# Ethics in the Digital Age: Privacy, Data Protection, and other Big Data Challenges

Johnny

…
Advisor: …

Big Data is ubiquitous in our modern digital society, but it raises many new ethical questions. In this seminar paper, we present an overview of ethical challenges connected to Big Data, as well as ways to approach them, based on the current state of scientific literature. We first give an introduction to the terms "Big Data" and "digital ethics". We subsequently discuss 5 ethical challenges that are significantly impacted by Big Data: *privacy*, *algorithmic bias*, *manipulation*, *accountability*, and *labour*. Finally, we examine approaches to these ethical challenges that have been recommended, proposed, and practiced by the industry, lawmakers, and researchers.

## 1 Introduction

Interconnected satellites and sensors, mobile phones, artificial intelligence, social networks, targeted advertisement and the Internet at large – these are some of the technological developments of recent decades that we associate with Big Data. Whether we realise it or not, we all feed into massive apparatuses that collect, process and analyse data at scale and repurpose the results to meet various ends. Since Big Data inherently affects both the individual and the collective in novel ways, it has spawned a discussion of countless issues from the perspective of ethics, as well as potential approaches for their resolution. In this overview, we aim to give an overview of the literature on ethics and Big Data by highlighting ethical issues and discussing potential solutions.

Before evaluating the specific ethical challenges and approaches for tackling them, it is crucial to gain clarity regarding the meaning of the central terminology used in this work: *Big Data* and *digital ethics*. This will provide a concrete idea of the social and technical sides of the issues at hand. It will also give an idea of the ethical theories in terms of which these issues and their proposed solutions may be analysed. The introduction to these two terms serves as a way to make the topic of discussion more concrete for the rest of the work.

## 1.1 Big Data

The term *Big Data*[1] saw its first uses in the 1980s, although it took until the middle of the 1990s for it to be associated with the phenomenon we have in mind today [19]. The phenomenon Big Data encompasses dimensions concerning the nature of the data in question itself, the technologies surrounding it ("Big Data technologies"), and the sociotechnological trends that led to the rapid growth of Big Data.

The "data" in Big Data are commonly characterised in the industry by 3 sub-dimensions: volume, velocity, and variety [29, 58]. Volume describes the amount of data involved, velocity the speed at which they are generated, and variety their heterogeneity regarding structure and format. Viewed in these dimensions, Big Data is large in volume, fast in velocity and (typically) high in variety, respectively [29, 58]. To make use of these unique characteristics, technologies have emerged that process Big Data at varying levels of specificity. Examples include large language models (LLMs) trained on data sets comprising large amounts of web pages [11], predictive surveillance systems employed by police [10], and targeted programmatic advertisement [36]. Machine learning in particular, while facing technical challenges regarding the aforementioned characteristics of Big Data, can profit from new sources of training data [34]. Finally, Big Data emerged from the Internet and the rapid digitalisation of the last decades [31], which is in part a social phenomenon. Most notably for the purposes of this work, the rise of social media has contributed greatly to user-generated Big Data [47, 61]. The ubiquity of "smart" devices, also known as the "Internet of Things" (IoT), is another example for such a development. Big Data amassed through these trends frequently feeds back into them. For example, data collected from user interaction on social networks are used to develop so-called *recommender systems* that determine what users see and get recommended in those same social networks [1].

## 1.2 Digital Ethics

Digital ethics, or the *ethics of computing* as systematised by Stahl, Timmermans, and Mittelstadt [54], can be understood as the body of work concerned with *ethical issues* caused or influenced by *technology* and how these issues can be interpreted from the perspective of *ethical theory*. It often also contains *recommendations* for how to mitigate those issues[2]. Stahl, Timmermans, and Mittelstadt [54] provide 5 categories for ethical issues in computing: *core ethical concepts* (issues rooted in moral philosophy) as well as *fundamental and theoretical*, *social and practical*, *legal*, and *technical* issues. Recommendations are things like "guidelines, tools, [...] awareness raising, contribution to policy and debate" [54], made in response to ethical issues. Ethical theories provide frameworks of "well-known positions [...] that are widely used in reflecting on why a particular action might be considered good or bad" [54]. They establish a foundation for evaluating ethical issues.

[1] While the word "data" is a plural form and will be treated as such, *Big Data* is considered a proper noun and used as singular throughout this work.

[2] The survey of Stahl, Timmermans, and Mittelstadt [54] lists two more items: *contribution* and *methodology*. These are more specific to research papers, while this work provides an overview reaching beyond research, hence they have been excluded.

Computing ethics can generally be categorised under the field of *applied ethics*, which deals with ethical analysis of specific areas and professions, in this case the area of digital technologies and work in computing [54].

## 1.3 Applying Ethical Theory to Big Data

Numerous connections between computing and traditional ethical theories have been made, in particular deontology and utilitarianism [54]. Deontology states that "moral quality of an action is to be located in the intention of the agent" [54], i.e. an action is morally good if the decision to perform it was motivated by good intentions. Closely connected to deontology is the work of Immanuel Kant, who is also referenced frequently in the ethics of computing [54]. "Kantian analysis argues that one should always respect the autonomy of other people, treating them as ends in themselves and never only as means to an end." [31]. On the basis of this observation, Herschel and Miori [31] view Big Data is highly problematic from a Kantian perspective. Big Data, by design, removes individuality by treating people "as data points" [31]. It depraves them of the autonomy over their data in order to serve an interest of the individual that is merely presumed by the entities behind Big Data [31].

In contrast to deontology, consequentialist theories like utilitarianism state that "the ethical value of an action is to be found in its consequences" [54] rather than the intentions behind it. Utilitarianism specifically is based on the idea that ethical value can be *measured*: "the aggregate amount of happiness minus the aggregate amount of pain caused by an action are the measure of its ethical quality" [54]. A utilitarian analysis of Big Data is not as straight-forward as a Kantian one, because the outcomes of Big Data need to be evaluated: does Big Data lead to a net increase in happiness? Herschel and Miori [31] argue that a utilitarian evaluation of Big Data is not feasible, since the topic is too vast and complex. It would be subject to bias, imprecision, and especially a societal lack of clarity regarding the meaning and pervasiveness of Big Data.

A similar argument is made by Zwitter [61]. He states that it is difficult to apply *any* traditional ethical theory to Big Data because Big Data challenges the individualist notion of "moral agency" that is central to those theories. Instead, he proposes that the interdependence of "Big Data collectors, Big Data utilizers, and Big Data generators" leads to "dependent agency", where the moral quality of an agent's action can no longer be evaluated in isolation from other actors [61].

Contemporary ethical theories exist that are better suited to analyse Big Data. Of particular note in this space is the work of Floridi, whose theories are among the most prolific in contemporary ethics of computing [54]. For example, he examines the aforementioned issues with moral agency in relation to Big Data in his work on *distributed morality* [61]. By far the most interesting ethical framework for a paper about Big Data and ethics is Floridi's *data ethics* [28]. Data ethics shifts the focus away from specific technology, hardware or software and instead focuses on data as "the true invariant" [28] across the ethics of computing. Data ethics consists of multiple dimensions: the ethics of data (issues pertaining to the handling of data itself), the ethics of algorithms (issues regarding the actions performed by technologies, particularly artificial intelligence), and the ethics of practices (issues regarding professional standards and policies) [28]. This maps quite well

to the approach of this paper: the ethics of data and algorithms are crucial for the issues outlined in section 2, while the ethics of practices inevitably comes up when discussing approaches to mitigate these issues. Floridi further embeds data ethics as a component of his theory of *digital governance*, where it "shapes [digital governance] and [digital regulation] through moral evaluation" [26]. Connected to this idea of an interplay between ethics, governance, and regulation is the distinction between *soft* and *hard* digital ethics. Hard ethics sets moral boundaries (i.e., determines what is right and what is wrong) that shape or influence regulation, while soft ethics concerns normative questions beyond the law [26]. The ideas of digital governance and the distinction between hard and soft digital ethics will play a role again when categorising and evaluating approaches in section 3.

# 2 Ethical Challenges of Big Data

As described in subsection 1.2, Big Data leads to the creation of new ethical issues such as algorithmic bias. It furthermore contributes to the vanishing of individualised ethics and thus changes some well-known ethical issues fundamentally, e.g. privacy. The following subsections each discuss a specific ethical issue, focusing on how it is impacted by Big Data in particular. Many more issues exist that are out of scope for this work but are worth examining in the context of Big Data nonetheless, such as theoretical issues like identity and agency, or more practical issues like consent and education [54]. In some instances, these issues intertwine.

## 2.1 Privacy

An immediate question that comes up in relation to Big Data is whether Big Data is compatible with privacy, which is a human right [41], a well-studied problem in ethics, and also the most commonly discussed ethical concept in the literature on computing ethics [54]. While Big Data is a framework for gaining statistical insights and therefore often allows for the removal of "individualised" data, i.e. data "connected to one specific person" [61], privacy is still a problem. Big Data operations may provide individualised privacy through mechanisms like anonymisation, but due to the correlation of one person's data with that of many others, incredibly detailed information can be obtained by association with *groups*. This is by design: without the ability to make inferences about groups, Big Data would be useless [61]. As mentioned in subsection 1.3, this poses a problem for ethics, which has traditionally been concerned with the individual.

The theory of group privacy (called "collective privacy" [38] in some instances) attempts to resolve this issue by recognising that Big Data technologies operate at the level of groups rather than that of individuals. In other words: individualised data is mostly irrelevant to Big Data, which more-so involves statistical patterns and groupings. Big Data's effects on the individual are largely incidental [57]. Group privacy shifts the focus "from 'their' to 'its' privacy with regard to the group" [57], i.e. the group is understood as an entity beyond the aggregated interests of its individual members.

The first observation of group privacy is that "groups" are *constructed by* Big Data. Different Big Data applications, even though they may operate on the same kind of data,

may select different "features of interest, according to some particular purpose" [57]. Such differences are enough to change the groups whose privacies are to be considered [38, 57]. This notion of a group is much more dynamic than that traditionally found in sociological group theory: Big Data technologies may create groups whose characteristics are beyond the perception of both its members and outside observers, which is needed to recognise a group in the traditional sense [33, 38]. For example, online targeted advertisement puts people in groups based on their shopping behaviour, even though nobody except the classification algorithm is aware of this group's existence or its constituents.

Traditional definitions of privacy that focus on the individual encompass multiple questions that are often intertwined with, and derive their moral underpinnings from, other ethical concepts such as individual autonomy and freedom. As pointed out in the previous paragraphs, individualised notions of privacy are problematic in the context of Big Data. Data privacy for example, i.e. the power over "authorship, movement, and modification" [54] of one's data, is called into question by Big Data through a disregard for autonomy. Similarly, personal privacy concepts like "the right to be left alone" or the freedom to self-determination are difficult to apply to Big Data because of the ubiquity of data processing and the inability of the individual to influence it [31, 33].

This leads to the question what group privacy entails exactly, and what a group privacy infringement looks like. Kammourieh et al. [33] give a very clear example of a group privacy problem in practice. They describe how, in 2013, data was released that contained detailed information about the work of New York taxi drivers, including which routes they took at what time and how many people were transported. This data was technically anonymised. However, it was possible to make inferences about *groups* of drivers. In this case, groups of devout Muslim drivers could be found by analysing the breaks they took. Since they would make regular stops for their daily prayers, this pattern was easily recognisable in the dataset [33]. The privacy of the individual was not infringed here. However, the privacy of the group of devout Muslim drivers was, as associations were created between the shared religious characteristic of that group and information whose relation would otherwise be unknown.

Credit scoring systems that take into account the geographical area or neighbourhoods that people are part of are another example of a data-driven system that creates a group privacy issue. The notable property of such a system is that the group-inducing characteristic (place of residence) is a purely collective one, yet it leads to a social and financial impact at the individual level. Therefore, a collective interest in group privacy can also make a difference for individuals [38].

In light of these observations, effective group privacy in the context of Big Data can be understood as a form of reduction, prevention, and limitation of potential harms to groups caused by Big Data [38]. From the positive angle, it also contains the right to sovereignty and self-determination, in particular for groups that are imperceivable constructions of Big Data and who are therefore not self-aware or able to represent themselves using traditional legal or moral instruments [33].

## 2.2 Algorithmic Bias

Big Data works on the basis of two presuppositions: knowledge of correlations allows us to build effective models of reality, and these correlations can be found by analysing large amounts of existing data [43]. Potentials for bias, i.e. the production of unfair disadvantages, can be identified in both of these components.

First, there is a fundamental criticism of using correlations as the basis for predictive models. Just because Big Data finds a correlation between two properties, this does not allow for conclusions regarding *causation*. Moreover, finding causes for relations in data is not even a *goal* of Big Data: its goal is "to predict and target, not to provide any sociological accounting for the reasons why people might seem to occupy particular patterns of life" [39, p. 40]. Applying Big Data's correlation-centered view to *science* can be particularly problematic. Fields like sociogenomics attempt to use data analysis to correlate genetic traits with social behaviours in humans. A disregard for causality in such contexts can be a catalyst for pseudoscientific tendencies [39, pp. 68f.].

A more concrete issue with basing models on mere correlations can be found in the concept of *proxies*, i.e. the substitution of a complex, qualitative property for a simpler numerical value that is assumed to be correlated to the real observation. For example, a property that is difficult to quantify is human health. So, in order to build predictive models for healthcare systems, data analysts have traditionally substituted "health" for the proxy of "health care cost". These models are thus designed to predict health care cost, but treated as if they actually predicted health, under the assumption that cost is an effective stand-in for level of illness [44]. Such proxies lead to weaker, more distant and discriminatory correlations [43, pp. 17f.] [39, p. 68]. The proxy of health care cost, in particular, has led to a significant racial bias: due to unequal access to health care in general, black patients in the US were assigned the same risk score (and thus, level of covered treatment) as white patients despite being significantly more ill [44].

Data in itself can already include biases. This begins with the *selection* of data, or what is "configured for 'capture'" [17]. For example, the data used to train OpenAI's GPT-3 language model included parts of Common Crawl, a dataset containing many pages of the world wide web [11]. While one might assume that the world wide web is a good approximation of the diversity of thought, people, and culture found in the real world, this is not the case. First, Internet access is not available everywhere and to everyone in the world, leading to an overrepresentation of younger people from developed countries [5]. Furthermore, datasets like these additionally get *curated* and filtered before being used as training data, which can accentuate disparities between apparent representation and reality even more [5]. This can happen when, for example, an LGBTQ forum is removed from the dataset, but a social network predominantly used by white men is kept, which would lead to the language model learning a bias towards white men. Many such biases are known to exist in large language models [5], even in subtle ways like prejudice based on dialect [32]. As another example, a data selection bias can also be found in commercial face recognition software, which performs significantly worse on dark-skinned women compared to light-skinned men [13].

Eliminating statistical representation bias from datasets is not sufficient for eliminating bias from data altogether, however. The context of dataset creation, i.e. the goals of

creating a dataset, the values associated with it, and the conditions of data work involved (see subsection 2.5), are another vector for introducing bias [18]. For example, the popular computer vision dataset ImageNet was created with the epistemological assumption that there is "an underlying and universal organization of the visual world in to clearly demarcated concepts" [18]. This is a problem, because the meaning of images cannot be determined from an absolute or objective viewpoint: lived experiences, situation, and context have an impact on how we understand images. This nuance is completely absent from ImageNet [18]. A similar problem comes to light with cultural differences between those that dictate the parameters of Big Data systems and those that are asked to do data work. For example, a US company may request from workers classification tasks such as deciding whether social media posts are hateful or not, but the US-centric worldview of what counts as hateful that is presumed self-evident may not be well understood by a Latin American worker [40].

Finally, there is a potential for a kind of historical bias in Big Data systems. Since classification and prediction rely on data collected and analysed in advance, Big Data's "modelling is stuck in abstractions drawn from the past, and so becomes a rearrangement of the way things have been rather than a reimagining of the way things could be" [39, p. 43]. This is especially a problem for machine learning models that are computationally expensive to train and hard to change after the fact. The notion of static data conflicts with changing societal views, because it can create "value-lock" [5]. Going even further, when Big Data's predictions are applied to decisions that impact the future, it can create a kind of feedback loop, where its models "help to create the environment that justifies their assumptions" [43, p. 29]. In other words, if biased, data-driven predictive technologies are used as a basis for decision making, they validate themselves by causing the manifestation of their own predictions.

## 2.3 Manipulation

Big Data is often discussed as a potential threat to democracy and political processes [22, 24, 30, 43, 46, 47]. This is often combined with the previously discussed issue of privacy. This section, however, focuses on one particular issue with regards to Big Data and politics: *manipulation*, i.e. ways in which Big Data can be used to undermine political discourse in general, and elections specifically.

The issue of Big Data becoming a tool for political manipulation first entered the broader public sphere following the *Cambridge Analytica* scandal, where user data from a Facebook application and other data brokers were secretly misappropriated to build profiles of 87 million Facebook users before the 2016 US presidential election [4, 14]. The purpose of this operation was to craft personalised political advertisements for potential Trump voters [14]. The data had been obtained through an innocuous app called "thisisyourdigitallife", which was connected to Facebook via its API and initially supposedly used for research [14].

Using Big Data to profile Internet users for the advertising purposes is what's known as *microtargeting* or *targeted advertisement*. It involves three parties: advertisers, who want to show advertisements to specific groups of users, ad platforms like *Google Ads* and *Meta ads*, who collect and analyse data about their users to make the resulting groupings available to advertisers, and users, who interact with online services and see targeted

advertisements [53]. The notion of *group* in this context as well as the privacy issues connected to it have been covered by this work in subsection 2.1. Microtargeting is common practice in the business models of all major advertisement-based social media services today: Facebook and Instagram use Meta Ads, YouTube uses Google Ads, X (formerly Twitter) and TikTok have their own platforms for targeted advertisements.

Doubt has been cast on whether the effectiveness of microtargeting lives up to what is promised by the ad platforms selling it, with some calling it "little more than 'snake oil'" [52]. Recent research appears to support this thesis, showing that while attribute-based targeting for political advertisement in general can be significantly more effective than displaying the same ads to absolutely everyone, no benefit is gained from the combination of many attributes to target more specific groups [56]. But whether political microtargeting is as effective as advertised or not, it sees widespread use in contemporary politics [24, 47], and concerns regarding privacy and transparency remain, both of which relate to fundamental ideas in democracy. Political microtargeting is not prevented by strong data protection laws [47], and can be used for discriminatory purposes, even if directly discriminatory attributes like race or gender are unavailable [53].

Big Data also bears potential for manipulation in different ways, e.g. through the amplification of certain political messaging in social media. A study of the Facebook pages of German political parties has shown that post recommender systems can be manipulated by "hyperactive users", i.e., users who like and comment significantly more than regular users, to favour certain political agendas [46].

## 2.4 Accountability

Data-driven algorithms and machine-learning-based artificial intelligence (AI) are used more and more frequently to automate decision making. When decisions emitted from such systems cause harm, who can be held responsible? This question is a multifaceted ethical issue and has been discussed in the literature on computing ethics primarily in the context of *agency*, where the question is whether and to which degree machines can be considered moral agents, i.e. take moral responsibility for actions [54]. This involves philosophical arguments about the nature of AI in relation to morality [27, 60] as well as discussion of the creation of a legal personhood for AI [12]. However, these discussions digress from the more limited question of accountability in Big Data that this work is concerned with: who is responsible for the results of using Big Data to justify decisions that affect humans?

One particularly problematic aspect of decisions made by Big-Data-based algorithms is their transparency. This is especially a problem with neural-network-based AI: AI always arrives at a specific, numerically discrete conclusion, but explaining *how* it arrived there is not part of its regular operation [39, pp. 20 ff.]. Even algorithms that are not based on neural networks can grow so complex that it becomes virtually impossible to keep track of how they operate. In addition to this, the workings of many Big Data technologies, such as social scoring systems, are kept secret intentionally. Scale and opacity are two characteristics of what O'Neil [43] calls "weapons of math destruction": data-driven models that do unfair and unjust damage to (parts of) society [43, p. 31]. The accountability issue in Big Data technologies can be exemplified even on a small scale. O'Neil [43] presents

a case from Washington D.C., where an algorithmic performance scoring system was used to decide which school teachers to lay off and which to keep. Teachers affected by this system had no way to get an explanation of their score. Furthermore, the algorithm lacked external verification or examination for potential flaws, effectively elevating a system that nobody truly understood as the single source of truth. Since one criterion for evaluation was students' scores in standardised tests, it became apparent that this system could be manipulated. In one case, this led to the unfair termination of a teacher who did an excellent job according to parents, students, and other teachers, but who was still assigned a low score by the algorithm, because other teachers had fabricated test results in the previous year [43].

Another, more severe example for a lack of accountability in data-driven systems is the LSI-R recidivism model, used in the US to predict how likely prisoners are to commit a crime after they regain their freedom ("risk scoring"). The system incorporates information that is unrelated to the crimes for which prisoners are convicted, such as their neighbourhood, first experience with police, or number of family members that have been convicted of a crime [43, pp. 25f.]. These factors, which are disconnected from the actual crime, might *correlate* with risk of recidivism, but using them to influence things like sentencing decisions is fundamentally unjust. On their own, they would even be inadmissible in court [43, p. 26]. This fact is obscured by using Big Data as an intermediary to turn qualitative analysis of a convict's situation into a numerical risk score. The algorithm is not accountable to regular transparency requirements and due process, and therefore its invocations "can have the force of the law without being of the law, and will create what we might call 'algorithmic states of exception'" [39, p. 76].

Frequently, such automated data-driven systems are designed with a supposed remedy, the so-called "human-in-the-loop": a person double-checking results with the intention to ensure the detection and prevention of mistakes made by the automated system. In cases like the teacher performance scoring, where a flawed algorithm causes harm, one initially plausible proposition might be to hold the humans supervising or operating it to account for its failures. However, this view is simplistic and does not sufficiently account for the complexities of socio-technical systems, mainly because there is an issue of balance: human operators typically do not have full control over, or even knowledge about, the automation involved in the systems they supervise. This is exacerbated by "the problem of many hands" [42], i.e. the fact that many disconnected actors who influence different parts of the systems are involved in Big Data: data brokers, data analysts, programmers, executives, regulators, and many more. This is one of several barriers that create gaps in the accountability of computer-based systems [23, 42]. While it is thus not generally appropriate to assign unique blame for system failures to individual humans [16], this idea of individual responsibility still dominates discussions around and coverage of accidents or injustices involving automated systems with a human in the loop [23].

Distributing control among autonomous systems and humans but retaining practically full responsibility for humans is a phenomenon which Elish [23] calls "moral crumple zones". While crumple zones in modern cars distribute force to protect human drivers in case of a crash, moral crumple zones are "ways in which automated and autonomous systems deflect responsibility in unique and structural ways, protecting the integrity of the technological system at the expense of the nearest human operator" [23]. A topical and

contemporary example of moral crumple zones is in the design of so-called "self-driving" vehicles, where the autonomous part of the system hands off controls to a human operator in case of a system failure or an imminent crash, transferring apparent responsibility to the human-in-the-loop at the last moment [23].

The gap in accountability of data-driven decisions does not only affect their human operators and humans harmed by them, but society's attitude towards decision making processes in general. Combined with complete faith in the correctness of the system, this gap can be viewed as an incubator for the emergence of "thoughtlessness". Thoughtlessness describes a mode of operation where processes (such as administrative decisions about employment) are neither fully understood nor questioned by their human operators, while consequences of these processes are ignored [39, p. 62]. Automated, data-driven decision making "amplifies this by adding computational opacity and technical authority" [39, p. 62], i.e. Big Data obfuscates the decision making process and evokes a sense of objectiveness or even infallibility.

## 2.5 Labour

AI is often presented as a fully autonomous system that does not require human intervention to function. A layperson may get the impression that AI is conceived of by collecting data, putting these data into a "training black box" and obtaining a trained model as a result. This is not the case, however. AI, and Big Data technologies more broadly, require different forms of human labour at various stages that is often made invisible [18]. While discussions of the relation between Big Data and labour in the literature has mostly been about *deployment* of AI in the workplace [50], this section is primarily concerned with the more hidden labour that goes into the *development* of AI, or what will later be defined as *data work*. This is not to say that the deployment of Big Data technologies in places of work is any less worthy of ethical evaluation, only that it is out of scope for this work.

Traditional machine learning involves the optimisation of a cost function based on preexisting knowledge of inputs and associated outputs ("training data"). A classic example for introductions to machine learning is a form of optical character recognition (OCR), where the goal is to create a program that recognises digits from pictures of handwriting. To train a neural network that can accomplish this task, one must provide a set of correctly labelled pairs between images of handwritten digits (inputs) and the actual corresponding digits (outputs). An example for a dataset that is commonly used for this purpose is the *MNIST* database [8]. This is what's known as *supervised* machine learning [59]. Crucially, supervised machine learning requires up front human labour: someone has to prepare the training data.

The required human efforts for AI change significantly as modern AI systems become increasingly versatile. Instead of fulfilling only a single task, such as recognising handwritten digits, AI systems are developed as much more general solutions for a broader range of problems. For instance, while LLMs like GPT-3 only employ *semi*-supervised learning [11], i.e., labelling is not required to the extent of all possible outputs, training a system on data at this scale still requires a significant amount of *data work*, i.e. "labor involved in the collection, curation, classification, labeling, and verification of data" [40], to obtain datasets that can be used for training.

In principle, data work can be carried out by anyone. Solving "Are you a robot?"-prompts on the Internet, for example, is a form of data work: classifying and labelling excerpts of Big Data that supports the development of computerised recognition systems [21]. Other contexts, such as medical applications of AI, require that trained professionals contribute to data work [7]. However, with today's ubiquity and generalised application of Big Data and AI, there is a trend towards the outsourcing and precarisation of data work [20, 40, 50].

In the labour market of data work, a distinction can be made between digital labour platforms [50] and business process outsourcing (BPO) companies [40]. Platforms are digital service providers where AI developers or companies post requests for data work and workers sign up to earn money by completing these tasks, such as labelling or curating data, from home. The workers do not have the legal status of employees, but are instead considered "independent contractors" and get paid on a per-task basis [20]. This is "digital piecework" [20]: working from home and getting paid per "micro-task" completed, without a guaranteed minimum wage or basic labour protections [20, 59]. On the other hand, BPO companies also offer piecework-as-a-service, but follow more traditional managerial structures and typically focus on a more specific data work service [40]

Digital piecework is not only used for the preparation of data for AI, but also *verification*, e.g. the correction of erroneous outputs by AI that is already in use [59]. Even *impersonation*, where pieceworkers play the role of supposed "AI" by performing micro-tasks that are too difficult for actual AI, are part of what labour platforms and BPO companies offer [40, 59]. An example for such a platform is the aptly named "Amazon Mechanical Turk" [59]. A more recent example of AI impersonation was the revelation that Amazon's "Just Walk Out" grocery store system, which supposedly used AI to keep track of what you put in your bag, actually employed hundreds of Indian workers remotely monitoring camera feeds [6]. Both impersonation and verification can be subsumed under the principle of a human-in-the-loop [59], which has been discussed in more detail in subsection 2.4.

Data work is usually not considered the foundational work that it is, and discussions of its conditions and efficacy are overshadowed by the larger interest in the algorithmic side of Big Data [18]. Data work has become "naturalized infrastructure": "unquestioned, unchallenged routines" [18]. This has led to an exploitation of labour in the global south. For example, following an economic crash, micro-work platforms saw an immense increase in skilled workers from Venezuela desperate to earn money [15]. BPO workers in Kenya spend 8 hours a day drawing boxes around objects in images for training autonomous vehicles. They receive little pay and face surveillance as well as high pressure and an unhealthy work environment [35]. Very recently, 97 Kenyan data workers addressed US president Joe Biden in an open letter, calling Big Tech's treatment of workers in Africa systematically abusive and exploitative [45]. They point to the psychological toll of data work like suffering PTSD from reviewing disturbing footage, as well as US tech companies' practices described as violations of national laws and a disregard for human and labour rights.

# 3 Approaches to Ethical Challenges

In the previous section, we presented five ethical issues that are created or exacerbated by Big Data: *group privacy*, *algorithmic bias*, *manipulation*, *accountability*, and *labour*. As illustrated, these issues do not exist only in theory, but already affect people in practice today. Consequently, researchers, regulators, and industry experts are looking for ways to solve these problems.

The resulting recommendations can be categorised as *technical approaches*, *regulatory approaches*, or a form of *soft ethics*. This categorisation is not exhaustive and there can be overlaps. For example, the ethical considerations for a technical approach may also be situated in soft ethics [3].

## 3.1 Technical Approaches

Ethics cannot "be programmed into" systems. Technical approaches to ethical challenges are therefore limited to a certain level of abstraction [28]: technology cannot be expected to eliminate ethical issues that are social in nature. Assuming the opposite would be falling into the trap of "tech solutionism": "the substitution of advanced technology for any serious attempt to address the structural causes of a [social] problem" [39, pp. 43f.]. Instead, technology can sometimes be used to *mitigate* ethical problems by eliminating certain *technical parent problems*. In other words, technology can have a positive impact on computing ethics by reducing the technical incubators for ethical issues.

For instance, take the problem of government surveillance. Surveillance is a well-discussed issue in ethics [54] and includes multiple components: autonomy, privacy, trust, public security, among others. While a general answer to the question of how much surveillance should be allowed is difficult to find, most agree that there must be areas where surveillance shouldn't reach. E.g., the Universal Declaration of Human Rights bans "arbitrary interference with [one's] [...] correspondence" [41]. End-to-end encryption (E2EE) is a technical approach to the ethical issue of digital surveillance. If implemented correctly and absent total control over all levels of digital infrastructure, E2EE makes it impossible for governments or anyone else to read private messages. It doesn't *solve* the underlying issue: surveillants have not lost their interest in reading private messages. However, it eliminates a certain technical basis for surveillance, namely the *ability* to read private messages. The ongoing campaign against E2EE by certain EU member states demonstrates its efficacy [51].

Many technical approaches to issue of privacy in Big Data focus on individual autonomy over personal data and the lack of trust towards data processors. The idea is that our current lack of autonomy facilitates privacy infringements, and that we must not rely on limitations that can be technically circumvented. For example, we are often asked if we consent to the use of cookies on websites, but there is no framework in place that ensures our choice is actually respected. We have to trust the website that it *actually* doesn't send us tracking cookies. Similarly, we cannot directly verify that the systems we interact with abide by their own privacy policies. When we upload something somewhere, we lose practical control over what is going to happen to it. Recent research investigating

the cookie notice compliance of websites suggests that these concerns are warranted: a majority of websites asking users for consent ignore the user's choice [9].

One proposed technical approach to mitigate or even solve issues regarding personal data ethics is a kind of "software exoskeleton" [3]. Autili et al. [3] propose a way to "build a software exoskeleton that enhances and protects humans by mediating their interactions with the digital world according to their ethics of actions and privacy of data". In other words, the aim is to create a personal barrier at the software-level that mediates all interactions with the digital world. It is supposed to make sure that the outside world respects the user's decisions regarding their personal data and moral convictions at all times [3]. Data are coupled with the actions taken on them, turning them into what the researches call "active data". E.g., a datum is encapsulated together with an action to delete it after a certain amount of time, strengthening technical guarantees regarding what processors do to personal data [3]. The hope, overall, is to improve both privacy and accountability in the infosphere: "on the one hand users will be protected by mediating their interactions with the digital world, and on the other hand they will be responsible of the consequences of theirs [sic] (ethical) decisions" [3].

A technical approach that is already being applied in practice and not too dissimilar from the idea of a virtual exoskeleton is Tim Berners-Lee's Solid project[3]. Here, all personal data are stored in a "pod" belonging to the subject, and the subject determines what data may leave or enter their pod. What mediates the subject's interactions with the digital world therefore isn't an exoskeleton, but an external mechanism provided by the entity hosting the pod. Pods can be hosted by subjects themselves or subjects can choose "managed" pods, similar to email. Issues of privacy and autonomy are not the only things the project focuses on: its higher-level goal is to restructure the web according to Berners-Lee's original vision of decentralisation and interoperability.

The fact that neural-network-based AI produces conclusions, but no explanations (see subsection 2.4) is a well-known issue in research, given that it obstructs deeper understanding of the insights that are supposed to be gained from AI. From this technical problem emerges the ethical issue of accountability. "Explainable AI" is both a technical and methodical approach to counter this and has become a very active field in research over the last years [37]. The goal is to find ways to reduce the opacity of AI by providing human-readable explanations for the conclusions it reaches. It is still unclear how explainability relates to other characteristics of "trustworthy AI". And even by itself, the challenge of explaining AI processes is far from being solved, with many open problems remaining [37].

## 3.2 Regulatory Approaches

Regulatory approaches to Big Data ethics are ways to impose certain moral imperatives and norms on society via means of lawmaking. Such laws are informed, changed, and created by "hard ethics", which itself emerges from society [26]. Regulation is not to be seen as the singular solution to ethical problems, however, since it is "insufficient to steer society in the right direction" [26]. Still, regulatory approaches play an important role in putting normative responses to ethical issues into practice.

[3] `https://solidproject.org/` (visited on 2024-06-15)

The General Data Protection Regulation (GDPR)[4] of the EU is one of the most cited pieces of regulatory action relating to digital privacy. Much like most technical approaches to privacy, it is committed to an individualised notion of privacy that is, in part, dismantled by Big Data. The relation of the GDPR with personal data ethics has been discussed meticulously, especially its conceptions of privacy by design, impact assessment and reaction, and individual data rights [55]. One article that is worth highlighting due to its higher relevance for the ethical issues of Big Data in particular is article 22: "Automated individual decision-making, including profiling" [Art. 22 GDPR]. Unlike most of the other, more general data protection provisions in the rest of the regulation, this article pays explicit attention to Big Data. In particular, it generally forbids the use of "solely automated processing" to make decisions significantly affecting people, in legal or other ways [Art. 22(1) GDPR]. Moreover, in cases where automated decision-making *is* allowed [Art. 22(2) GDPR], it gives data subjects the right to "obtain human intervention on the part of the controller, to express [their] point of view and to contest the [automated] decision" [Art. 22(3) GDPR]. This is a reasonable regulatory approach to the issue of accountability. However, some criticism of this article is merited: the rights are limited to cases of *solely* automated decision-making. Thus, conversely, if a human operator is involved in the process (human-in-the-loop), the article becomes inapplicable, despite such systems remaining problematic regarding accountability. Art. 22(2) combined with Art. 9 imposes further restrictions on the use of sensitive personal data such as ethnicity or religion with the goal of "prevent[ing], *inter alia*, discriminatory effects" [rec. 71 GDPR]. Notwithstanding, we have seen in previous sections that the absence of *directly* discriminatory data does not prevent bias altogether.

Further regulation that affects other ethical issues with Big Data is currently being worked on or has recently been adopted in the EU.

In March of 2024, the EU parliament adopted the "AI Act", a regulation with the stated goal of ensuring responsible and safe use of socially applied AI, as well as some fundamental restrictions on what AI may be used for [48]. While the intention to regulate AI was broadly welcomed, there is strong criticism of the final version of the law. Human rights organisation *Algorithm Watch* warns about loopholes and "blanket exemptions" for national security agencies, a disregard for the rights of people outside the EU, the act's weak stance against surveillance, and its initially strong obligations for high-risk AI in the private sector being watered down to mere self-regulation [2].

Another EU regulation that recently entered into force is the "regulation on the transparency and targeting of political advertising"[5], which formulates norms for the use of targeted advertisment in the political sphere, attempting to counter the issue of manipulation. This regulation has been criticised. The motives for the regulation show a dissonance between targeted advertisement as manipulation on the one hand, and as a useful tool for campaigning and movements on the other hand. There is also an exemption for the use of political targeted advertising by state actors [24].

[4] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

[5] Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising [2024] OJ L 2024/900

The recently adopted proposal for a "platform work directive"[6] affects the issues surrounding data work that were described in this work, as well as issues concerning algorithms in the workplace. It aims to prevent circumvention of labour protections that is common practice via the "self-employed" status, and establish rules for using algorithms to monitor and score workers, following a similar approach to automatic decision-making as the GDPR [49].

## 3.3 Soft Ethics

Complimentary to the concept of hard ethics mentioned in the previous section, there is "soft ethics", which comprises norms to evaluate behaviour "*over and above* the existing regulation, not against it, or despite its scope, or to change it, or to by-pass it" [26]. Put simply, soft ethics is concerned with what norms should be adhered to after compliance with regulation, i.e. after having done the "bare minimum". These norms can provide a general approach to Big Data technologies, e.g. through industry standards for self-regulation or in the form of a broader moral framework.

Stahl and Wright [55] provide an overview of different soft ethics initiatives that aim to improve the ethics of Big Data. As a rather broad example for industry standards concerning data ethics, they mention the "information security, cybersecurity and privacy protection" family of ISO standards (ISO 27000 series) as well as the "Ethical design and application of robots and robotic systems" standard by the British Standards Institution. Other institutions like the IEEE and the Association for Computer Machinery have also started initiatives regarding ethics and Big Data, explicitly addressing issues mentioned in this work such as privacy and accountability of algorithms. The argument put forth by Stahl and Wright [55] is that these efforts can be unified under the framework of responsible research and innovation (RRI), which provides principles and methods for dealing with digital ethics. Its integration into research in the context of the Human Brain Project has demonstrated that RRI can be successfully used as an ethical guideline for Big Data research in practice [55].

In "Resisting AI", McQuillan [39] argues for "demand[ing] the possible not the probable" [39, p. 115] with regards to AI. Contemporary Big Data systems are flawed in that they always reproduce the past using statistical methods (the probable) instead of providing qualitative insights into what can and should change (the possible). The approach presented in the book is to put *care* before efficiency: "[t]he computations of large-scale models operate at millions of floating point operations per second, not 'at the pace of the toddling child or the elderly person with emphysema'" [39, p. 116]. This ethics of care can be understood as a type of soft ethics that is applied to society at large, rather than focusing only on the responsibility of Big Data controllers. Mutual aid, i.e. concrete actions of solidarity and social cohesion, can be seen as a practical tactic to respond to Big Data's ethical issues from the bottom up [39, pp. 119f.].

[6] Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on improving working conditions in platform work [2021] COM/2021/762 final

# 4 Summary and Conclusion

Big Data has become a massive sociotechnological phenomenon over the last two decades. It pervades many systems we interact with and rely on in our daily lives. Due to its societal impact, it is important to identify current issues with Big Data through the lens of digital ethics.

We have seen that Big Data is a multidimensional term, comprising technologies, certain types of data, and contemporary social trends. Digital ethics, or the ethics of computing, is a body of literature concerned with the interplay of technologies, issues, theories, and recommendations. Ethical evaluation of Big Data can be achieved by concretising the scope of computing ethics. Some theoretical perspectives on Big Data exist, both traditional and contemporary. Traditional ethical theories are often difficult to apply to Big Data due to their individualistic preconceptions.

Five ethical issues are uniquely affected by Big Data. While privacy is one of the most commonly discussed issues, *group* privacy is a recontextualisation that fits the age of Big Data more neatly. The privacy of groups is a concern that's not directly coupled to the privacy of individuals, and Big Data has great potential to infringe this novel kind of privacy. Algorithmic Bias describes unfair disadvantages created by Big Data systems and can be found in many forms and applications that already have a troubling real-world impact. Concerning the issue of manipulation, we focused on the use of microtargeting for political purposes, which is common practice and has been widely criticised since the *Cambridge Analytica* scandal, though evaluations of its efficacy in this context are still inconclusive. The issue of accountability is often connected to a lack of transparency in Big Data and leads to the question of who is responsible for decisions made or influenced by automated systems. We have seen that the "human-in-the-loop" principle is not a sufficient answer to this question. Finally, the new kind of labour ("data work") that comes with Big Data raises questions about labour and human rights. Workers who are essential for technological development are treated poorly and made invisible.

Several approaches to the ethical issues of Big Data have been proposed, through research, legislation, and the industry. They can be categorised as technical approaches, regulatory approaches, or soft ethics, oriented by Floridi's theory of digital governance and data ethics. On the technical side, there exist concepts for encapsulating personal privacy and explainable AI to combat privacy and accountability issues, respectively. In recent years, various digital regulation has been proposed and implemented in the EU to respond to issues of privacy, manipulation, and labour, among others. Soft ethics concepts have emerged in the form of the responsible research and innovation framework and industry standards such as ISO 27000. More generally, there are calls to implement a care-based approach to Big Data and AI.

The discourse around digital ethics and Big Data will only grow as the sociotechnological significance of Big Data increases. Especially the last two years have seen an immense rise in hype of AI, which is a kind of Big Data technology. Big Data systems are being promised as solutions to all kinds of problems, be it technical, social, medical, educational, or other. It is thus crucial to maintain a critical view on Big Data and be vigilant of the harms it can cause. Enforcement of ethics cannot be limited to a retroactive mode of operation. Instead, digital ethics needs to play a central role at every step of the way, from the process of

coming up with a problem domain, to the design and implementation of Big Data systems, should they be used. Big Data is not a one-size-fits-all solution. It must never be blindly applied without regard for ethical consequences. Digital regulation, technical boundaries and soft ethics can help us move towards a world where Big Data is no longer a magnet for ethical issues.

## References


[1] Flora Amato et al. "Recommendation in Social Media Networks". In: *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*. IEEE, Apr. 2017. DOI: 10.1109/bigmm.2017.55.

[2] Nikolett Aszódi. *EU's AI Act fails to set gold standard for human rights*. Apr. 2024. URL: https://algorithmwatch.org/en/ai-act-fails-to-set-gold-standard-for-human-rights/ (visited on 06/16/2024).

[3] Marco Autili et al. "A Software Exoskeleton to Protect and Support Citizen's Ethics and Privacy in the Digital World". In: *IEEE Access* 7 (2019), pp. 62011–62021. DOI: 10.1109/access.2019.2916203.

[4] Nadeem Badshah. "Facebook to contact 87 million users affected by data breach". In: *The Guardian* (Apr. 2018). URL: https://www.theguardian.com/technology/2018/apr/08/facebook-to-contact-the-87-million-users-affected-by-data-breach (visited on 06/10/2024).

[5] Emily M. Bender et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. ACM, Mar. 2021. DOI: 10.1145/3442188.3445922.

[6] Alex Bitter. "Amazon's Just Walk Out technology relies on hundreds of workers in India watching you shop". In: *Business Insider* (Apr. 2024). URL: https://www.businessinsider.com/amazons-just-walk-out-actually-1-000-people-in-india-2024-4 (visited on 06/07/2024).

[7] Claus Bossen et al. "Data work in healthcare: An Introduction". In: *Health Informatics Journal* 25.3 (Aug. 2019), pp. 465–474. DOI: 10.1177/1460458219864730.

[8] L. Bottou et al. "Comparison of classifier methods: a case study in handwritten digit recognition". In: *Proceedings of the 12th IAPR International Conference on Pattern Recognition (Cat. No.94CH3440-5)*. ICPR-94. IEEE Comput. Soc. Press, 1994. DOI: 10.1109/icpr.1994.576879.

[9] Ahmed Bouhoula et al. "Automated Large-Scale Analysis of Cookie Notice Compliance". In: *Proceedings of the 33rd USENIX Security Symposium*. Prepublication. 2024. URL: https://www.usenix.org/conference/usenixsecurity24/presentation/bouhoula (visited on 07/07/2024).

[10] Sarah Brayne. "Big Data Surveillance: The Case of Policing". In: *American Sociological Review* 82.5 (Aug. 2017), pp. 977–1008. DOI: 10.1177/0003122417725865.

[11] Tom B. Brown et al. “Language models are few-shot learners”. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS ’20. Vancouver, BC, Canada: Curran Associates Inc., 2020. ISBN: 9781713829546. DOI: 10.5555/3495724.3495883.

[12] Joanna J. Bryson, Mihailis E. Diamantis, and Thomas D. Grant. “Of, for, and by the people: the legal lacuna of synthetic persons”. In: *Artificial Intelligence and Law* 25.3 (Sept. 2017), pp. 273–291. DOI: 10.1007/s10506-017-9214-9.

[13] Joy Buolamwini and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Ed. by Sorelle A. Friedler and Christo Wilson. Vol. 81. Proceedings of Machine Learning Research. PMLR, 23–24 Feb 2018, pp. 77–91. URL: https://proceedings.mlr.press/v81/buolamwini18a.html (visited on 07/07/2024).

[14] Carole Cadwalladr and Emma Graham-Harrison. “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach”. In: *The Guardian* (Mar. 2018). URL: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (visited on 06/10/2024).

[15] Angela Chen. “Desperate Venezuelans are making money by training AI for self-driving cars”. In: *MIT Technology Review* (Aug. 2019). URL: https://www.technologyreview.com/2019/08/22/65375/venezuela-crisis-platform-work-trains-self-driving-car-ai-data/ (visited on 06/09/2024).

[16] Mark Coeckelbergh. “Moral Responsibility, Technology, and Experiences of the Tragic: From Kierkegaard to Offshore Engineering”. In: *Science and Engineering Ethics* 18.1 (Sept. 2010), pp. 35–48. DOI: 10.1007/s11948-010-9233-3.

[17] Nick Couldry and Ulises A. Mejias. “Data Colonialism: Rethinking Big Data’s Relation to the Contemporary Subject”. In: *Television & New Media* 20.4 (Sept. 2018), pp. 336–349. DOI: 10.1177/1527476418796632.

[18] Emily Denton et al. “On the genealogy of machine learning datasets: A critical history of ImageNet”. In: *Big Data & Society* 8.2 (July 2021), p. 205395172110359. DOI: 10.1177/20539517211035955.

[19] Francis X. Diebold. “On the Origin(s) and Development of the Term “Big Data””. In: *SSRN Electronic Journal* (2012). DOI: 10.2139/ssrn.2152421.

[20] Veena Dubal. “The time politics of home-based digital piecework”. In: *Center for Ethics Journal: Perspectives on Ethics, Symposium Issue “The Future of Work in the Age of Automation and AI.”* (2020), p. 50. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3649270 (visited on 07/07/2024).

[21] Josh Dzieza. “Why CAPTCHAs have gotten so difficult”. In: *The Verge* (Feb. 2019). URL: https://www.theverge.com/2019/2/1/18205610/google-captcha-ai-robot-human-difficult-artificial-intelligence (visited on 06/07/2024).

[22] Theoharris-William Efthymiou–Egleton et al. “Big Data and Democracy”. In: *HAPSc Policy Briefs Series* 1.2 (Dec. 2020), p. 18. DOI: 10.12681/hapscpbs.26473.

[23] Madeleine Clare Elish. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction". In: *Engaging Science, Technology, and Society* 5 (Mar. 2019), pp. 40–60. DOI: 10.17351/ests2019.260.

[24] Malte Engeler. "Wie viel kostet eine Wahl? Wie die Europäische Verordnung über Targeting bei politischer Werbung am Spagat zwischen Markt und Demokratie scheitert". In: *Grundrechte-Report 2024*. FISCHER Taschenbuch, May 2024, p. 193. ISBN: 978-3-596-71084-3. URL: https://netzpolitik.org/2024/grundrechte-report-2024-wie-viel-kostet-eine-wahl/ (visited on 07/07/2024).

[25] Luciano Floridi. "Information ethics: On the philosophical foundation of computer ethics". In: *Ethics and Information Technology* 1.1 (1999), pp. 33–52. DOI: 10.1023/a:1010018611096.

[26] Luciano Floridi. "Soft ethics, the governance of the digital and the General Data Protection Regulation". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2133 (Oct. 2018), p. 20180081. DOI: 10.1098/rsta.2018.0081.

[27] Luciano Floridi and J.W. Sanders. "On the Morality of Artificial Agents". In: *Minds and Machines* 14.3 (Aug. 2004), pp. 349–379. DOI: 10.1023/b:mind.0000035461.63578.9d.

[28] Luciano Floridi and Mariarosaria Taddeo. "What is data ethics?" In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (Dec. 2016), p. 20160360. DOI: 10.1098/rsta.2016.0360.

[29] Google. *What is Big Data?* URL: https://cloud.google.com/learn/what-is-big-data (visited on 05/21/2024).

[30] Dirk Helbing et al. "Will Democracy Survive Big Data and Artificial Intelligence?" In: *Towards Digital Enlightenment*. Springer International Publishing, Aug. 2018, pp. 73–98. ISBN: 9783319908694. DOI: 10.1007/978-3-319-90869-4_7.

[31] Richard Herschel and Virginia M. Miori. "Ethics & Big Data". In: *Technology in Society* 49 (May 2017), pp. 31–36. DOI: 10.1016/j.techsoc.2017.03.003.

[32] Valentin Hofmann et al. *Dialect prejudice predicts AI decisions about people's character, employability, and criminality*. Preprint. 2024. DOI: 10.48550/ARXIV.2403.00742.

[33] Lanah Kammourieh et al. "Group Privacy in the Age of Big Data". In: *Group Privacy: New Challenges of Data Technologies*. Springer International Publishing AG, 2017. Chap. 3, pp. 37–66. ISBN: 978-3-319-46608-8.

[34] Alexandra L'Heureux et al. "Machine Learning With Big Data: Challenges and Approaches". In: *IEEE Access* 5 (2017), pp. 7776–7797. DOI: 10.1109/access.2017.2696365.

[35] Dave Lee. "Why Big Tech pays poor Kenyans to teach self-driving cars". In: *BBC* (2018). URL: https://www.bbc.com/news/technology-46055595 (visited on 06/09/2024).

[36] Heejun Lee and Chang-Hoan Cho. "Digital advertising: present and future prospects". In: *International Journal of Advertising* 39.3 (July 2019), pp. 332–341. DOI: 10.1080/02650487.2019.1642015.

[37] Luca Longo et al. “Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions”. In: *Information Fusion* 106 (June 2024), p. 102301. DOI: 10.1016/j.inffus.2024.102301.

[38] Alessandro Mantelero. “From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era”. In: *Group Privacy: New Challenges of Data Technologies*. Springer International Publishing AG, 2017. Chap. 8, pp. 139–158. ISBN: 978-3-319-46608-8.

[39] Dan McQuillan. *Resisting AI. An Anti-fascist Approach to Artificial Intelligence*. Bristol University Press, July 2022. ISBN: 978-1529213508.

[40] Milagros Miceli and Julian Posada. “The Data-Production Dispositif”. In: *Proceedings of the ACM on Human-Computer Interaction* 6.CSCW2 (Nov. 2022), pp. 1–37. DOI: 10.1145/3555561.

[41] United Nations. *Universal Declaration of Human Rights*. 1948. URL: https://www.un.org/en/about-us/universal-declaration-of-human-rights (visited on 05/28/2024).

[42] Helen Nissenbaum. “Accountability in a computerized society”. In: *Science and Engineering Ethics* 2.1 (Mar. 1996), pp. 25–42. DOI: 10.1007/bf02639315.

[43] Cathy O’Neil. *Weapons of math destruction. how big data increases inequality and threatens democracy*. Penguin Books, 2016. ISBN: 978-0-14-198541-1.

[44] Ziad Obermeyer et al. “Dissecting racial bias in an algorithm used to manage the health of populations”. In: *Science* 366.6464 (2019), pp. 447–453. DOI: 10.1126/science.aax2342.

[45] *Open letter to President Biden from tech workers in Kenya*. May 2024. URL: https://www.foxglove.org.uk/open-letter-to-president-biden-from-tech-workers-in-kenya/ (visited on 06/09/2024).

[46] Orestis Papakyriakopoulos, Juan Carlos Medina Serrano, and Simon Hegelich. “Political communication on social media: A tale of hyperactive users and bias in recommender systems”. In: *Online Social Networks and Media* 15 (Jan. 2020), p. 100058. DOI: 10.1016/j.osnem.2019.100058.

[47] Orestis Papakyriakopoulos et al. “Social media and microtargeting: Political data processing and the consequences for Germany”. In: *Big Data & Society* 5.2 (July 2018), p. 205395171881184. DOI: 10.1177/2053951718811844.

[48] European Parliament. *EU AI Act: first regulation on artificial intelligence*. June 2023. URL: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (visited on 06/16/2024).

[49] European Parliament. *Parliament adopts Platform Work Directive*. Apr. 2024. URL: https://www.europarl.europa.eu/news/en/press-room/20240419IPR20584/parliament-adopts-platform-work-directive (visited on 06/16/2024).

[50] Julian Posada. “The Future of Work Is Here: Toward a Comprehensive Approach to Artificial Intelligence and Labour”. In: *Ethics of AI in Context (2020)* (July 2020). DOI: 10.48550/ARXIV.2007.05843. arXiv: 2007.05843 [cs.CY].

[51] Markus Reuter and Daniel Leisegang. "Going Dark: EU States Push for Access to Encrypted Data and Increased Surveillance". In: *netzpolitik.org* (June 2024). URL: https://netzpolitik.org/2024/going-dark-eu-states-push-for-access-to-encrypted-data-and-increased-surveillance/ (visited on 06/14/2024).

[52] Annika Richterich. "How Data-Driven Research Fuelled the Cambridge Analytica Controversy". In: *Partecipazione e Conflitto* 11 (2018), pp. 528–543. DOI: 10.1285/I20356609V11I2P528.

[53] Till Speicher et al. "Potential for Discrimination in Online Targeted Advertising". In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Ed. by Sorelle A. Friedler and Christo Wilson. Vol. 81. Proceedings of Machine Learning Research. PMLR, 23–24 Feb 2018, pp. 5–19. URL: https://proceedings.mlr.press/v81/speicher18a.html (visited on 06/10/2024).

[54] Bernd Carsten Stahl, Job Timmermans, and Brent Daniel Mittelstadt. "The Ethics of Computing: A Survey of the Computing-Oriented Literature". In: *ACM Computing Surveys* 48.4 (Feb. 2016), pp. 1–38. DOI: 10.1145/2871196.

[55] Bernd Carsten Stahl and David Wright. "Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation". In: *IEEE Security & Privacy* 16.3 (May 2018), pp. 26–33. DOI: 10.1109/msp.2018.2701164.

[56] Ben M. Tappin et al. "Quantifying the potential persuasive returns to political microtargeting". In: *Proceedings of the National Academy of Sciences* 120.25 (June 2023). DOI: 10.1073/pnas.2216261120.

[57] Linnet Taylor, Luciano Floridi, and Bart van der Sloot. "Introduction: A New Perspective on Privacy". In: *Group Privacy: New Challenges of Data Technologies*. Springer International Publishing AG, 2017. Chap. 1, pp. 1–12. ISBN: 978-3-319-46608-8.

[58] Sherry Tiao. *What Is Big Data?* Mar. 2024. URL: https://www.oracle.com/big-data/what-is-big-data/ (visited on 05/21/2024).

[59] Paola Tubaro, Antonio A Casilli, and Marion Coville. "The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence". In: *Big Data & Society* 7.1 (Jan. 2020), p. 205395172091977. DOI: 10.1177/2053951720919776.

[60] Wendell Wallach. "Implementing moral decision making faculties in computers and robots". In: *AI & SOCIETY* 22.4 (Mar. 2007), pp. 463–475. DOI: 10.1007/s00146-007-0093-6.

[61] Andrej Zwitter. "Big Data ethics". In: *Big Data & Society* 1.2 (July 2014), p. 205395171455925. DOI: 10.1177/2053951714559253.