Americans than on Caucasians (Klare et al. 2012). As the use of face recognition technology is expected to progress significantly in the next years, it is fundamental to ensure that no ethnic group will benefit from this technology less than other groups. The inclusiveness principle is at the core of The Algorithmic Justice League (AJL) was launched by Joy Buolamwini in November 2016. AJL provides a free platform to detect algorithmic bias that “can result in exclusionary experiences and discriminatory practices” and create “inclusive training sets.”⁹

The principle of inclusiveness does not apply exclusively to facial or physiognomic traits but to any other ethically relevant social bias that may intendedly or unintendedly emerge during CT development. These include cultural, political and language bias etc. An example of minimization of cultural and language bias is the internationalization strategy outlined by Open AI’s Software Requirements Specification. As the specification states: modules should be internationalized, in the sense that they “need to conform to the local language, locales, currencies etc., according to the settings specified in the configuration file or the environment in which they are running in.”¹⁰ The principle of inclusiveness is strictly related to the transparency principle. In fact, building algorithms which explain their reasoning and decision making is the best way to guarantee that hidden biases will be understood and promptly eliminated. In addition, it is important to create larger, more inclusive and diverse data sets with which to train the algorithms. Pluralism and diversity are critical notions for implementing the principle of inclusiveness in cognitive technology.

The principle of user-centeredness advocates that emerging cognitive technologies should be designed, developed and implemented according to the users’ needs and personal choices. User-centered approaches to the development of cognitive technology are necessary to guarantee that end-users (as widely as possible characterized, in accordance with the principles of openness and inclusiveness) are involved in the design, development and implementation of cognitive technologies on an equal footage. This principle has both methodological and social relevance.

Methodologically, user-centered approaches have been observed to increase the capacity of cognitive technology to fulfill the needs and wishes of end-users, reduce friction in human–machine interaction, facilitate usability hence increase overall user satisfaction (Ienca et al. 2017; Kübler et al. 2014). User-centered approaches have been observed to increase technology uptake and social adoption among end-users (De Vito Dabbs et al. 2009). Furthermore, such approaches ensure that technology is truly designed for the benefit of users instead of making users passive buyers of novel commercial products. For example, Kübler et al. have showed that user-centered design is a viable and effective approach to evaluate the usability of BCI-controlled applications, including among vulnerable end-users severe impairment (Kübler et al. 2014). Similar approaches have been pursued with BCIs based on event related potentials for brain spelling (Kaufmann et al. 2012) and painting (Zickler et al. 2013). User-centeredness is particularly important in relation to cognitive technologies developed for assisting patients with cognitive disorders. In fact, these people (e.g. older adults with dementia) are often frail and vulnerable individuals, hence entitled to utmost respect of their needs and wishes (Ienca et al. 2016). At the level of technology implementation, the principle of user-centeredness would prescribe increased individual control over one’s own cognitive processing and the adaptation of cognitive technologies to the needs, wishes and capabilities of individual users.

Finally, the principle of convergence can be described both in a narrow and in a broad sense. In the narrow sense, convergence is the principle of interoperability, intercommunication and ease of integration among all components of cognitive technology (i.e. the cognitive *tools* or *modules*): in order to reach the common goal of measuring, enhancing or emulating cognition, all cognitive tools must, at some important level, speak the same language and behave in a mutually consistent manner. It is worth noting, however, that excessive interoperability might result in increased data insecurity, hence must be carefully balanced over other ethical principles and technical safeguards. In a broader and more abstract sense, it is also the principle of converging different types of cognitive technology, especially neurotechnology, on the one hand, and artificial intelligent systems on the other hand. As described in the first section of this paper, such convergence is already occurring. For example, BCIs have been combined with artificial intelligent systems (environment-sensing, obstacle-avoidance and pathfinding capabilities) to achieve shared control and context based filtering of user commands, hence enhance the overall performance of the brain–machine combination (Millán et al. 2010; Tonin et al. 2010). In addition, a proposals to make this link closer and more reliable via brain-computer interaction are being pursued by various companies including Facebook, Neuralink, Kernel and Emotiv (Gent 2017). Similar convergence-aimed solutions have

⁹ See <http://www.ajlunited.org/the-coded-gaze>

¹⁰ See <http://openai.sourceforge.net/OpenAI-srs.html>

been pioneered also at the microscopic level. A promising example is a minimally invasive three-dimensional interpenetration of electronics within artificial structures or biological brains (Liu et al. 2015). Such mesh-brain implants have already demonstrated the capacity to successfully integrate into a mouse brain and enable neuronal recordings (Fu et al. 2016). While convergence in the narrow sense is necessary to guarantee the successful functioning of cognitive technology, broad-sense convergence might, on the medium-to-long term, empower individuals and provide ultimate control and protection against malevolent applications of cognitive technology.

Conclusions

Cognitive technologies have the potential to accelerate technological innovation and provide significant benefit for individuals and societies. At the same time, due to their dual-use potential, they can be potentially coopted by State and non-State actors for non-benign purposes including cybercrime, cyberterrorism, cyberwarfare and mass surveillance. In light of the recent global crisis of democracy, increased militarization of the digital infosphere, and concurrent potentiation of cognitive technologies, it is important to proactively design strategies that can mitigate emerging risks and align the future of CT with the basic principles of liberal democracy in free and open societies.

In this paper, I described a proactive approach to the democratization of CT based on six normative ethical principles: avoidance of centralized control, openness, transparency, inclusiveness, user-centeredness and convergence. This approach is designed to universalize and evenly distribute the potential benefits of CT and mitigate the risk that such emerging technological trend could be coopted by State or non-State actors in ways that are inconsistent with the principles of liberal democracy or detrimental to individuals and groups. While this paper offered a preliminary and general characterization of how to democratize cognitive technology, future research is required to expand this proposal into a comprehensive ethical, legal and political framework.

Compliance with ethical standards

Conflict of interests The author declares no conflict of interest.

References

Beynon, M., Nehaniv, C. L., & Dautenhahn, K. (2003). Cognitive Technology. Instruments of Mind. In 4th International Conference, CT 2001 Coventry, UK, August 6–9, 2001 Proceedings (Vol. 2117), Springer.

- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: OUP
- Bostrom, N. (2017). Strategic implications of openness in AI development. *Global Policy*, 8(2), 135–148.
- Bostrom, N., & Sandberg, A. (2009). Cognitive enhancement: Methods, ethics, regulatory challenges. *Science and Engineering Ethics*, 15(3), 311–341. <https://doi.org/10.1007/s11948-009-9142-5>.
- Buch, E., Weber, C., Cohen, L. G., Braun, C., Dimyan, M. A., Ard, T., ... Fourkas, A. (2008). Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke. *Stroke*, 39(3), 910–917.
- Callam, A. (2015). Drone wars: Armed unmanned aerial vehicles. *International Affairs Review*, 18(3), 122–132.
- Cavuoto, J. (2012). Regulatory efficacy. Neurotech Report. Retrieved on December 12, 2017 from <http://www.neurotechreports.com/pages/publishersletterOct12.html>.
- Chacko, P., & Jayasuriya, K. (2017). Trump, the authoritarian populist revolt and the future of the rules-based order in Asia. *Australian Journal of International Affairs*, 1–7.
- Christiano, T. (1993). *Social choice and democracy. The idea of democracy* (pp. 173–195). Cambridge: Cambridge University Press.
- Christiano, T. (2004). The authority of democracy. *Journal of Political Philosophy*, 12(3), 266–290.
- Clark, A. (2001). Reasons, robots and the extended mind. *Mind & Language*, 16(2), 121–145.
- Clark, A. (2003). *Natural-Born Cyborgs: Minds, technologies, and the future of human intelligence*. Oxford: Oxford University Press.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7–19.
- Collomb, A., & Sok, K. (2016). Blockchain/Distributed Ledger Technology (DLT): What Impact on the Financial Sector? *Communications & Strategies*, 103, 93.
- Dascal, M., & Dror, I. E. (2005). The impact of cognitive technologies: Towards a pragmatic approach. *Pragmatics & Cognition*, 13(3), 451–457.
- De Vito Dabbs, A., Myers, B. A., Curry, M., Dunbar-Jacob, K. R., Hawkins, J., Begey, R. P. A., & Dew, M. A. (2009). User-centered design and interactive health technologies for patients. *Computers, Informatics, Nursing: CIN*, 27(3), 175. <https://doi.org/10.1097/NCN.0b013e31819f7c7c>.
- Deibert, R. (2015). The geopolitics of cyberspace after Snowden. *Current History*, 114(768), 9.
- Dupont, B. (2013). Cybersecurity futures: How can we regulate emergent risks? *Technology Innovation Management Review*, 3(7), 6.
- Ehrenfeld, J. M. (2017). Wannacry, cybersecurity and health information technology: A time to act. *Journal of Medical Systems*, 41(7), 104.
- Europol. (2016). 2016 Internet Organised Crime Threat Assessment (IOCTA) (pp. 0890–8044). Retrieved from <https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2016>.
- Farah, M. J., Illes, J., Cook-Deegan, R., Gardner, H., Kandel, E., King, P., ... Wolpe, P. R. (2004). Neurocognitive enhancement: what can we do and what should we do? *Nature Reviews Neuroscience*, 5(5), 421.
- Ferbrache, D. (1992). Historical perspectives. In *A pathology of computer viruses* (pp. 5–30). London: Springer.
- Fernandez, A., & Nikhil, S. (2015). *Pervasive neurotechnology*. San Francisco: SharpBrains.
- Fitz, N. S., & Reiner, P. B. (2016). Perspective: Time to expand the mind. *Nature*, 531(7592), S9–S9. <https://doi.org/10.1038/531S9a>.
- Floridi, L. (2010). *The Cambridge handbook of information and computer ethics*. Cambridge: Cambridge University Press.
- Floridi, L. (2014a). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford: OUP.

- Floridi, L. (2014b). The latent nature of global information warfare. *The Philosophers' Magazine*, 67, 17–19.
- Forge, J. (2010). A note on the definition of “Dual Use”. *Science and Engineering Ethics*, 16(1), 111–118. <https://doi.org/10.1007/s11948-009-9159-9>.
- Fu, T.-M., Hong, G., Zhou, T., Schuhmann, T. G., Viveros, R. D., & Lieber, C. M. (2016). Stable long-term chronic brain mapping at the single-neuron level. *Nature Methods*, 13(10), 875–882. <https://doi.org/10.1038/nmeth.3969>.
- Gans, E., Roberts, D., Bennett, M., Towles, H., Menozzi, A., Cook, J., & Sherrill, T. (2015). Augmented reality technology for day/night situational awareness for the dismounted Soldier. In *Display technologies and applications for defense, security, and avionics IX; and head-and-helmet-mounted displays XX* (Vol. 9470, p. 947004). Bellingham: International Society for Optics and Photonics.
- Gent, E. (2017). Brain-hacking tech gets real: 5 companies leading the charge. Retrieved on August 2, 2017 from <https://www.livescience.com/59326-companies-investing-in-brain-hacking-tech.html>.
- Gershgorin, D. (2016). The US government seriously wants to weaponize artificial intelligence. Quartz. Retrieved from <https://qz.com/767648/weaponized-artificial-intelligence-us-military/>.
- Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553), 452–459. <https://doi.org/10.1038/nature14541>.
- Gilbert, F. (2017). Deep brain stimulation: Inducing self-estrangement. *Neuroethics*, 11(2), 157–165.
- Gil de Zúñiga, H., Veenstra, A., Vraga, E., & Shah, D. (2010). Digital democracy: Reimagining pathways to political participation. *Journal of Information Technology & Politics*, 7(1), 36–51.
- Giordano, J. (2014). *Neurotechnology in National Security and Defense: Practical Considerations*. Neuroethical Concerns: CRC Press.
- Goering, S., & Yuste, R. (2016). On the necessity of ethical guidelines for novel neurotechnologies. *Cell*, 167(4), 882–885.
- Göhring, D., Latotzky, D., Wang, M., & Rojas, R. (2013). Semi-autonomous car control using brain computer interfaces. *Intelligent Autonomous Systems*, 12, 393–408.
- Gorayska, B., & Mey, J. L. (1996). Cognitive technology. In K. S. Gill (Ed.), *Information society: New Media, ethics and postmodernism* (pp. 287–294). London: Springer.
- Halperin, J. M., & Healey, D. M. (2011). The influences of environmental enrichment, cognitive enhancement, and physical exercise on brain development: Can we alter the developmental trajectory of ADHD? *Neuroscience and Biobehavioral Reviews*, 35(3), 621–634. <https://doi.org/10.1016/j.neubiorev.2010.07.006>.
- Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., ... Zwitter, A. (2017). Will democracy survive big data and artificial intelligence. *Scientific American*, p. 25.
- Helbing, D., & Pournaras, E. (2015). Build digital democracy: Open sharing of data that are collected with smart devices would empower citizens and create jobs. *Nature*, 527(7576), 33–35.
- Hodges, A. (2012). *Alan Turing: the enigma*. New York: Random House.
- Hughes, V. (2010). Head case. *Nature*, 464(7287), 340.
- IBM-THINK. (2017). Transparency and trust in the cognitive era. Retrieved on August, 2017 from <https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/>.
- Ienca, M. (2016). Meet tomorrow's world: A meeting on the ethics of emerging technologies. Retrieved on March 26, 2017 from <http://www.theneuroethicsblog.com/2016/12/meet-tomorrows-world-a-meeting-on-ethics.html>.
- Ienca, M. (2018). Cognitive technology and human-machine interaction: The contribution of externalism to the theoretical foundations of machine and cyborg ethics. *Annals of the University of Bucharest—Philosophy Series*; Vol 66 No 2 (2017): Annals of the University of Bucharest: Philosophy Series.
- Ienca, M., Fabrice, J., Elger, B., Caon, M., Pappagallo, A. S., Kressig, R. W., & Wangmo, T. (2017). Intelligent assistive technology for Alzheimer's disease and other dementias: A systematic review. *Journal of Alzheimer's Disease*, 56(4), 1301–1340. <https://doi.org/10.3233/jad-161037>.
- Ienca, M., & Haselager, P. (2016). Hacking the brain: Brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, 18(117), 117–129.
- Ienca, M., Jotterand, F., & Elger, B. S. (2018). From healthcare to warfare and reverse: How should we regulate dual-use neurotechnology? *Neuron*, 97(2), 269–274. <https://doi.org/10.1016/j.neuron.2017.12.017>.
- Ienca, M., Jotterand, F., Vică, C., & Elger, B. (2016). Social and assistive robotics in dementia care: Ethical recommendations for research and practice. *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-016-0366-7>.
- Ikegami, S., Takano, K., Saeki, N., & Kansaku, K. (2011). Operation of a P300-based brain-computer interface by individuals with cervical spinal cord injury. *Clinical Neurophysiology*, 122(5), 991–996.
- Illes, J., & Bird, S. J. (2006). Neuroethics: A modern context for ethics in neuroscience. *Trends in neurosciences*, 29(9), 511–517.
- Inglehart, R., & Norris, P. (2016). Trump, Brexit, and the rise of Populism: Economic have-nots and cultural backlash. KS Working Paper No. RWP16-026. Available at SSRN: <https://ssrn.com/abstract=2818659> or <https://doi.org/10.2139/ssrn.2818659>. Retrieved on April 12, 2017 from <http://wotantue.us/Trump-Brexit-Populism.pdf>.
- Jamieson, M., Cullen, B., McGee-Lennon, M., Brewster, S., & Evans, J. J. (2014). The efficacy of cognitive prosthetic technology for people with memory impairments: A systematic review and meta-analysis. *Neuropsychological Rehabilitation*, 24(3–4), 419–444.
- Jotterand, F., & Ienca, M. (2017). The Biopolitics of neuroethics. In E. Racine & J. Aspler (Eds.), *Debates about neuroethics: perspectives on its development, focus, and future* (pp. 247–261). Cham: Springer.
- Kaufmann, T., Völker, S., Gunesch, L., & Kübler, A. (2012). Spelling is just a click away – a user-centered brain-computer interface including auto-calibration and predictive text entry. *Frontiers in Neuroscience*. <https://doi.org/10.3389/fnins.2012.00072>.
- Kiger, P. J. (2017). 5 Ways society will be affected by cognitive technology. Retrieved on December 12, 2017 from <http://electronicshustuffworks.com/future-tech/5-ways-society-will-be-affected-by-cognitive-technology.htm>.
- Kirkpatrick, D. (2007). Cognitive technology threat warning systems (CT2WS). Retrieved on August 2, 2017 from <https://web.archive.org/web/20080204203721/http://www.darpa.mil/baa/BAA07-25.html>.
- Kirkpatrick, K. (2016). Battling algorithmic bias: How do we ensure algorithms treat us fairly? *Communications of the ACM*, 59(10), 16–17.
- Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V., & Jain, A. K. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security*, 7(6), 1789–1801. <https://doi.org/10.1109/TIFS.2012.2214212>.
- Kosmyna, N., Tarpin-Bernard, F., & Rivet, B. (2015). Towards brain computer interfaces for recreational activities: Piloting a drone. Paper presented at the Human-Computer Interaction.
- Kübler, A., Holz, E. M., Riccio, A., Zickler, C., Kaufmann, T., Kleih, S. C., ... Mattia, D. (2014). The user-centered design as novel perspective for evaluating the usability of BCI-controlled

- applications. *PLoS ONE*, 9(12), e112392. <https://doi.org/10.1371/journal.pone.0112392>.
- Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T. M., Pawelzik, H., Schalk, G., ... Wolpaw, J. R. (2005). Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10), 1775–1777.
- Lacy, M., & Prince, D. (2018). Securitization and the global politics of cybersecurity. *Global Discourse*, 8(1), 100–115.
- Langfitt, F. (2013). China beware: A camera may be watching you. *NPR*, 29, 40.
- Langleben, D. D., Hakun, J. G., Seelig, D., Wang, A.-L., Ruparel, K., Bilker, W. B., & Gur, R. C. (2016). Polygraphy and functional magnetic resonance imaging in lie detection: A controlled blind comparison using the concealed information test. *The Journal of Clinical Psychiatry*, 77(10), 1372–1380.
- Langleben, D. D., Loughhead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, 26(4), 262–272. <https://doi.org/10.1002/hbm.20191>.
- Lee, T.-S., Goh, S. J. A., Quek, S. Y., Phillips, R., Guan, C., Cheung, Y. B., ... Chin, Z. Y. (2013). A brain-computer interface based cognitive training system for healthy elderly: A randomized control pilot study for usability and preliminary efficacy. *PLoS ONE*, 8(11), e79419.
- Liberati, G., Dalboni da Rocha, J. L., van der Heiden, L., Raffone, A., Birbaumer, N., Belardinelli, O. M., & Sitaram, R. (2012). Toward a brain-computer interface for Alzheimer's disease patients by combining classical conditioning and brain state classification. *Journal of Alzheimer's Disease*, 31(Suppl 3), 211–220. <https://doi.org/10.3233/jad-2012-112129>.
- Limited, C. T. C. (2014). China telecom 2014 annual work conference highlights [Press release]. Retrieved on August 2, 2017 from <http://www.irasia.com/listco/hk/chinatelecom/press/p140103.htm>.
- Liu, J., Fu, T.-M., Cheng, Z., Hong, G., Zhou, T., Jin, L., ... Lieber, C. M. (2015). Syringe-injectable electronics. *Nature Nanoelectronics*, 10(7), 629–636. <https://doi.org/10.1038/nnano.2015.115>.
- Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society*, 1(2), 2053951714541861.
- Manuti, A., & de Palma, P. D. (2018). The cognitive technology revolution: A new identity for workers. In A. Manuti & P. D. de Palma (Eds.), *Digital HR: A critical management approach to the digitization of organizations* (pp. 21–37). Cham: Springer.
- Mao, C.-C., Chen, C.-H., & Sun, C.-C. (2017). Impact of an augmented reality system on learning for army military decision-making process (MDMP) course. In M. Soares, C. Falcão, & T. Z. Ahrum (Eds.), *Advances in ergonomics modeling, usability & special populations: Proceedings of the AHFE 2016 international conference on ergonomics modeling, usability & special populations, July 27–31, 2016, Walt Disney World®, Florida, USA* (pp. 663–671). Cham: Springer.
- Martinovic, I., Davies, D., Frank, M., Perito, D., Ros, T., & Song, D. (2012). On the feasibility of side-channel attacks with brain-computer interfaces. Paper presented at the USENIX Security Symposium.
- Mascarenhas, H. (2016). Elon Musk's \$1bn non-profit launches 'gym' to train AI with Atari games. *International Business Times*.
- Matusitz, J. (2005). Cyberterrorism: How can American foreign policy be strengthened in the Information Age? *American Foreign Policy Interests*, 27(2), 137–147.
- McCane, L. M., Heckman, S. M., McFarland, D. J., Townsend, G., Mak, J. N., Sellers, E. W., ... Vaughan, T. M. (2015). P300-based brain-computer interface (BCI) event-related potentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-matched controls. *Clinical Neurophysiology*, 126(11), 2124–2131.
- Millán, J. d. R., Rupp, R., Mueller-Putz, G., Murray-Smith, R., Giugliemina, C., Tangermann, M., ... Leeb, R. (2010). Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in Neuroscience*, 4, 161.
- Miranda, R. A., Casebeer, W. D., Hein, A. M., Judy, J. W., Krotkov, E. P., Laabs, T. L., ... Sanchez, J. C. (2015). DARPA-funded efforts in the development of novel brain-computer interface technologies. *Journal of Neuroscience Methods*, 244, 52–67.
- Mitterlehner, B. (2014). *Cyber-Democracy and Cybercrime: Two Sides of the Same Coin. Cyber-Development, Cyber-Democracy and Cyber-Defense* (pp. 207–230). New York: Springer.
- Moor, J. H. (2005). Why we need better ethics for emerging technologies. *Ethics and Information Technology*, 7(3), 111–119.
- Moore, B. E. (2013). *The brain computer interface future: time for a strategy*. Alabama: Air University.
- Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1), 82–90.
- National Academies of Sciences, E., and Medicine. (2012). *The safety promise and challenge of automotive electronics insights from unintended acceleration*. Washington, DC: The National Academies Press.
- Nissenbaum, H. (2005). Where computer security meets national security. *Ethics and Information Technology*, 7(2), 61–73.
- Øines, S., Ubacht, J., & Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. *Government Information Quarterly*, 34(3), 355–364. <https://doi.org/10.1016/j.giq.2017.09.007>.
- Open, A. I. (2016). Retrieved from <https://openai.com/about/>.
- Paganini, P. (2014). New powers for the Russian surveillance system SORM-2. Security Affairs. Retrieved on August 2, 2017 from <http://securityaffairs.co/wordpress/27611/digital-id/new-power-s-sorm-2.html>.
- Posner, R. A. (2009). *Antitrust law*: University of Chicago Press.
- Powell, C., Munetomo, M., Schlueter, M., & Mizukoshi, M. (2013). Towards thought control of next-generation wearable computing devices. Paper presented at the International Conference on Brain and Health Informatics.
- Pycroft, L., Boccard, S. G., Owen, S. L. F., Stein, J. F., Fitzgerald, J. J., Green, A. L., & Aziz, T. Z. (2016). Brainjacking: Implant Security Issues in Invasive Neuromodulation. *World Neurosurgery*, 92, 454–462. <https://doi.org/10.1016/j.wneu.2016.05.010>.
- Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, 36(4), 105–114.
- Sapaty, P. (2015). Military robotics: Latest trends and spatial grasp solutions. *International Journal of Advanced Research in Artificial Intelligence*, 4(4), 9–18.
- Sarewitz, D., & Karas, T. H. (2007). Policy implications of technologies for cognitive enhancement. SAND2006-7909. Sandia National Laboratories Albuquerque, New Mexico 87185. Retrieved from <http://prod.sandia.gov/techlib/access-control.cgi/2006/067909.pdf>.
- Schatsky, D., Muraskin, C., & Gurumurthy, R. (2015). Cognitive technologies: The real opportunities for business. *Deloitte Review*, 16, 115–129.
- Selgelid, M. J. (2009). Governance of dual-use research: an ethical dilemma. *Bulletin of the World Health Organization*, 87(9), 720–723.
- Sentientia, W. (2004). Neuroethical considerations: cognitive liberty and converging technologies for improving human cognition. *Annals of the New York Academy of Sciences*, 1013(1), 221–228.

- Statt, N. (2017). Elon Musk launches Neuralink, a venture to merge the human brain with AI. The Verge. Retrieved on August 2, 2017 from <http://www.theverge.com/2017/3/27/15077864/elon-musk-neuralink-brain-computer-interface-ai-cyborgs>.
- Stewart, I. I. I. C. (2017). Electoral Vulnerabilities in the United States: Past, Present, and Future. MIT Political Science Department Research Paper(5).
- Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, 81(4), 457.
- Taddeo, M. (2017). Cyberwar: How to regulate nation state warfare on the internet. Retrieved on February 12, 2018 from <http://www.scienceviewsthenews.com/cyberwar-how-to-regulate-nation-state-warfare-on-the-internet/>.
- Taddeo, M., & Floridi, L. (2018). Regulate artificial intelligence to avert cyber arms race. *Nature*, 556(7701), 296–298. <https://doi.org/10.1038/d41586-018-04602-6>.
- Tennison, M. N., & Moreno, J. D. (2012). Neuroscience, ethics, and national security: The state of the art. *PLoS Biology*, 10(3), e1001289. <https://doi.org/10.1371/journal.pbio.1001289>.
- Tonin, L., Carlson, T., Leeb, R., & Millán, J. d. R. (2011). Brain-controlled telepresence robot by motor-disabled people. Paper presented at the Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE.
- Tonin, L., Leeb, R., Tavella, M., Perdakis, S., & Millán, J. d. R. (2010). The role of shared-control in BCI-based telepresence. Paper presented at the Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on.
- Vincent, J. (2018). *Badly implemented AI could 'jeopardize democracy'*. The Verge.
- Walker, W. R., & Herrmann, D. J. (2004). *Cognitive technology: Essays on the transformation of thought and society*. Jefferson, NC: McFarland.
- WEF (2016). *The global risks report 2016* (11th ed.). World Economic Forum. Retrieved from <https://www.weforum.org/reports/the-global-risks-report-2016>.
- White, A. (2014). Future special operations protection systems (tactical assault light operator suit). *Military Technology*, 38(12), 70–73.
- Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global Catastrophic Risks*, 1(303), 184.
- Yuste, R., Goering, S., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., ... Illes, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature News*, 551(7679), 159.
- Zickler, C., Halder, S., Kleih, S. C., Herbert, C., & Kübler, A. (2013). Brain painting: Usability testing according to the user-centered design in end users with severe motor paralysis. *Artificial Intelligence in Medicine*, 59(2), 99–110. <https://doi.org/10.1016/j.artmed.2013.08.003>.