Neuroethics (2021) 14:83–98
https://doi.org/10.1007/s12152-018-9371-x

ORIGINAL PAPER

Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices

Philipp Kellmeyer

Received: 5 February 2018 / Accepted: 11 May 2018 / Published online: 19 May 2018

Abstract The focus of this paper are the ethical, legal and social challenges for ensuring the responsible use of "big brain data"—the recording, collection and analysis of individuals' brain data on a large scale with clinical and consumer-directed neurotechnological devices. First, I highlight the benefits of big data and machine learning analytics in neuroscience for basic and translational research. Then, I describe some of the technological, social and psychological barriers for securing brain data from unwarranted access. In this context, I then examine ways in which safeguards at the hardware and software level, as well as increasing "data literacy" in society, may enhance the security of neurotechnological devices and protect the privacy of personal brain data. Regarding ethical and legal ramifications of big brain data, I first discuss effects on the autonomy, the sense of agency and authenticity, as well as the self that may result from the interaction between users and intelligent, particularly closed-loop, neurotechnological devices. I then discuss the impact of the "datafication" in basic and clinical neuroscience research on the just distribution of resources and access to these transformative technologies. In the legal realm, I examine possible legal consequences that arises from the increasing abilities to decode brain states and their corresponding subjective phenomenological experiences on the hitherto inaccessible privacy of these information. Finally, I discuss the implications of big brain data for national and international regulatory policies and models of good data governance.

Keywords Brain data · Neurotechnology · Big data · Privacy · Security · Machine learning

P. Kellmeyer (✉)
Department of Neurosurgery, Epilepsy Center, University of Freiburg – Medical Center, Engelbergerstr. 21, D-79106 Freiburg im Breisgau, Germany
e-mail: philipp.kellmeyer@uniklinik-freiburg.de

P. Kellmeyer
Cluster of Excellence BrainLinks-BrainTools, University of Freiburg, Freiburg im Breisgau, Germany

P. Kellmeyer
Institute for Biomedical Ethics and the History of Medicine, University of Zurich, Zurich, Switzerland

Introduction

We currently witness converging technological macrotrends—big data, advanced machine learning, and consumer-directed neurotechnological devices—that will likely lead to the collection, storage, and analysis of personal brain data on a large scale. In basic and applied neuroscience, this impending age of "Big Brain Data" may lead to important breakthroughs, particularly for our understanding of the brain's structure and function, for identifying new biomarkers of brain pathology, as well as for improving the performance of neurotechnological devices (such as brain-computer interfaces, BCIs). But the same technology, when applied in consumer-directed neurotechnological devices, whether for entertainment, the interactive use of web services, or other purposes, may lead to the uncontrolled

collection and commodification of neural data that may put vulnerable individuals at risk with respect to the privacy of their brain states.

Big data refers to collecting and storing vast amounts of data, for example from wearable devices (e.g. "fitness trackers"), electronic health records, or our online footprint from using web-based software services. This growing mountain of data, however, would not be of much use was it not for the advanced machine learning algorithms, specifically artificial neural networks (ANN) for "deep learning" and related methods, that are now available for analyzing this data. Most of the personal information that web-based software companies gather today is based on our voluntarily submitting our data—mostly by yielding to convoluted and mostly inscrutable "end-user license agreements" (EULAs). What we not yet have on a large-scale, but what many device and software companies are now actively developing, are consumer-directed wearable devices for recording and uploading our brain activity, mostly based on electroencephalography (EEG) [1, 2]. In combination with other wearable sensors for tracking biometric data, these devices will provide particularly rich multivariate data troves for the "personal sensing" of an individual "physiome", for the (online) decoding of person's (neuro)physiological state and behavior [3], and for making predictions on future states or behavior, an application that is studied particularly intensively in the area of mental health [4–7]. Meanwhile, companies are using powerful algorithms for "deep learning" to create facts on the ground[1] and invest heavily in leveraging these methods for consumer and health-care applications, especially in basic and clinical neuroscience [9, 10].

This "datafication" [11] across all areas of research and technological development—in which data not only refers to but enacts and guides social life [12]—puts established modes for normative reflection, deliberative value formation, as well as legislative and policy responses under pressure. At the same time, finding sustainable political and legislative responses to this transformation and hedge the relentless stream of highly personalized data against misuse and exploitation is becoming more and more difficult. For one, in order to be able to understand the benefits and risks and to then formulate an adequate regulatory response, lawmakers and politicians (as well as the general public that elects these officials) need to have at least a basic understanding of the complexity of the technologies involved. This is important for governments (or supranational bodies) in order not to succumb to indiscriminate techno-alarmism and pass "laws of fear" [13] that stifle important scientific and technological progress on the one hand, while at the same time not to display a blind techno-enthusiasm that ignores or downplays important risks.

Preferably, a democratic society should provide the necessary space and time for an inclusive and participatory bottom-up deliberative process that involves all stakeholders in the debate on how to regulate and govern the use of personal brain data. In the spirit of such a "reflexive modernization" [14]—that is not taming (human) nature with technology (a defining feature of industrialization and the modern era), but shaping technology through user-centered and value-based design—I will discuss some important ethical and legal ramifications of this profound technological transformation.

Specifically, the aim and scope of this paper is to give a comprehensive overview of (a) big data and machine learning as the driving technologies behind the 'dataification' in basic and clinical neuroscience and consumer-directed neurotechnology, (b) some pertinent ethical, legal, social and political challenges that arise from the collection, storage, and analysis of large amounts of brain data from clinical and consumer-directed neurotechnological devices.

Of the many threads and challenges that this emerging techno-social constellation offers, I will focus here on the normative implications, both from an ethical and legal perspective, of big brain data. To this end, I will examine ethical and legal implications in areas in which I believe emerging big data / machine learning applications will have a particularly profound influence. The selection of topics—such as the privacy of brain data, or the problem of bias in machine learning—is therefore motivated mostly by the likely impact of the technological transformation rather than inherent commonalities between these areas of concern (e.g. in terms of ethical theory or political philosophy).

Potential Benefits and Risks of Big Data Analytics in Basic and Clinical Neuroscience

A Brief Introduction to Big Data and Advanced Machine Learning

Before detailing the current use of big data and machine learning in basic and clinical neuroscience, let me first

[1] Such as DeepMind's Go-playing program AlphaGo Zero recently beating it's previous version AlphaGo by 100–0 games [8].

provide some brief definitions of recurring concepts and techniques from computer science:

Artificial intelligence (AI) is a term in computer science and robotics that refers to an embodied (machine/robot) or non-embodied (software program) system that can reason, learn, and plan, and which exhibits behavior which we associate with biological intelligent systems (such as humans) [15].

Big data refers to the collection and/or systematic storage of large amounts of (labeled or unlabeled) data for the purpose of finding hitherto unknown patterns, relationships or other informative features by computational analysis, often involving advanced machine learning algorithms.

Machine learning refers to a programming approach in computer science in which the behavior of a program is not fully determined by the code but can adapt its behavior (i.e. learn) based on the input data ("Learning without being programmed"). The first such program was designed to play the game of checkers (1959) foreshadowing the whirlwind successes of recent deep learning networks in beating humans in games.

Deep learning is a particular variant of machine learning which is often modelled on artificial neural networks (ANN). A typical ANN architecture consists of interconnected nodes – representing artificial neurons – with an input layer, hidden layers and an output layer. In the hidden layers, the data from the input layer undergo linear or nonlinear transformations multiple times (hence "deep"). The power of the ANN for solving data-driven tasks like pattern recognition lies in their ability for reinforcement learning at different levels of abstraction through recurrent modelling. Specific variants of such deep learning architectures, for example convolutional neural networks (ConvNet), have recently been particularly successful in applied machine learning across many research fields (such as neuroscience [16, 17]) and industrial sectors. Historically, many machine learning algorithms were developed to address pattern recognition and classification problems in computer vision and speech recognition. Therefore, detecting features and classes in a large amount of images is still one of the most widely used applications.

Factors Determining the Scope and Limits of Machine Learning Approaches to Data Analysis

Across all of the many different problems, or "use-cases" for which advanced machine learning methods are now employed, we find some commonalities that define the power and limits of these methods:

- Deep learning works particularly well in data-rich (*big data*) environments for recognizing patterns and generating predictions, tasks that are generally difficult, very time-consuming or even impossible for humans. Imagine you were asked to differentiate between thousands of animal species by looking at millions of animal images in a short time or learn to play world-class Go by mining databases with millions of recorded games and moves.
- *Scalability*, the ability to apply algorithms to very large amounts of data while retaining reasonable computation times and storage requirements, is another important feature of recent advances in applied machine learning. In data-rich environments, scalable machine learning algorithms become ever more accurate and more usable with the increasing data size.[2] In spite of these impressive achievements, there are still significant challenges and limits for advanced machine learning:

As we have discussed, advanced machine learning is particularly powerful for analyzing large amounts of data. Consequently, in all scenarios in which only few data are available, these methods are substantially less effective. For clinical applications, for instance, rare diseases or rare genotypes would be examples of such *data paucity*.

One important challenge, therefore, is to devise the right computational model for addressing any particular use-case given the data at hand. This manual tinkering, including also the labeling and/or annotation of data for learning and so-called hyperparameter setting, takes a lot of human resources, knowledge and time and is also error-prone.

The effectiveness of machine learning for data analysis and classification also relies on finding the optimal learning scenario for any given problem. For clinical use-cases, the most effective applications of advanced machine learning have so far relied on a so-called supervised (or semi-supervised) learning scenario and clinical questions related to digitized images. In *supervised learning*, an algorithm trains with labeled data, for example magnetic resonance images (MRI) of the brain

[2] For example, while a couple of years ago the program *Google Translate* often provided clunky, slightly "off" translations, the current version (as well as similar programs such as the *DeepL Translator*) can automatically recognize many input language and provide usable translations in over 100 languages.

that have been labeled as either normal or abnormal by a radiologist. After learning, the algorithm then analyzes a new data set and can identify abnormal images with high precision.[3]

Finally, most current machine learning programs still have difficulties with *transfer learning*, applying knowledge extracted from one set of problems to a new challenge [19]. An algorithm that is effective for classifying brain images may not perform particularly well on other types of data.

Benefits in Using Big Data and Advanced Machine Learning in Basic and Clinical Neuroscience

For an informed risk-benefit-analysis, it is important—in my view—to appreciate the actual and potential benefits for patients that big data and advanced machine learning may offer in basic and clinical neuroscience.[4]

While my observations here focus on the area of neuroscience, we should acknowledge that advanced machine learning has revolutionized basic and clinical research across all areas in biomedicine and turbocharged the emerging field of "precision medicine" [20]. To provide just a few recent examples: such algorithms have been shown to achieve dermatologist-level accuracy in classifying skin lesions as cancerous [21], to be able to predict the outcome of antiepileptic drug treatment [22] or to predict the prognosis of small-cell lung cancer from images of pathological tissue samples [23].

For basic and translational neuroscience these methods, particularly deep learning, yield important advances too. Researchers from the University of Freiburg (Germany), for example, used a convolutional neural network (ConvNet) in 2017 for deep learning to decode movement-related information from EEG data [16] and for operating an autonomous robot [24]. ConvNets for deep learning were also successfully used to predict signal processing in primate visual cortex [25] and to predict human brain responses (from functional magnetic resonance imaging, fMRI) to processing natural images [26].

In clinical neuroscience and translational neurotechnology, we see similar advances in leveraging advanced machine learning for diagnostic classification and for predicting disease outcomes or therapeutic responses. The computing revolution mentioned above results in a surge of digital health-related data, from single data points (e.g. lab parameters), to continuous data from monitoring devices over days or weeks (such as continuous ECG monitoring from intensive care or EEG in epilepsy centers) to complete electronic health records (eHR), which can be used for data and text mining [27]. It is therefore not difficult to imagine how these streams of data may inform advanced computational analyses and even outperform human diagnosticians in many medical disciplines.

Most recent efforts in leveraging advanced machine learning in clinical neuroscience have been particularly fruitful in neuroimaging-based diagnosis (and/or prediction) in neurology and psychiatry. In neurology, such approaches have already been able to: detect morphological brain changes typical of Alzheimer's disease from neuroimaging [28, 29], to predict brain tumor response to chemotherapy from brain images [30], or distinguish typical from atypical Parkinson's syndromes [31]. In psychiatric research, examples for leveraging machine learning are the prediction of outcomes in psychosis [32], the persistence and severity of depressive symptoms [33], and the prediction of suicidal behavior [7].

Risks of Big Data Analytics and Advanced Machine Learning in Basic and Clinical Neuroscience Research

While acknowledging the many actual and potential benefits of big data analytics with advanced machine learning, it is equally important to discuss some inherent risks of this approach. The sheer scale of the technological transformation discussed here across so many sectors of society naturally invites scrutiny and caution in risk assessment. With an eye on the scope of the paper, however, I will focus on identifiable and concrete risks to individuals,

[3] *Unsupervised learning*, in contrast, does not use labels. The algorithm therefore gets no input on what the "right" pattern of data is, but explores the data and finds recurring patterns and structures by itself, for example finding clusters of "similar" brain images. In situations in which labeling data is unfeasibly time consuming or expensive, as in labeling all CCTV images in London from a given day with whether they contain faces or not, *semi-supervised* algorithms can use both labeled data (images with and without faces) and unsupervised data (the remaining images) for analysis. An intermediate approach, using both supervised and semi-supervised machine learning, was recently used to create a spectacular new brain map in which the semi-supervised algorithm identified ninety-seven new anatomically distinct areas [18]. Ideally, a machine learning program would recognize and decide the optimal model parameters itself and unsupervised learning with convolutional networks for deep learning is an interesting frontier in that respect.

[4] As basic neuroscience, I refer here to research on basic mechanisms and functions in the central (and peripheral) nervous system, from the (sub)cellular level to large-scale brain networks, whereas clinical neuroscience refers to all research related to pathological changes in the nervous system and therapeutic interventions (particularly translational neurotechnology).

particularly patients and research subjects (rather than transformative effects on society as a whole).

For the time being, I see no immediate risks for the momentary well-being of research subjects *during* experiments in which big data and advanced machine learning are later used for "offline" analysis of brain data. In such typical research scenarios, data are collected from many individuals, collated on local servers or in cloud-based data repositories, and then analyzed, for example with advanced machine learning algorithms. The collection and storage of neural (and other) personal data does, however, carry certain risks with respect to data privacy which I will discuss next. In subsequent sections, I will then examine, how real-time interaction between users and a neurotechnological device, particularly in closed-loop systems, may affect the autonomy, sense of agency and other aspects of a user's experience.

Some Ethical and Legal Implications of Big Brain Data

The development described above will quite likely have a transformative effect on research practices in neuroscience as well as clinical neurology and psychiatry. Likewise, consumer-directed neurotechnological devices will create new ways in which users may interact with systems for entertainment, personal computing and mobile devices. Of the many ethical and legal challenges that emerge from this techno-social constellation, I will limit my analysis here to a few issues that I find particularly pressing—fully acknowledging that this selection of topics is neither particularly comprehensive nor representing anything other than my current personal interests.

On the Security of Neurotechnological Devices and the Privacy of Brain Data

First of all, storing an individual's brain data on local or web-based servers / repositories makes these data vulnerable to unintentional data exposure, intentional data leaks and (cyber) attacks ("hacking"). Furthermore, cross-referencing biometric data with other types of data may allow for the de-anonymization[5] of personalized data—i.e. exposing the identity of research subjects or patients. This de-anonymization may then leave individuals vulnerable to identity theft or other criminal acts by third persons (e.g. holding a person ransom by threatening to release potentially damaging information, such as on brain pathology) [34].

While this is a general problem when data records from research participants (or patients) are stored electronically, the highly personalized nature of brain data (see e.g. the possibility for "brain fingerprinting" [35])—much like genomic data—may increase the identifiability of individuals.

Case Example: Deep Learning for Brain-Computer Interfacing in a Severely Paralyzed Patient

For the individual user of a clinical or consumer-directed devices that processes large amounts of neural data, we may discern the following scenario as an example of the importance of privacy of brain data.

In the case of a patient with locked-in syndrome—that is severe paralysis through extensive damage to the brainstem—who uses a spelling system operated by a brain-computer interface that uses deep learning for analyzing her neural data, the following concern might apply:

You may imagine, that if a BCI spelling system was used consistently for some time, it is quite likely that the BCI user conversed with different people at different times; with relatives, nurses, doctors, friends, visitors and others. These different conversations will have been different in topic and their level of intimacy. Perhaps the patient would or would not mind if a mundane conversation with her nurse to adjust the bed was to be read by another person, but she might very well object if an intimate discussion with her husband, for example about her fear of death, would be read by anyone else. Furthermore, as we have discussed above, combining the continuous neural recordings with the spelling content provides a powerful source for unmasking the user's identity.

Therefore, limiting and securing access to the patient's data is an important prerequisite for preserving privacy. For the log files, we might ask, for example, whether they should be preserved at all, deleted after a pre-specified time, or remain fleeting—like our spoken conversations usually are? Or should they be recorded permanently but only accessible to the BCI user via a password? What happens when these records may become relevant in a legal context? Imagine that the locked-in patient may no longer be able to use the BCI and a medical emergency occurs, for example a life-

[5] Or rather „de-pseudonymization", given that a lot of personal data from participants is pseudonymized rather than anonymized (which is often not feasible in studies).

threatening pneumonia requiring artificial ventilation in an induced coma. Now, the husband of the patient—having become the legal representative via an advanced directive—wishes no further treatment claiming this to represent the wish of the patient. The doctor however remembers conversing with the patient a couple of weeks before, when she was still able to use the BCI, where she told him to treat any medical emergencies exhaustively. If the case is brought before a judge, will he have the right to subpoena the BCI spelling log files?

Neurohacking and the Emergence of "Neurocrimes"

Another important threat to data privacy and security for the individual patient / user is posed by "neurohacking".[6] If the BCI system, in the case presented here, was connected to a web-based cloud server for storing and analyzing the brain recordings, the data could get exposed either intentionally (e.g. a rogue employee of the server company who sells the data), released accidentally or be accessed and/or stolen via hacking into the server.

The feasibility of hacking such active medical devices has already been demonstrated for implantable cardioverter-defibrillator (ICD) systems [36–38]—Halperin et al. [38] demonstrate how to use equipment from a general electronics store to remotely hack into a wireless ICD—and it seems likely that BCI systems could be equally vulnerable to electronic attacks.

Such unwarranted access to one's neural recordings and other types of personalized information (e.g. the spelling logs of the BCI system) would be a valuable data trove for persons with malicious intent. For example, a hacker could use the highly personalized information for holding a person at ransom (threatening release of the personal information) or could disable the BCI and demand a ransom for unlocking the device and/ or its operating software ("ransomware"). In a BCI system that is used for controlling a robotic prosthesis, a hacker could similarly take control of the prosthesis and threaten or cause harm to the user or other persons. Therefore, the safeguarding of these highly personalized biometric (and other) data and ensuring device and system security is an important area of concern and merits an in-depth examination from a legal, forensic and technological perspective to prevent such potential "neurocrimes".

Technological Barriers and Opportunities for Safeguarding Neurotechnological Devices and Brain Data

Given this importance of safeguarding neurotechnological devices as well as servers and software programs for processing brain data, let us first briefly look at the main technological barriers.

First, the collection, aggregation and (real-time) analysis of large amounts of data requires massive storage and processing units, which today are most often provided by services for server-based cloud computing.[7] While personal devices can be secured quite effectively against unwarranted access,[8] cloud-based software repositories are much more difficult to secure. Many of the technology giants that are moving into the consumer neurotechnology market—such as Facebook and Google—have traditionally been software rather than hardware-based enterprises [40]. Therefore, it remains to be seen whether these companies can develop strong safeguards at the hardware level to secure such devices from unwarranted access.

At the software level, Alphabet and other companies are very active in developing new paradigms for securing personal data. One particularly ingenious idea is the concept of federated learning. In federated learning, the algorithm for machine learning with a person's neural data would operate locally on the neurotechnological device and only share certain, non-personalized, inferences on the data with a central server for further data processing [41]. Such a local encapsulation, when coupled with strong device-level hardware and software security and strong encryption of the transferred data, could make such a system much less vulnerable to device hacking and cyber-attacks. Similarly, other technologies, such as blockchain and "differential privacy"[9] [42] could be used for the granular auditing and tracking of brain data.

[6] I use "neurohacking" here in the sense of gaining illicit access to a neurotechnological device or a software program that processes neural data. In the literature and the media, neurohacking is sometimes discussed in the sense of "hacking your brain", i.e. referring to neuroenhancement.

[7] The biggest three cloud computing services are currently provided by Amazon Web Services, Microsoft and Google; cp. https://due.com/blog/biggest-cloud-computing-companies-earth/, accessed January 10th 2018

[8] Exemplified by the fact that hacking the Apple smartphone of the perpetrators of the San Bernardino mass shooting in 2016 cost the FBI $1.3 m [39]

[9] See also a current whitepaper by Nissim et al. (2017) on the subject: https://privacytools.seas.harvard.edu/files/privacytools/files/nissim_et_al_-_differential_privacy_primer_for_non-technical_audiences_1.pdf; accessed Jan. 10th 2018

Some observers have noted, however, that it might not be ideal if users will have no other choice in the future, than to leave both their data *and* the responsibility for data safety in the hands of one company [43]. An alternative could be the creation of so-called data banks, companies that specialize in data security and act as intermediaries for using brain data for research, clinical, or consumer purposes. While this idea merits some further investigation, in my opinion, I would, as others have, worry that such a system could also abet the process of privatization and commodification of health and biometric data [44].

With respect to legislative and regulatory implications, I would argue that legislators should therefore mandate strong security requirements, such as device-level encryption and hardware protection, end-to-end encryption for data transfer, as well as methods for auditing data trails for clinical and consumer-directed neurotechnological devices.

Privacy of Personal Brain Data: Psychological and Social Barriers and Opportunities

In addition to the technical challenges, there are also important psychological and social barriers for safeguarding brain (and other personal) data from unwarranted access.

From the formative, comparatively open years of the early internet to the cordoned off web dominated by oligopoly and "data capitalism" of today, user attitudes towards data privacy as well as the political and legal frameworks for the security of personal data have changed substantially. The current fabric of the web is characterized by a triad of corporatization, commercialization and monopolization for providing content and services in which personal data has become the most important commodity. Surveys in the U.S. suggest that concerns about this commodification of personal data and repeated instances of massive data leaks influence the privacy concerns of internet users. While the main topic of internet users' concern—the disclosure and trading of personally identifiable information (PII)—has not changed over time (according to a study comparing 2002 and 2008 [45]), the level of concern indeed has risen substantially in this period.

One would think that this gradual "Snowdenization", the rising awareness and concerns in society about the mass collection, dissemination and misuse of PII, would perhaps have created a fertile ground for a level-headed and evidence-based debate about the future handling of personal brain data. Yet, at the same time, don't we often wonder why users of online software services seem, on average, to care little about their personal data trails? From an individual psychology point of view, it seems that on social media the actual (or perceived) psychological rewards for using the services often outweigh the possible threats to privacy for the users [46].

Furthermore, the near ubiquitous use of social media for communication may impel vulnerable individuals, such as teenagers or individuals with psychiatric disorders (e.g. social anxiety or depression), to use these services to avoid social exclusion or ostracism and thereby compromise on possible privacy concerns.

Another psychological barrier could be that still today many services have a default opt-in (rather than opt-out) policy concerning the use of personal data. Furthermore, even if there is a default opt-in environment, the EULAs of the web services are often difficult to understand and navigate [47]. Moreover, opting out of data sharing with the service providers may also worsen the usability and consumer experience (or even prevent the usage of these services).

Counteracting these social and psychological pressures would require the restructuring of many basic design and programming features of device- and web-based software services. For a start, we may need to consider to move from an opt-out to an opt-in environment in any context in which sensitive personal information, particularly biodata (and especially brain data), are being transferred. Furthermore, companies could be incentivized (if not obliged) to improve the EULAs of such services, allowing for a granular and transparent consent process for the users. Moreover, in order to move from opaque, black box neurotechnology to transparent systems, users and patients should have the right to know whenever they interact with an intelligent system, who trained the system and what data was used for training.

Transparent EULAs in and of themselves are a necessary but not sufficient step towards improving the consent process for ceding personal brain data, however. This needs to be complemented, in my view, by increasing the average level of basic understanding of the capabilities and limitations of big data and advanced machine learning—"data literacy"— in society. Many educational initiatives in different countries already work toward this goal but I would submit here that the "long view" in shaping future educational policies should include brain data as an emerging (and perhaps special) class of bio(-medical) data (and commodity) [48].

Further questions such as whether and to what degree this juridification of the processing of brain data should remain solely in the hands of national governments and/or should also be codified in international treaties and international public law, as well as whether brain data are a different kind of biometric data that may require special "neurorights" [49] exceed the scope of my discussion here, but are certainly important issues for further (comparative) legal scholarship.

Decision-Making and Accountability in Intelligent Closed-Loop Neurotechnological Devices

When humans and intelligent medical devices work in concert—take a closed-loop brain-computer interface that uses deep learning for decoding a user's EEG data—mutual adaptivity may greatly increase the effectiveness of the intended use, for example by increasing the decoding performance over time. Advanced machine learning, moreover, is also a powerful method for the analysis of brain data, such as EEG or fMRI, in real time ("online"), for example to control of a robot with brain activity [24]. As such closed-loop interaction unfolds in real-time, there is the risk that the output of highly adaptive algorithms for deep learning—which are by their nature evolving and thus unpredictable—may harm participants or patients. In cases in which such intelligent closed-loop devices do not only decode neural data for specific purposes but may also actively interfere with brain states, for example by delivering electric stimulation to the cortex, and the decision if and with what intensity was determined solely by the device, the system gains decision-making capacity. Elsewhere, together with colleagues, I have discussed the problem of an "accountability gap" that may arise in such cases in which an (semi)autonomous intelligent system is granted decision-making capacity based on an evolving and adaptive algorithm [50]. In this paper, we have argued that the regulatory process for approving such closed-loop neurotechnological devices should take these possible effects into account and that further research is absolutely crucial in studying these effects in such devices—whether for medical or non-medical use.

I would add here the importance of developing guidelines and models for promoting a design and development process for neurotechnological devices that is centered on the needs, capabilities and preferences of the intended end-users. To date, most such devices as well as complementary assistive technology, such as robotic systems that may be controlled by a BCI, undergo a top-down design and development process with little input from the end-user perspective.

Possible Neurophenomenological Effects of Closed-Loop Neurotechnological Devices

While it is one thing to employ such closed-loop interaction to optimize the performance of a medical device, for example for regulating seizures in patients with otherwise treatment-resistant epilepsy, using the same capabilities in consumer-directed neurotechnological devices may result in many unintended adverse effects. Especially closed-loop interaction—i.e. changing the parameters of a device based on the real-time sensing of neurophysiological data—may adversely affect the phenomenological experience of individuals, for example by altering the sense of agency, a subject's sense of authenticity and autonomy, or the self [50–52].

Take the simple, non-brain related, example of the now familiar algorithms for "optimizing" web searches or for making recommendations for buying items in online shops. If the algorithm recommends a certain item based on my previous purchases (and other users' purchases), to what degree does a purchase based on this recommendation reflect a choice based on my preferences (momentary or long-term), and to what degree is the decision shaped by the algorithm?

Similarly, we may examine such effects in (thus far hypothetical) closed-loop consumer-directed neurotechnological devices. Imagine, if you will, a wearable EEG system that is connected to your PC and web-based cloud server, that continuously analyzes your neural data with deep learning and modifies the content of your social media feed (or other software) services adaptively based on this analysis. If the underlying choices (and biases) for classifying your neural data in certain ways that the system makes to "optimize" your user experience remain unknown or opaque, to what degree can you trust the system that the modification of what you experience by the system reflects your true preferences rather than inherent biases in the algorithm (or the data it uses for learning)? If learning occurs not only with a single user's data but across many users in the cloud, how personalized are the choices the system recommends (or makes) really? Would the co-evolving adaptivity between user and algorithm over time blur the line between the users original and/or genuine mental landscape (preferences, attitudes, opinions, desires and so forth) perhaps even her cognitive abilities on

the one side, and the system's biased inferences on the other side?

One might ask,[10] of course, whether and how this 'co-adaptation' between a human and an intelligent (closed-loop) system substantially differs from the 'standard' adaptation of our preferences, attitudes and behavior as we engage with the world. Of course, we also see adaptation between humans and other (data-driven) systems, for example conventional advertising or standard treatment models for common diseases. I would submit, however, that there are indeed substantial, i.e. non-trivial, differences between such established modes of interaction and the co-adaptation between humans and a (closed-loop) system based on big data analytics and advanced machine learning, particularly: (a) The aim of traditional models of data-driven analytics is to derive common parameters / features from data from many individuals to build system that responds well for the average user; in the case of advertisement, for example, a product would be tested on many individuals and then modified according to the average preferences of consumers. In emerging systems based on big data / machine learning, the aim is often to achieve a maximally individualized response based on pattern classification and predictive analysis; again in the case of advertisement, to develop highly personalized 'targeted' advertising that can adapt to a consumers changing preferences. For analytics based on brain data, for example, a brain-computer interface for paralyzed patients, such a system—because it continuously analyzes brain responses and can thus adapt to neurophysiological changes—would be much more adaptable to (e.g. disease-related) changes to brain signals over time. (b) In the case of a closed-loop system involving brain data, this co-adaptation is taken even further: A brain-computer interface, for example, based on measurements of bioelectric brain activity from an implanted electrode on the brain surface (which has the capability for delivering electrical stimulation to the cortex), could modify brain activity in real-time through delivering electrical stimulation based on the measured brain activity—a capability that would elude traditional open-loop BCI systems. (c) More mundanely perhaps, the now ubiquitous algorithms that provide recommendations for further purchases in online stores (either based on individual data: "Because you bought item X, you might also like item Y."; or based on data analysis over many individuals: "People who bought item X also bought item Y.") can produce eerie effects on our sense of autonomy (and directionality) and authenticity in making choices / decisions: Would I like item Y equally if it had not been flagged by the algorithm? How can I stay open-minded and/or change my preferences if chance encounters with unusual items are more or less eliminated by the 'filter bubble' of the algorithmic shopping assistant? Such unease from a close co-adaptation between an algorithm and human interactors, of course, now emerges in many other domains, such as information ('fake news') or political opinion formation.

Such continued interactions between a user and an 'intelligent' neurotechnological device may thus have a profound and potentially transformative effect on the experience of authenticity, the sense of agency, the active self and other aspects [53]. Therefore, I would recommend to make the study of these "neurophenomenological" effects an integral part of user-centered research on and the development of neurotechnological devices, particularly devices for closed-loop interaction.

The Problem of Bias in Applications Based on Big Data and Machine Learning

As mentioned in the previous section, the influence of bias on machine learning and (closed-loop) neurotechnology may be substantial. Bias in (data) science denotes systematic skews in the way data is collected (e.g. 'selection bias', in which particular sources of data are systematically, though mostly unconsciously, ignored), annotated, categorized and so forth. Importantly, however, bias also (and to particularly deleterious effect) operates at the level of human cognition. Convergent research in behavioral psychology and cognitive science has revealed the important and universal influence of cognitive biases on human decision-making and choice-taking. Cognitive biases are mental shortcuts (or heuristics) that all humans are inclined to take in evaluating problems or making decisions. The *availability heuristic*, for example, refers to the overreliance on readily available information and the *recency bias* describes the reliance on often highly memorable, recent events or information when making decisions[11] [54]. Such human cognitive biases are particularly

[10] I would like to thank an anonymous reviewer for raising this point.

[11] When making travel plans, for example, we are more likely to take the car than flying if a major plane crash occurred a few days before, despite the irrefutable fact that the probability of being harmed in a car accident – independent of any particular plane crash having taken place – is, was and will be much higher than being harmed on any particular flight.

problematic for techniques, such as machine learning, that rely on large amounts of data that are annotated and categorized by humans. In other words, bias is a ubiquitous and almost inescapable phenomenon which may skew the basis for learning (and subsequent 'behavior') of devices based on big data and machine learning at many levels: at the level of the data collection, annotation and categorization; through biases of the programmers; or the biases of the users of such a system [55]. In the case of AI-based decision support systems for clinicians—a system that analyze a patient's data and may give advice for further tests or recommend treatments—biases in the training data for the underlying artificial neural networks may lead to skewed decision-making [56].

If an ANN for skin cancer detection, for example, was mostly trained on images from light-skinned individuals, it might perform better in screening light-skinned than dark-skinned individuals, which would effectively introduce an ethnic bias into the diagnostic procedure [57].

Distributive Justice: Impact on Research in Data-Poor Areas

As we have discussed, big data and machine learning are particularly effective methods for analyses that are easy for machines but difficult for humans. As researchers increasingly leverage the power of this approach for a variety of problems, we might see whole research programs in neuroscience move into data-rich environments, simply in order to be able to employ these methods to their maximal potential (much in the way that gene editing based on CRISPR/Cas9 is currently transforming research in cell biology and molecular biomedicine).

For basic and clinical research on issues that are not blessed with plentiful data, for example on rare (so called "orphan") diseases or research in countries with no significant digital infrastructure and a lack of economic resources, these new methods will not be so readily applicable and/or available. It remains to be seen, whether the promise and success of the "datafication" of medicine will drain human and financial resources from such areas or whether the macro-level research policies, at the international, governmental and funding agency level, will find ways for commensurate funding schemes to allow different approaches to flourish.

Mens Mea: On the Legal Status of Brain States

Given the technological development outlined above, it seems realistic to anticipate an increase in the ability to correlate particular brain states more and more reliably with concurrent "mental states"[12] through advanced machine learning on brain data. This, in turn, could breach the hitherto closed off sanctum of one's thoughts and feelings—particularly those mental states (often denoted as phenomenological consciousness) that are not accompanied by overt behavior or peripheral physiological state changes and were thus far unobservable. This scenario raises important questions regarding the privileged privacy of one's mental states, the right to not disclose one's thoughts or feelings, especially (but not only) in a legal context.

The main question in this regard may well be, whether brain states and inferences on those brain states re their corresponding mental states (through decoding) existing legal concepts and instruments are sufficient to govern the fair use of these data in the courtroom (as well as in the context of policing or criminal investigations). On the one hand, we might ask again whether individuals should have the (inalienable?) right to not have their mental states decoded? If so, would this amount to a (new) fundamental human right or would existing legal frameworks be sufficient to deal with this question (see [49] for an excellent discussion). On the other hand, what if such methods could also be used for exculpatory purposes in favor of the defendant? Today, neuroimaging, particularly for demonstrating structural anomalies in a defendant's brain, for example, is overwhelmingly used for establishing brain damage as a mitigating factor in criminal cases (see [58] for a recent overview). Should a defendant therefore not have the possibility (if not right) to use decoding methods based on advanced machine learning for establishing mitigating factors? To this end, it should also be discussed whether the existing (in the US legal system anyway) so-called Daubert standard for scientific admissibility is applicable to decoding brain states with the deep learning, given the concerns about the "black box" characteristics of many such machine-learning-based decoding architectures [59].

[12] I use "mental states" here as a token for denoting first-person phenomenological experiences, thoughts and feelings—without "cryptodualistic" intentions.

Jurisprudence and legal philosophy has long known the concept of *mens rea*—the guilty mind—for determining a defendant's responsibility (and thus culpability) for his or her actions. Perhaps it is now the time to intensify the discussion of *mens mea*—the concept of one's mind as a protected sanctum of thoughts and feelings—with respect to the legal status of brain data and mental states. In terms of civil liberties, we encounter two main scholarly debates on the freedom of our mental states and capacities: (1) The *mens mea* question mentioned above (often framed in terms of 'freedom of thought'), i.e. the freedom *from* unwanted interference with one's mental states and/or cognitive capacities by others (i.e. 'negative liberty') [60–65]; and (2) the positive freedom *to* (for some involving the fundamental right to) maximize / fully realize one's cognitive capacities, involving the right to employ methods for cognitive / neural enhancement (also referred to as 'cognitive liberty') [66–70].

For comparative purposes, it might be interesting to look at the ongoing debate in forensic science and criminal law around the acceptability—both in terms of scientific standards and from a normative point of view—of using DNA analysis for identifying phenotypical traits (e.g. eye or skin color) for identifying suspects and as evidence in the courtroom [71].

Some Thoughts on Regulating and Governing Big Brain Data

In a recent policy paper [57] together with colleagues, we have pointed out the importance of updating existing and/or developing new guidelines for research and development of clinical and consumer-directed neurotechnology that acknowledges the challenges outlined above.

To this, I would like to add the following thoughts. First, I would like to voice concern regarding the usurpation of the ethical and legal discourse by the private sector around the question of brain data privacy and the safety of AI [72]. While the participation and active engagement, preferably beyond the minimal standards of "business ethics", of the industries that actively shape the development of this technology is highly commendable and important (in the spirit of an inclusive deliberation process), I think it is important to closely monitor the ways in which these companies may dominate these discourses by spending a lot of resources on the issue. If for, for example, the Ethics Board of such companies as DeepMind remains shrouded in intransparancy with respect to their personnel and mission, it is difficult to see the *raison d'être* of such entities [72]. Arguably, there is a discernable difference between running a corporate policy of honest and transparent participation in public discourse on the one side, and engaging in lobbyism and opinion-mongering on the other, and I hope the companies in question will adhere to the former, rather than the latter, form of corporate social responsibility. To this end, the citizenry should actively participate in the public discussion on neurotechnology and AI and engage the companies in critical discourse on their corporate strategies and policies.

Another important and largely unresolved question concerns the adequate classification of different types and sources of highly personal data, particularly biomedical data, with respect to the appropriate (and proportional) legal and regulatory frameworks. For example, most would agree that results from blood tests or data from wearable fitness trackers constitute biomedical data. What about movement data from a person's phone GPS sensors or person's text (or image or voice) entries in her social media account, however? In recent studies, researchers were able to infer suicidality from automated, machine-learning-based analyses of electronic health records [7] and even from user entries in Facebook [73].

Should the casual texts we disseminate via social media thus be considered as biomedical data if they turn out to be highly valuable for AI-based predictive analyses with implications for a person's well-being? These questions do point to the fact that—perhaps counterintuitively— there is no generally accepted definition, let alone granular classification, of biomedical data as a particular class of data.

In the absence of such a generally accepted, the question whether brain data should be treated just like any other type of data (and it's 'value' solely determined by economic or other parameters), or whether it should be considered to be a special class of data must remain unresolved for the time being.

Given the breathtaking speed with which new methods and devices for gathering massive amounts of highly personalized data enter our lives, however, I would suggest that it is important to develop a comprehensive classification system that precisely defines biomedical data. Better yet, this system should also enable an evidence-based risk stratification (e.g. in terms of the potential for misuse and other risks for the individual from which the data was collected).

Any coordinated effort of classifying biomedical data, of course, will not occur in a normative vacuum but

will be motivated by ulterior (ethical, legal and/or political) goals. For example, from a deontological perspective, the goal for such a classification could be to maximize each person's individual rights, such as civil liberties (e.g. in terms of data ownership), whereas, from a utilitarian perspective the focus could be to maximize the benefit of big data analytics for society and the average individual. In terms of ensuring a transparent and accountable process, developing a biomedical data classification should be managed by institutions that are democratically legitimated, such as commissions in (or between) democratic states or supranational institutions (e.g. the European Commission).

Furthermore, when it comes to forging an international consensus process on how to shape and regulate research and development of neurotechnologies and AI (and Big Brain Data for that matter), we should acknowledge that there are important differences, mostly for historical and systemic political reasons, in the ways in which different nations and supranational bodies address the question of technology and risk assessment.

Without being able to map the full extent of the problem here, let us briefly look at differences between the US and European approach in risk assessment and regulation: While the reality is of course much more nuanced, it seems fair to say that—from a historical perspective—the European take on risk regulation has relied more on the precautionary principle[13] than the US approach. Historically, we may understand the emergence of precaution as a regulatory strategy against the globalization of large-scale technological hazards such as nuclear proliferation, climate change, genetic engineering and the like, in the modern era. The extensive and ongoing legal and political struggle between the US and the EU on how to regulate genetically modified organisms (GMO) perhaps provides a case in point [74]. In some grand sociological theories, this "risk society" is even considered to be the constitutive condition of modernity.[14] In this discussion, the distinction between hazards and risks is important for salvaging precaution from being scrapped away as a "paralyzing" principle, as exposed incisively in Cass Sunstein's book "Laws of Fear: Beyond the Precautionary Principle" [13]. While I concur with Sunstein's main criticism that precaution in and of itself – without considering feasibility, cost-effectiveness and other contingencies – is at best ineffectual and may even prevent important progress or be harmful [76] I nevertheless think that precaution is an important mechanism to hedge rapid technological developments against unintended (and unforeseen) adverse consequences of neurotechnology and AI and merits further legal and sociological study.

[13] The "precautionary principle" denotes an approach to risk assessment and regulatory policy that is based on proceeding cautiously in allowing the use of emerging technologies—often proportional to the uncertainty regarding the likelihood of known hazards and acknowledging the possibility of unknown hazards.

[14] An influential concept in that vein is the notion—most prominent in the works of the German sociologist Ulrich Beck—as living in a "risk society" (German: *Risikogesellschaft*) as a constitutive condition of modernity [75].

Summary and Conclusions

To summarize, let me point out that my main concern with the Big Brain Data scenario sketched here is not the underlying technology—neurotechnological devices, big data and advanced machine learning—but rather the uncontrolled collection of brain data from vulnerable individuals and the unregulated commodification of such data. We have seen that the attitudes of technology users towards the privacy of PII and device security may vary substantially, from uncritical enthusiasm to broad skepticism and every stance in between. If we accept the basic premise of living in a techno-consumerist society predicated upon a growth model of (data) capitalism, we need to find some discursive space to accommodate both the "enthusiasts'" stance of cognitive liberty—the freedom to shape their selves by means of new technologies—and the "critics'" stance of acknowledging inherent risks of emerging clinical and consumer neurotechnology and proceeding with precaution in the development and application of these devices.

I would submit that those two stances or "conceptual lenses" [77] are not incommensurable in terms of how they might inform and guide our legislative, regulatory and political response (and preemptive strategies) for governing the use of brain data. Despite the differences between these stances, I hope that both sides could agree on some basic guiding principles that may hedge and facilitate this process of deliberation and, ultimately, decision-making:

(1) To maximize the knowledge on the technical and (neuro)scientific aspects as well as medical, social and psychological effects of such devices on the individual user and, at the macro level, on societal norms, legal and political processes. This entails

making qualitative, participatory and user-centered research a central and indispensable part of the design, development and application of clinical and consumer-oriented neurotechnological devices.

(2) To avoid the instrumentalization of neurotechnology, machine learning and big data (in accordance with Kant's "formula of humanity"), i.e. treating the users of such devices not merely as a means (e.g. to maximize profits through targeted advertising) but as an end, by measurably improving their social, psychological and medical well-being and thereby promoting human flourishing.

(3) To integrate the ethical, legal, philosophical and social aspects of (neuro)technological research and development, machine learning and big data into the curricula of disciplines that participate in / contribute to the development of such devices; i.e. computer science, engineering, neurobiology, neuroscience, medicine and others.

(4) To explore inclusive and participatory models that combine expert knowledge and opinions with a bottom-up process of public opinion formation to inform the political and legal deliberation and decision-making process. Such a model of *indirect normativity*, i.e. specifying processes rather than values, could perhaps enhance the acceptance and safety of these emerging technologies and also satisfy some stakeholders' need for a precautionary approach.

Finally, the broader (and, again taking the "long view", perhaps decisive) question that is highlighted by the ascent of big data and machine learning across all sectors in society is, in my view, how we as the public—a collective of responsible social and political beings[15]—can determine and shape the beneficial use of this powerful technology in society, how we can be the sculptors of this process rather than mere data sources[16] and spectators.

Acknowledgments This work was (partly) supported by the German Ministry of Education and Research (BMBF) grant 13GW0053D (MOTOR-BIC) to the University of Freiburg – Medical Center and the German Research Foundation (DFG) grant EXC 1086 BrainLinks-BrainTools to the University of Freiburg, Germany.

[15] In Aristotle's sense of a person as a ζῷον πολιτικόν (*Zoon politikon*), a social *and* political creature.

[16] In analogy to Yuval Harari's notion of humans as „biochemical algorithms"[11]

References

1. Kumari, Preeti, Lini Mathew, and Poonam Syal. 2017. Increasing trend of wearables and multimodal interface for human activity monitoring: A review. *Biosensors and Bioelectronics* 90: 298–307. https://doi.org/10.1016/j.bios.2016.12.001.
2. Piwek, Lukasz, David A. Ellis, Sally Andrews, and Adam Joinson. 2016. The rise of consumer health wearables: Promises and barriers. *PLoS Medicine* 13: e1001953. https://doi.org/10.1371/journal.pmed.1001953.
3. Price, Nathan D., Andrew T. Magis, John C. Earls, Gustavo Glusman, Roie Levy, Christopher Lausted, Daniel T. McDonald, Ulrike Kusebauch, Christopher L. Moss, Yong Zhou, Shizhen Qin, Robert L. Moritz, Kristin Brogaard, Gilbert S. Omenn, Jennifer C. Lovejoy, and Leroy Hood. 2017. A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nature Biotechnology.* 35: 747–756. https://doi.org/10.1038/nbt.3870.
4. Kreitmair, Karola V., Mildred K. Cho, and David C. Magnus. 2017. Consent and engagement, security, and authentic living using wearable and mobile health technology. *Nature Biotechnology* 35: 617–620. https://doi.org/10.1038/nbt.3887.
5. Li, Xiao, Jessilyn Dunn, Denis Salins, Gao Zhou, Wenyu Zhou, Sophia Miryam Schüssler-Fiorenza Rose, Dalia Perelman, et al. 2017. Digital health: Tracking Physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biology* 15: e2001402. https://doi.org/10.1371/journal.pbio.2001402.
6. Mohr, David C., Zhang Mi, and Stephen M. Schueller. 2017. Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology* 13: 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949.
7. Barak-Corren, Yuval, Victor M. Castro, Alison G. Solomon Javitt, Yael Dai Hoffnagle, Roy H. Perlis, Matthew K. Nock, Jordan W. Smoller, and Ben Y. Reis. 2016. Predicting suicidal behavior from longitudinal electronic health records. *American Journal of Psychiatry* 174: 154–162. https://doi.org/10.1176/appi.ajp.2016.16010077.
8. Simonite, Tom. 2017. This more powerful version of AlphaGo learns on its own. WIRED.COM, October 18.
9. Marblestone, Adam H., Greg Wayne, and Konrad P. Kording. 2016. Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience* 94. https://doi.org/10.3389/fncom.2016.00094.
10. Hern, Alex. 2017. Google's DeepMind plans bitcoin-style health record tracking for hospitals. *The Guardian*, March 9, sec. Technology.

11. Harari, Yuval Noah. 2016. Homo Deus: A brief history of tomorrow. Harvill Secker.
12. Gray, Jonathan. 2016. Datafication and democracy: Recalibrating digital information systems to address broader societal interests. *Juncture* 23: 197–201. https://doi.org/10.1111/newe.12013.
13. Sunstein, Cass R. 2005. Laws of fear: Beyond the precautionary principle. Cambridge University Press.
14. Beck, Ulrich, Anthony Giddens, and Scott Lash. 2007. *Reflexive modernization: Politics, tradition and aesthetics in the modern social order*. Cambridge, UK: Polity Press.
15. Russell, Stuart, and Peter Norvig. 2013. *Artificial Intelligence: A Modern Approach*. New international edition. Prentice Hall.
16. Schirrmeister, Robin Tibor, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping* 38: 391–5420. https://doi.org/10.1002/hbm.23730.
17. Burget, Felix, Lukas Dominique Josef Fiederer, Daniel Kuhner, Martin Völker, Johannes Aldinger, Robin Tibor Schirrmeister, Chau Do, et al. 2017. Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills. arXiv:1707.06633 [cs].
18. Glasser, Matthew F., Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, et al. 2016. A multi-modal parcellation of human cerebral cortex. Nature advance online publication. https://doi.org/10.1038/nature18933.
19. Lu, Jie, Vahid Behbood, Peng Hao, Hua Zuo, Shan Xue, and Guangquan Zhang. 2015. Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems* 80. 25th Anniversary of Knowledge-Based Systems: 14–23. doi:https://doi.org/10.1016/j.knosys.2015.01.010.
20. Hodson, Richard. 2016. Precision medicine. *Nature* 537: S49–S49. https://doi.org/10.1038/537S49a.
21. Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542: 115–118. https://doi.org/10.1038/nature21056.
22. Colic, Sinisa, Robert G. Wither, Min Lang, Liang Zhang, James H. Eubanks, and Berj L. Bardakjian. 2017. Prediction of antiepileptic drug treatment outcomes using machine learning. *Journal of Neural Engineering* 14: 016002. https://doi.org/10.1088/1741-2560/14/1/016002.
23. Yu, Kun-Hsing, Zhang Ce, Gerald J. Berry, Russ B. Altman, Ré Christopher, Daniel L. Rubin, and Michael Snyder. 2016. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications* 7: 12474. https://doi.org/10.1038/ncomms12474.
24. Burget, F., L. D. J. Fiederer, D. Kuhner, M. Volker, J. Aldinger, R. T. Schirrmeister, C. Do, et al. 2017. Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills. In, 1–6. IEEE. 10.1109/ECMR.2017.8098658.
25. Kriegeskorte, Nikolaus. 2015. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science* 1: 417–446. https://doi.org/10.1146/annurev-vision-082114-035447.
26. Güçlü, Umut, and Marcel A.J. van Gerven. 2014. Unsupervised feature learning improves prediction of human brain activity in response to natural images. *PLoS Computational Biology* 10: e1003724. https://doi.org/10.1371/journal.pcbi.1003724.
27. Adkins, Daniel E. 2017. Machine learning and electronic health records: A paradigm shift. *American Journal of Psychiatry* 174: 93–94. https://doi.org/10.1176/appi.ajp.2016.16101169.
28. Karwath, Andreas, Markus Hubrich, Stefan Kramer, and the Alzheimer's Disease Neuroimaging Initiative. 2017. Convolutional neural networks for the identification of regions of interest in PET scans: A study of representation learning for diagnosing Alzheimer's disease. In *Artificial Intelligence in Medicine*, 316–321. Cham: Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-319-59758-4_36.
29. Moradi, Elaheh, Antonietta Pepe, Christian Gaser, Heikki Huttunen, and Jussi Tohka. 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104: 398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002.
30. Roberts, Thomas A., Ben Hipwell, Giulia Agliardi, Angela d'Esposito, Valerie Taylor, Mark F. Lythgoe, and Simon Walker-Samuel. 2017. Deep learning diffusion fingerprinting to detect brain tumour response to chemotherapy. *bioRxiv*: 193730. https://doi.org/10.1101/193730.
31. Salvatore, C., A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M.C. Gilardi, and A. Quattrone. 2014. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive Supranuclear palsy. *Journal of Neuroscience Methods* 222: 230–237. https://doi.org/10.1016/j.jneumeth.2013.11.016.
32. Young, Jonathan, Matthew J. Kempton, and Philip McGuire. 2016. Using machine learning to predict outcomes in psychosis. *The Lancet Psychiatry* 3: 908–909. https://doi.org/10.1016/S2215-0366(16)30218-8.
33. Kessler, R.C., H.M. van Loo, K.J. Wardenaar, R.M. Bossarte, L.A. Brenner, T. Cai, D.D. Ebert, I. Hwang, J. Li, P. de Jonge, A.A. Nierenberg, M.V. Petukhova, A.J. Rosellini, N.A. Sampson, R.A. Schoevers, M.A. Wilcox, and A.M. Zaslavsky. 2016. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry* 21: 1366–1371. https://doi.org/10.1038/mp.2015.198.
34. Al-Azizy, Dalal, David Millard, Iraklis Symeonidis, Kieron O'Hara, and Nigel Shadbolt. 2015. A literature survey and classifications on data Deanonymisation. In *Risks and Security of Internet and Systems*, ed. Costas Lambrinoudakis and Alban Gabillon, 36–51. Lecture notes in computer science 9572. Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-31811-0_3.
35. Kumar, Kuldeep, Laurent Chauvin, Matthew Toews, Olivier Colliot, and Christian Desrosiers. 2017. *Multi-modal brain fingerprinting: a manifold approximation based framework. bioRxiv*: 209726. https://doi.org/10.1101/209726.

36. Hern, Alex. 2017. Hacking risk leads to recall of 500,000 pacemakers due to patient death fears. *The Guardian*, August 31, sec. Technology.
37. Maisel, William H.M.D. 2005. Safety issues involving medical devices: Implications of recent implantable cardioverter-defibrillator malfunctions. [editorial]. *JAMA* 294: 955–958.
38. Halperin, D., T. S. Heydt-Benjamin, B. Ransford, S. S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. H. Maisel. 2008. Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 129–142. doi:https://doi.org/10.1109/SP.2008.31.
39. Yadron, Danny. 2016. "Worth it": FBI admits it paid $1.3m to hack into San Bernardino iPhone. *The Guardian*, April 21, sec. US news.
40. Strickland, Eliza. 2017. Facebook announces "typing-by-brain" project. *IEEE Spectrum: Technology, Engineering, and Science News*. (April 20).
41. Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. 2016. Practical secure aggregation for federated learning on user-held data. *arXiv: 1611.04482 [cs, stat]*.
42. Differential Privacy. 2018. https://privacytools.seas.harvard.edu/differential-privacy. Accessed January 25.
43. Alphabet and Amazon want to protect you from hackers. That's a blessing and a curse. 2018. MIT technology review. https://www.technologyreview.com/the-download/610061/alphabet-and-amazon-want-to-protect-you-from-hackers-thats-a-blessing-and-a/. Accessed January 29.
44. Wilbanks, John T., and Eric J. Topol. 2016. Stop the privatization of health data. *Nature News* 535: 345–348. https://doi.org/10.1038/535345a.
45. Anton, Annie I., Julia B. Earp, and Jessica D. Young. 2010. How internet users' privacy concerns have evolved since 2002. *IEEE Journals & Magazine. IEEE Security & Privacy* 8: 21–27. https://doi.org/10.1109/MSP.2010.38.
46. Debatin, Bernhard, Jennette P. Lovejoy, Ann-Kathrin Horn, and Brittany N. Hughes. 2009. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication* 15: 83–108. https://doi.org/10.1111/j.1083-6101.2009.01494.x.
47. Luger, Ewa, Stuart Moran, and Tom Rodden. 2013. Consent for All: Revealing the Hidden Complexity of Terms and Conditions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2687–2696. CHI '13. New York, NY, USA: ACM. https://doi.org/10.1145/2470654.2481371.
48. Chaudhry, Amir, Jon Crowcroft, Heidi Howard, Anil Madhavapeddy, Richard Mortier, Hamed Haddadi, and Derek McAuley. 2015. Personal Data: Thinking Inside the Box. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*, 29–32. AA '15. Aarhus, Denmark: Aarhus University Press. doi:https://doi.org/10.7146/aahcc.v1i1.21312.
49. Ienca, Marcello, and Roberto Andorno. 2017. Towards new human rights in the age of neuroscience and neurotechnology. Life Sciences, Society and Policy 13: 5https://doi.org/10.1186/s40504-017-0050-1.
50. Kellmeyer, Philipp, Thomas Cochrane, Oliver Müller, Christine Mitchell, Tonio Ball, Joseph J. Fins, and Nikola Biller-Andorno. 2016. The effects of closed-loop medical devices on the autonomy and accountability of persons and systems. *Cambridge quarterly of healthcare ethics: CQ: the international journal of healthcare ethics committees* 25: 623–633. https://doi.org/10.1017/S0963180116000359.
51. Gilbert, Frederic. 2015. A threat to autonomy? The intrusion of predictive brain implants. *American Journal of Bioethics Neuroscience* 6: 4–11. https://doi.org/10.1080/21507740.2015.1076087.
52. Dings, Roy, and Leon de Bruin. 2016. Situating the self: Understanding the effects of deep brain stimulation. *Phenomenology and the Cognitive Sciences* 15: 151–165. https://doi.org/10.1007/s11097-015-9421-3.
53. Sarajlic, Eldar. 2015. Do predictive brain implants threaten Patient's autonomy or authenticity? *American Journal of Bioethics Neuroscience* 6: 30–32. https://doi.org/10.1080/21507740.2015.1094538.
54. Tversky, Amos, and Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5: 207–232. https://doi.org/10.1016/0010-0285(73)90033-9.
55. Knight, Will. 2017. Biased algorithms are everywhere, and no one seems to care. *MIT Technology Review*, July 12.
56. Chen, Jonathan H., and Steven M. Asch. 2017. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *The New England Journal of Medicine* 376: 2507–2509. https://doi.org/10.1056/NEJMp1702071.
57. Yuste, Rafael, Sara Goering, Blaise Agüera y Arcas, Guoqiang Bi, Jose M. Carmena, Adrian Carter, Joseph J. Fins, et al. 2017. Four ethical priorities for neurotechnologies and AI. *Nature News* 551: 159. https://doi.org/10.1038/551159a.
58. Kellmeyer, Philipp. 2017. Ethical and legal implications of the methodological crisis in neuroimaging. Cambridge quarterly of healthcare ethics: CQ: the international journal of healthcare ethics committees 26: 530–554. https://doi.org/10.1017/S096318011700007X.
59. Zhou, Lina, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos. 2017. Machine learning on big data: Opportunities and challenges. 237: 350–361. https://doi.org/10.1016/j.neucom.2017.01.026.
60. Lavazza, Andrea. 2018. Freedom of thought and mental integrity: The moral requirements for any neural prosthesis. *Frontiers in Neuroscience* 12. https://doi.org/10.3389/fnins.2018.00082.
61. Bublitz, Jan Christoph, and Reinhard Merkel. 2014. Crimes against minds: On mental manipulations, harms and a human right to mental self-determination. *Criminal Law and Philosophy* 8: 51–77. https://doi.org/10.1007/s11572-012-9172-y.
62. Blitz, Marc Jonathan. 2010. Freedom of thought for the extended mind: Cognitive enhancement and the constitution. *Wisconsin Law Review* 2010: 1049.
63. Halliburton, Christian M. 2007. Letting Katz out of the bag: Cognitive freedom and fourth amendment Fidelity. *Hastings Law Journal* 59: 309.
64. Boire, Richard G. 2005. Searching the brain: The fourth amendment implications of brain-based deception detection devices. *The American Journal of Bioethics* 5: 62–63. https://doi.org/10.1080/15265160590960933.

65. Kerr, Orin S. 2004. The fourth amendment and new technologies: Constitutional myths and the case for caution. *Michigan Law Review* 102: 801–888. https://doi.org/10.2307/4141982.
66. Ienca, Marcello. 2017. The right to cognitive liberty. *Scientific American* 317: 10–10. https://doi.org/10.1038/scientificamerican0817-10.
67. Bublitz, Christoph. 2016. Moral enhancement and mental freedom. *Journal of Applied Philosophy* 33: 88–106. https://doi.org/10.1111/japp.12108.
68. Barfield, Woodrow. 2015. Cognitive liberty, brain implants, and Neuroprosthesis. In *Cyber-Humans*, 101–133. Cham: Copernicus. https://doi.org/10.1007/978-3-319-25050-2_4.
69. Bublitz, Christoph. 2015. Cognitive liberty or the international human right to freedom of thought. In Handbook of Neuroethics, 1309–1333. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4707-4_166.
70. Sententia, Wrye. 2006. Neuroethical considerations: Cognitive liberty and converging technologies for improving human cognition. *Annals of the New York Academy of Sciences* 1013: 221–228. https://doi.org/10.1196/annals.1305.014.
71. Kayser, Manfred, and Peter M. Schneider. 2009. DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International. Genetics* 3: 154–161. https://doi.org/10.1016/j.fsigen.2009.01.012.
72. Shead, Sam, 10:00 26. 03.2016, and 314. 2016. The biggest mystery in AI right now is the ethics board that Google set up after buying DeepMind. *Business Insider Deutschland*. http://www.businessinsider.de/google-ai-ethics-board-remains-a-mystery-2016-3. Accessed September 22.
73. Kelion, Leo. 2017. Facebook artificial intelligence spots suicidal users. *BBC News*, March 1, sec. Technology.
74. Tait, Joyce. 2001. More Faust than Frankenstein: The European debate about the precautionary principle and risk regulation for genetically modified crops. *Journal of Risk Research* 4: 175–189. https://doi.org/10.1080/13669870010027640.
75. Beck, Ulrich. 1992. *Risk Society: Towards a New Modernity*. London ; Newbury Park, calif: SAGE Publications Ltd.
76. Sunstein, Cass. 2005. The precautionary principle as a basis for decision making. *The Economists' Voice* 2: 1–10.
77. Parens, Erik. 2014. *Shaping our selves: On technology, flourishing, and a habit of thinking*. Oxford University Press.