sciendo

Ethics & Bioethics (in Central Europe), 2025, 15 (1–2), 107–119 DOI:10.2478/ebce-2025-0002

Data Ethics: Issues related to data biases and the application of traditional ethical theories on AI Ethics

Anestis Karastergiou[1]

Abstract

The purpose of this article is to present and evaluate some issues related to data ethics in the era of artificial intelligence (AI). First, a philosophical definition of AI is attempted following the classic portrayal of AI presented by John Searle. This sets the frame for understanding AI and its basic components, machine learning, and big data. A distinction of the basic, traditional ethical theories (deontological ethics, utilitarianism, and virtue ethics) follows to draw an ethical schema via which to assess the discussion of data ethics in this paper. Then, the basic biases that are relevant to data gathering and processing are thoroughly presented with some significant examples to enhance understanding of these issues. After explaining the biases related to data, an assessment of how these could be eliminated follows based on the simple ethical schema presented at the initial stages of the article. None of these ethical theories suffice to eliminate data biases. However, their importance lies in the fact that they can be used as ethical methodological tools to assess data bias and understand the ways in which it may hinder the presumed objectivity of AI.

Keywords: ethics, AI ethics, data ethics, big data, machine learning, biases

Data ethics is a new and relatively controversial field. If the broader scientific mindset is taken into account, the reasons for this controversy seem obvious. To be specific, operating in the fields of Big Data, Machine learning, Deep learning, or generally in many advanced computational fields, has led scientists to believe that biases can be restricted to the structure of algorithms, the operations of neural networks or the interpretations of statisticians. From this point of view, data seem to be neutral and unbiased. It may come as no surprise to many that scientists try to solve real world problems by restructuring an algorithm to become as unbiased as possible. However, the main concern of this project is to find out if this way of thinking is actually the right one. Can the restructuring of an algorithm eliminate bias? Is it possible to get rid of biases by making sure that the interpretation of a neural network is unbiased? These questions and many more ultimately rest on our response to the overarching question; Can data be biased? This question will be explored in detail in this project. A short answer, contrary to what many might think, is yes. However, it remains to be shown why this is so. What are the nuances of the problem? Is it possible to structure data in a way that eliminates bias? A response to these questions may be much harder than it seems. Thus, in this paper data ethics related to the collection, accumulation, and usage of data will be explored. It will be shown how biases can actually emerge from data processing, while attempting to open the black box of data in weak, according to Searle, AI applications.

Defining AI from a philosophical perspective, an exercise in vanity

Not surprisingly, defining AI is a notoriously difficult task in which the present paper will endeavor briefly and with caution. Starting with John Searle's distinction between strong and weak AI, it is possible to set the appropriate frame in which to talk about AI and data ethics. According to the philosopher, for "weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool" (Searle, 1980, p. 417). He goes on to mention that "for example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion" (Searle, 1980, p. 417). In other words, AI is a tool we have in our hands and a very

[1] National and Kapodistrian University of Athens; ankarast@uoa.gr; ORCID: 0000-0002-0398-5613

powerful one for that matter. We construct the algorithms and the neural networks, we feed it data and it provides us with outputs which can vary according to its abilities to learn and adapt to new problems. It is quite obvious that this type of AI is closer to machine and deep learning instances. However, strong AI is on a whole separate level. As Searle puts it, "according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states" (Searle, 1980, p. 417). He continues "in strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations" (Searle, 1980, p. 417). To rephrase, the states of strong AI equal the 'mental' states of the mind, strong AI equals conscience. Let us not delve into this dubious philosophical issue, but focus on the basic assumption that will be used for this project. Strong AI is unattainable up to now and it seems that it will remain so for the near future, to say the least. However, weak AI already exists, it has many applications in our daily lives, and it is bound to expand in the near future. A reason for this is closely related to the main theme of the project at hand. Weak AI applications, machine, and deep learning instance need data, Big data. Therefore, data or Big data ethics are quite relevant when talking about machine and deep learning.

Data ethics: a simple definition and approach via classical ethics

When defining data ethics, one has to actually define two terms and then merge these definitions. What is the data we refer to? What is ethics? To begin with the first question, when we talk about data in this context, we mostly refer to Big data. It can be defined, according to the SAS (Statistical Analysis System) company, as "a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves" (Big Data: What it is and why it matters?, n.d.). In addition, Oracle's definition is as follows: "big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address business problems you wouldn't have been able to tackle before" (What Is Big Data? | Oracle, n.d.). One uptake from these definitions is, of course, that Big data consists of huge amounts of data which can be collected and analyzed in unprecedented levels using modern technology. The importance of Big data in AI applications can be stressed more by recalling the words of Reter Norvig, Google's director, who said "We don't have better algorithms. We just have more data" (Sanders, 2014, p. 7). In other words, data expansion is crucial to machine and deep learning.

Turning now to another very important aspect of the previous definitions. They both point to the use of data by organizations. Processing data can lead to better decisions, which is something obvious but not quite specific as 'better' implies the existence of a purpose and an agent. It is a better decision related to what goal and for whom? But SAS goes on to say that it is "what organizations do with the data that matters". This is where Ethics come into play. Use of Big data is not neutral, it is susceptible to numerous ethical issues as shall be seen. However, what is Ethics? It is not an easy question to answer and, obviously, if one attempts to do so most of the nuances related to the philosophical implications of the term will be left aside. However, that is exactly what will be done here in order to contextualize the discussion. The various ethical theories will not be defined, as machine and deep learning were not defined. Instead, an ethical schema will be created and it will be seen how it is relevant to Big data.

Supposing that there are three main branches of Ethics that are of interest in this paper. First, deontocracy or deontological ethics, with its most prominent representative to be Immanuel

Kant. He spoke of the moral duty to do the right thing and devised the categorical imperative, which simply refers to acting as if one's actions could become a universal moral law and to never use human beings merely as means but always as an end (Kant, 2012, pp. 21–24, 64–67, 74–75; Rohlf, 2023; Alexander & Moore, 2021). Applications of deontological ethics in AI have been proposed in the literature, as Kant's principles could be used to develop a set of rules that guide machine learning, from the accumulation of data to the structure of the algorithm (Hooker & Kim, 2018; Prabhumoye et al., 2021). Then, there is consequentialism and mostly utilitarian ethics, a very influential branch of philosophy. Its most prominent representatives are the philosophers Jeremy Bentham and John Stuart-Mill. Simply put, according to utilitarianism, the right thing to do is what maximizes happiness and minimizes pain (see: Driver, 2022; Crimmins, 2023; Sinnott-Armstrong, 2023). Bentham tried to quantify the maximization of happiness or benefit for the majority of people via his felicific calculus, an algorithm that could be used to find what maximizes happiness and at the same minimizes pain (Bentham, 2012, pp. 25–33). However, due to the subjective nature of the notion of happiness and the experience of pain, this attempt was heavily criticized (Crimmins, 2023). Even from early on, John Stuart-Mill recognized the existence of higher and lower pleasures, as opposed to Bentham's quantification (Brink, 2022). Utilitarian principles have been used to assess ethical concerns regarding AI, even if sometimes the need to transform these principles to the reasons for the decision of choosing the act instead of relying solely on its consequences (Roff, 2020). Finally, the simple ethical schema used here is completed with virtue ethics. With the ancient Greek philosopher Aristotle as the most prominent representative, virtue ethics focus on a person's character to assess morality (Hursthouse & Pettigrove, 2023). For Aristotle, virtue is a habit (*hexis* or *ἕξις*) or a disposition according to most English translations of the ancient Greek texts (Kraut, 2022), meaning that one becomes a virtuous individual by means of acting virtuously (Shields, 2023). Virtue ethics have been used in research on AI ethics mostly to refer to issues regarding the human agent in the era of AI, which is, of course, a very important aspect for data ethics (Bauer, 2020; Farina et al., 2022; Hagendorff, 2022).

One can imagine a scenario related to AI, where these ethical theories could be applied and could lead to very different interpretations regarding the ethical nature of the act. Suppose one is a machine learning engineer and attempts to train a Large Language Model, a very sophisticated neural network. This model is a black box for them as they do not have *a priori* knowledge of the outcomes given the input data. However, they have to account for the data inserted into the machine and faces the dilemma, should they try to do everything they can to clean the data and account for any sort of bias beforehand? To make matters worse, this model will be used to evaluate the ability of people to gain access to medical treatment at an international level. Applying deontological ethics, they should try to make sure that the data is by no means biased and that everyone has an equal say, i.e., everyone's data is included in the process and treated fairly. First, this is mandated by the categorical imperative to never treat anyone merely as a means and to act as if one's actions could become a universal moral law. However, this is virtually unattainable here and may result in inaction. Applying the utilitarian ethical methodology here could be tricky as this consequentialist approach implies that one can assume the outcome with some degree of certainty. Nevertheless, Betham's felicific calculus could surmise that acting now on the data they have access to is the best for most people, as the model will learn and be implemented fast giving access to medical treatment to more people. In other words, the short-term benefit would be greater if the model is implemented fast, even if it means that some proportion of people is treated unfairly. Virtue ethics would mandate that the machine learning engineer should build their virtuous character by acting virtuously. In this case, they should try to find all the data and account for biases even if it means postponing the use of the model.

Evidently, the above ethical schema seems a bit naïve. However, it will be used to talk about the ethics of data usage and the moral agents implied in those. Echoes of these branches of ethics are also present in business ethics, of course. However, another distinction relevant mostly to this field is that of the stockholder, stakeholder, and social ethical theory. To understand this distinction, one can imagine circles of ethical proximity, much like Singer's ethical kinship ones (Singer, 2017). In the first case, the stockholders are the sole moral agents. The second expands to stakeholders who can be various ethical players. Finally, the third talks about a business's ethical responsibility towards society. In a sense, we are all stakeholders in this last one. In what follows the ethical issues related to data will be defined, mostly the biases that occur and apply these ethical frameworks in order to propose ways to eliminate data biases (Moriarty, 2021). Applying the concept of business ethics to the machine learning engineer above, they should have to choose between these theories to find the most appropriate one to act upon. For instance, they may think that the best approach for society on the whole would be to try and find all the data possible. However, according to the stakeholder theory and more so in the case of the stockholder theory, this is not necessary.

Bias in the collection of data

Statisticians call it "selection bias", and it can actually occur in any type of research, be it quantitative or qualitative. Simply enough, here the partially collected, biased, data is referred to. Having said that, does this partial selection take place? Is it just a reflection of the researcher's biases? Well, it could definitely be so. No matter how impartial a researcher aspires to be, it is quite common to fall victim to their subconscious, probably socially constructed, biases. To elaborate; on a high level racial, gender, or ethnicity biases come into play without the researcher even noticing that they are viewing data through already hardcoded prisms. These biases can be implemented both in the collection of data and the interpretation of the results. For instance, let us consider collecting data on math test scores. There may be a tendency to stratify the results through the prism of race, i.e., Asian students would be expected to have better scores, therefore any small deviation from the expected outcome, and it may also, happen that the results, the output of the survey, are interpreted with the same prism, falling victim to confirmation bias. Although knowing these biases, they can be accounted for, people and researchers are no exception and susceptible to the bias blind spot, i.e., one is fast to detect bias in others but not on oneself with profound effects on everyday life and scientific research (Scopelliti et al., 2015). What about selection biases which are independent from the user? For example, one may have more records of urban killing in the US rather than rural ones (Price & Ball, 2014). It does not mean that the rate of these killings is lower in rural areas, but they are just not recorded. In collecting these records, one is bound to omit those events for which there is no data, thus one is *de facto* partial in a survey and the conclusions that may follow. Another example would be big versus small events. This type of selection bias is called "event size bias" (Price & Ball, 2014). Big events are more likely to be recorded than small events. Thus, one has a huge database of big to mega events, but a relatively small one of those events which are small, local for instance. Yet again, the bias lies in the structure of collecting data.

Bias in structured data

Distinguishing between unstructured and structured data is crucial for machine learning applications, among others. Unstructured is any kind of data, which is stored in its native format, without any further processing. However, structured data can be fully or semi structured. A semi structured data set is one that is organized to some extent, yet some more processing is necessary. A fully structured data set is one that is highly organized, factual, and to the point. It can be assumed that the biases which occur during the collection of data are mostly related to unstructured data. When it comes to structured data sets, these may be masked,

making it very difficult to actually understand their existence and impact. One could see this process as a kind of black boxing. To elaborate on this, through data extraction unstructured data transforms into a structured form. Schematically presented, this occurs through segmentation, i.e., finding the starting and ending boundaries of the text snippets that will fill a database field", classification, i.e., determining "which database field is the correct destination for each text segment", association which "determines which fields belong together in the same record", normalization, i.e., putting "information in a standard format in which it can be reliably compared" and deduplication which "collapses redundant information so you don't get duplicate records in your database" (McCallum, 2005, pp. 49–51). Thus, the structured data set appears neat and tidy. Various gender, race, and so on biases which were probably embedded in unstructured data seem to be eliminated. However, is this so? Apparently not. Structuring data does not actually eliminate biases. It masks them through a process of black boxing. For instance, data collected through a racial prism now appears to be normalized so as to be reliably compared with other data.

Bias in the interpretation of data

The nuances of data interpretation are probably too many to be presented thoroughly in such a small space. However, the present paper will try to focus on primary issues related to interpretation and its potential biases (Kaptchuk, 2003). Mostly when talking about interpretation, to the evaluation of the results are referred to and to be more specific, interpreting data coming will be talked about with the outputs of machine and deep learning applications in mind.

To begin with, some trivial, yet important cases of biased data interpretation will be presented. Yet again, confirmation bias is probably too hard to avoid. Researchers may have many different prisms from which they view data. Although here race and gender prisms will not be discussed. Assume that these stereotypes are accounted for. However, researchers still have expectations about the output of their research. Sometimes, following the Popperian methodology of falsification may be too an idealistic imperative. Pressing deadlines, necessity of results in order to secure research funding, and many more are factors which contribute to the confirmation bias kicking in. The sketch in the image below is reflective of what is being discussed here. It is not uncommon or even uncalled for in some cases, to 'cook' results in order to confirm your hypotheses or to actually make them 'look right'. If they are not, you just need to 'stir the pile' again and again until you reach the expected outcomes.

Consider the case of a machine learning application on a high level, just for the purposes of this mind experiment. How does it work? We program it using algorithms; thus, we create neural networks (let's say a type of algorithm) which may have multiple layers to foster the various modes of machine learning. Then we constantly feed the machine with data, and it gives us outputs. Depending on the type of learning, supervised, unsupervised, semi-supervised, or reinforcement learning, and the capabilities of the machine, (deep learning instances are generally more versatile in learning), feeding and labelling data increases the effectiveness of the machine, thus giving us better, that is more accurate, results (Structured vs. Unstructured Data, 2018). Now, how could it be for the machine to be wrong and how can we know it? First, increasing machine complexity has an impact on transparency, creating a black box through a process of unknowing. Simply put, we are not always cognizant of how the machine reaches a certain output from a given input. There may be many variables which contribute to the result, but what is the actual significance of its one? Increasing complexity and machine opaqueness leads to a conundrum, should we continue to 'stir the pile' until the machine learns what we want it to, until the machine conforms to our expectations, or should we continue testing the machine and accept the results which may be far from our expectations? If we want to open the black box of machine learning, I guess we have to choose between 'guiding' this unboxing procedure from an output point of view, thus dominating the procedure implementing our

expectation bias or we can let it reveal itself in these outputs. To do the latter, one has to apply immense hours of work and be impartial to one's interpretations. One way to achieve it though would be to give open access to the algorithm, to the machine, so that anyone could experiment with it. A type of citizen science perspective which could enhance transparency without limiting complexity at the same time. Impartiality is, of course, virtually unachievable, but partiality could be diffused by letting more people from diverse backgrounds into the interpretation of the results, i.e., achieving relative impartiality in hindsight (Pasquale, 2015, pp. 72–80).

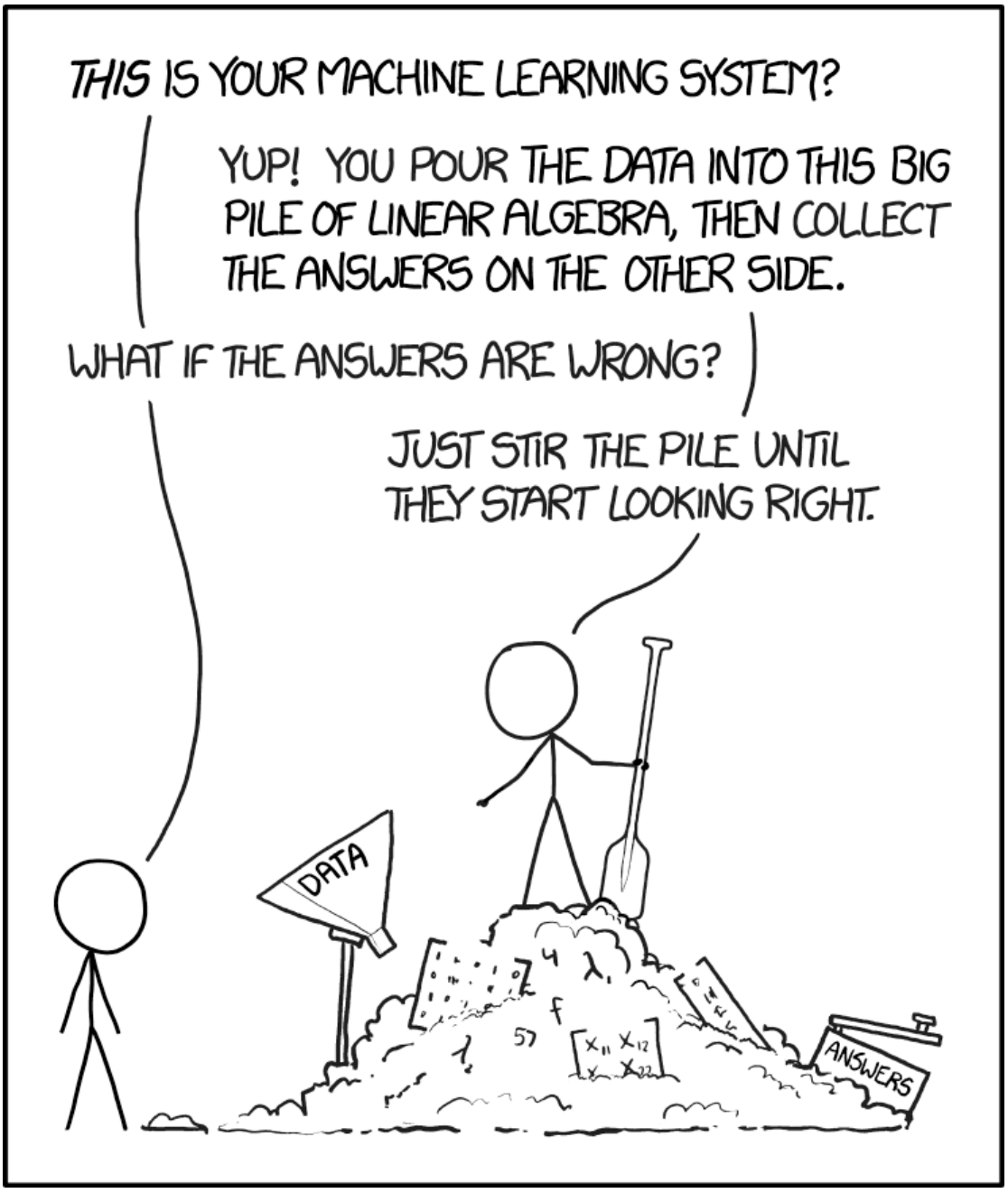

Image retrieved from: https://towardsdatascience.com/survey-d4f168791e57.

Another aspect of the interpretation bias should be pointed out in order to see the particular nuances related to machine learning (Kaptchuk, 2003). When 'stirring the pile' one may 'cook' the inputs as well as the outputs. More specifically, if one finds that a certain input or labelling the input in a certain way may influence the outputs towards the expected results, they may tend to do so for the reasons presented above. This is an application of the confirmation,

interpretation bias from the endpoint to the beginning of the machine learning process. Moreover, the very algorithms may be biased towards a specific result. However, the specifics of this will not be elaborated upon, because how data becomes biased is of more interest here and the unknowing procedures, the black boxing related to it. In the next section, some cases related to biases in data usage from Amazon's machine learning applications and social media algorithms will briefly be mentioned. It is obvious up to now, that many aspects of data biases are left aside in this project. Nevertheless, in the next few sections an attempt will be made to include a broader range of biases related to Big data.

Bias in the usage and the creation of data

From the sections above, it can be seen that the usage of data, be it collection or interpretation of it has already been talked about. However, in this section reference will be made to the creation and usage of data in the case of social media and Amazon's face recognition algorithms. The difference here lies in how data is actually created and used in real time. The focus will be on biases which are embedded in the very process of creating data and how relevant some known biases are in the usage of data to create a face recognition machine learning application.

It is well documented that only few social media users create most of its contents (Baeza-Yates, 2018). Starting with the response bias, these few, probably celebrities or influencers, among others, are more eager to respond and create a vast amount of content be it visual or textual. This response bias is also related to the omitted voices bias, which, as its name suggests, points to the fact that a significant percentage of users do not contribute in a significant way to the creation of data. Nowadays Big data created in social media is already vast and it keeps increasing exponentially. Moreover, it has many and diverse usages. Data coming from social media users can be utilized to create personalized marketing content, to segment the market according to certain characteristics, to optimize research results and many more. Knowing that the above biases are embedded in the very process of creating this type of data and that they actually affect usage is important in order to unpack the black box of creating data in this case. However, how is it a black box? The reason is simple enough and quite obvious. Having a very high number of users, social media are perceived as instruments of democratization as they can give voice, accessibly and cheaply to every person in the world. However, political, economic, and social factors actually hinder such an aspiration. Many countries and, of course, a large number of people do not have access to the technological means necessary to use social media. Additionally, some countries have policies which ban its citizens from using social media. Even among social media users, many are banned due to censorship issues or not conforming with the rules of conduct. From those left, it has already been pointed out that only a few have a disproportionate contribution to the creation of content. In this sense, the black boxing of media content creation emerges by omission, i.e., despite social media aspirations for openness and inclusion, many people are left out of its operations (Pasquale, 2015, pp. 72–96).

To make matters worse, in a sense, age, gender, and race representation in social media is quite restricted. Older people are largely excluded, willingly or due to lack of computer literacy, from most social media platforms (Park, 2012). In the middle age range, most people use both LinkedIn and Facebook, but younger people show a preference for Facebook and Instagram, among others (Hargittai, 2020). The level of education and income is also a factor which may determine the preference of a specific social media platform. Regarding Facebook, those who make less are more likely to use it and the level of education does not seem to play a very important role (Hargittai, 2020). However, these change in the case of LinkedIn and Twitter, among others. Moreover, in Hargittai's study it was shown that women are much more likely to use Facebook, with all other variables controlled for (Hargittai, 2020). One could go on in this vein, but this brief presentation has already given the schema to acknowledge how various

biases come into play into data creation. One could also point out that social media users are more eager to present themselves in a more sophisticated, unrealistic way in order to be more likable. Thus, they filter their posts and are generally biased towards their actions in the virtual world.

All the above and many more that could be added, contribute to what can be called here bias in the creation and usage of data. Up to now, the focus has been on the creation of data. Now, a brief example from the usage from Amazon's Face Recognition tech will be taken (Spichkova et al., 2019). Notorious for its racial bias and, specifically, for its anti-black bias, this technology was withdrawn from police use after a relatively short period of use and because of the social upheaval against it (Williams, 2020). Using Deep Vision AI, this technology was quick to analyze pictures and identify faces (Spichkova et al., 2019). What it actually did was to identify disproportionally more black than white people as criminal perpetrators (Williams, 2020). To get the relative schema, a racial bias against black people which was and still is largely embedded in US society was reflected in police arrests. Of course, discrimination against black people has affected their economic and social status, thus more black people were likely to occupy poor households. To generalize, crime and namely petty criminal offences are also related to economic status. Thus, police arrests in the US were disproportionally more for black people. Amazon's Recognition used police databases to feed its machine, deep in this case, learning technology. As most criminal investigators know, one of the best, if not the sole reliable factor for predicting future criminal offences is past offences. Therefore, a racial bias can be seen here which was embedded in the data feeding the machine and followingly biased usage of this data as its results confirmed the expectation of prejudiced, even subconscious, policing. Social upheaval made the opening of this black box a necessity and finally a reality. Yet, it was specific and relatively easy to pinpoint. There are many other black boxes related to AI applications, as we have seen somewhat already, which are hard to find and much more so to pry them open.

Understanding bias

Understanding bias in Big data and AI is crucial for various reasons. It challenges the interpretation of a neutral machine and the dogmatic belief that using data and technology for decision making is *de jure* better and impartial. The response bias, omitted user bias, how the confirmation bias may kick in, selection bias, societal bias in the case of social media content creation and the interpretation bias have already been presented. Drift bias, which is created by the system drift, the transformation of the system due to updates in its programs and, of course, its algorithms that changes how the users interact with it could also be mentioned here. Apart from these types of biases, which are somewhat specific to data and AI issues, many of the cognitive biases are applicable here. One of them was the confirmation bias, for instance.

However, it is more important to explain a basic tenant of the present discussion rather than trying to present a catalogue of these biases and how they may or may not relate to data and AI. This is none other than the man behind the machine. It may be naively put here, but the metaphor of the mechanical Turk is particularly relevant here and it can be a bridge to the ethics discussion which will follow. Constructed by Wolfgang von Kempelen to impress Empress Maria Theresa of Austria, the mechanical Turk was an automaton which was capable of playing chess, albeit only in appearance. In most cases, behind the machine and the orientally dressed homunculus, was a dwarf sized human being who actually played the game (Standage, 2002, pp. 15–99). I find this metaphor especially revealing as technology, from Descartes and the automatons of Vesalius, and mostly modern technology is not regarded as a tool, but as a dehumanizing and mechanizing cultural instrument. The proximity to the human agent is broadened to a large extent. As the famous saying goes, 'out of sight, out of mind'. Taking the human agent out of the picture makes the machine appear independent, absolutely autonomous.

If ethics is a human issue, machine ethics would be an oxymoron. However, if a machine is a tool, even a very sophisticated one, at the disposal of humans, then Data and AI ethics are now more relevant than ever. This is why we should start from understanding and accounting for the biases which come into play in AI and Big Data and then proceed to an application of ethical theories to contextualize and eliminate them, should there be a way to do so. I strongly believe that machines, even modern AI applications, are another aspect of what Hegel called the ruse of reason (Maybee, 2020). To put it as naively as possible, they are tools in human disposal, means to achieve one's goals. If reason, in the Hegelian sense, is the driving force of history, intentionality, in the broader of senses philosophically, but very specifically in some cases of machine construction, is projected on the machine.

Eliminating bias

Evidently, the first step in eliminating bias is to find it. However, even if one can specifically pinpoint the bias and analyze it, it is not always obvious how it can be eliminated or why it should be so. One may wonder why I even posed the second question. The mandate to be impartial, i.e., eliminate biases, is considered self-evident. No researcher would proudly say "I am biased and it's ok". Maybe so. Although, this comes from scientific deontology, the daughter of Kant's deontocracy in a broad sense. It is the duty of the researcher to be unbiased or to check for biases. Therefore, even in the simplest of all examples, one can see that behind what is called scientific methodology lies a moral imperative, an ethical theory. Kant's deontocracy, though, and its categorical imperative is too hard to attain in most cases (Johnson & Cureton, 2021). This can be applied on the impartiality mandate. Is it possible for everyone and at any time to be impartial? Probably not. I would suggest going beyond this and apply other ethical theories to account for and eliminate biases.

One of the consequentialist concepts, utilitarianism, is the weapon of choice for business ethics as it provides the framework for a cost benefit analysis. The goal here is to maximize benefit and minimize damage for the largest amount of stakeholders possible. Due to the shortage of space, how cynical this theory can be will be presented. Consider the case of Amazon's Recognition. Now let it be assumed that pointing to black perpetrators and arresting them is a benefit for most people in a white dominated society. Black people are a minority, which is not actually realistic for the US, and, thus, the damage inflicted on them is marginal in comparison to the benefit derived for society. Maybe this is too extreme an example, but I think that applying Bentham's felicific calculus here could have extremely inhumane results, even 'unethical' in the sense of the word in common parlance (Sinnott-Armstrong, 2023).

I was fast to eliminate two basic ethical theories, but I would not want to do so entirely. They are still important as frameworks and can give us significant ethical insights. Let me proceed now to another ethical theory or frame, virtue ethics. A daughter of Aristotle, the importance of virtue ethics was renewed as it fits effectively the liberal, individualistic sociopolitical regime of the so-called Western world. Not to mention that the image of virtuous leader is not so much an Aristotelian concept as a Confucian one. According to this theory, a virtuous person is constituted through acting ethically, virtue is a disposition (*ἕξις*). One should aspire to be a virtuous person and, thus, one has to act ethically to do so (Hursthouse & Pettigrove, 2023). However, choosing the right thing to do, choosing to do good, is not always evident and it is highly context specific. To be more precise, in the context of business, AI, and data ethics, the moral agent needs to be pointed to and the recipient of the act in order to actually do good (Farina et al., 2022).

For the latter, stockholder, stakeholder, and social theory can be used. Depending on which one is used, the moral agents can be specified, and one can understand the intentionality of the 'good' action. For instance, suppose one takes the stockholder theory. In this case, the good act is relative to the good perceived only by those people who actually hold stocks in the company.

In the case of data, what is good is what lies within the interests of, let's say, the data analyst. Stakeholder theory expands the group of interest to anyone who has some connection to the data. Social theory or social contract theory supposes that data gathering and usage should be accountable to society. In other words, data processing should be focused on increasing benefit for the society as a whole (Moriarty, 2021). Taking the last ethical theory into account, one can see that with society's interest in view, including of course minorities, the right thing to do has actually gained the necessary perspective. Adding to this, in order to answer the question "who decides what is right for the society?", human rights theory can actually be used as a lens through which we assign value to our actions. In a democratic society, the right to equal representation, justice, doing no unnecessary harm, and the like should guide our actions. Processing data is no exception to this (Nickel, 2019).

However, the above presentation draws on some simplified views of the very complex ethical landscape regarding AI and data ethics. The ethical guidelines that are proposed to evaluate data and find the moral agent that should act ethically are inefficient in most cases as any deviation from them does not actually produce any results (Hagendorff, 2020). Moreover, philosophical ethics may provide some very useful guidelines, but they do not suffice to account for the complexities regarding data ethics and, generally, ethics in the era of AI. The need for an interdisciplinary approach is necessary, along with politicizing AI ethics to account for the nuances of AI applications in various circumstances (van Maanen, 2022). In this paper, the discussion about AI ethics may draw mostly from the philosophy of ethics, but it also incorporates psychology, by discussing biases, and the Science and Technology Studies (STS) by addressing the black box concept and tool to assess the relationship between philosophy and technology. The ethical frameworks taken from classical ethical theories presented here can be used as methodological tools to address issues related to data ethics, something that can be extremely useful in broadening our understanding of the problems that relate to modern technology (Markham, Tiidenberg & Herman, 2018).

Conclusion

Summarizing the present brief presentation of some issues related to data ethics, biases in AI applications, and ethical issues concerning data collection and usage, it is quite obvious that this is a new and expanding field where more research should be done concerning ethical issues particular to it. There is no neutral unbiased data and using machine learning applications to improve impartiality in decision making is not a panacea. Biases may be embedded in data processing and sometimes they are so hard coded in society and consequently scientific practice that they may be too hard to find and account for. Opening the black box of data and AI should be the primary focus, but it should be done through a data ethics perspective. All of the ethical theories presented very briefly above can provide us with useful frameworks to achieve our goal and a combination of them would be very helpful to move from increasing transparency to equal and just use of data.

Moreover, data ethics should implement a multidisciplinary approach with some of the ethical theories presented here as the backbone of the ethical frameworks to be implemented. STS methodology is already incorporated in the discussion of data ethics (Heeney, 2017) and rightly so as it provides a useful lens to understand the technological black box that needs to be made visible for an ethical assessment of technology. Additionally, drawing from psychology, this paper focused on the biases related to data gathering and processing as they form a very important aspect of the discussion regarding data ethics. Finally, more studies about data ethics should try to incorporate different views from various disciplines, such as philosophy of ethics, philosophy of technology, STS, psychology, sociology, anthropology, political science, computer science, and many more.

References

ALEXANDER, L. & MOORE, M. (2021): Deontological ethics. In: E. N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. [online] [Retrieved April 13, 2023]. Available at: https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/

BAEZA-YATES, R. (2018): Bias on the Web. In: *Communications of the ACM*, 61(6), pp. 54–61.

BAUER, W. A. (2020): Virtuous vs. utilitarian artificial moral agents. In: *AI & SOCIETY*, 35(1), pp. 263–271. [online] [Retrieved 20 May, 2024]. Available at: https://doi.org/10.1007/s00146-018-0871-3

BENTHAM, J. (2012): *An introduction to the principles of morals and legislation*. New York: Dover Publications Inc.

BIG DATA: WHAT IT IS AND WHY IT MATTERS. (n.d.): [online] [Retrieved April 14, 2021]. Available at: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

BRINK, D. (2022): Mill's moral and political philosophy. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Fall 2022). Metaphysics Research Lab, Stanford University. [online] [Retrieved May 20, 2024]. Available at: https://plato.stanford.edu/archives/fall2022/entries/mill-moral-political/

CRIMMINS, J. E. (2023): Jeremy Bentham. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University. [online] [Retrieved April 18, 2023]. Available at: https://plato.stanford.edu/archives/fall2023/entries/bentham/

DRIVER, J. (2022): The history of utilitarianism. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University. [online] [Retrieved March 13, 2023]. Available at: https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history/

FARINA, M., ZHDANOV, P., KARIMOV, A. & LAVAZZA, A. (2022): AI and society: A virtue ethics approach. In: *AI & SOCIETY*. [online] [Retrieved 20 May, 2024]. Available at: https://doi.org/10.1007/s00146-022-01545-5

HAGENDORFF, T. (2020): The ethics of AI ethics: An evaluation of guidelines. In: *Minds and Machines*, 30(1), pp. 99–120. [online] [Retrieved January 23, 2023]. Available at: https://doi.org/10.1007/s11023-020-09517-8

HAGENDORFF, T. (2022): A virtue-based framework to support putting AI ethics into practice. In: *Philosophy & Technology*, 35(3), p. 55. [online] [Retrieved 20 May, 2024]. Available at: https://doi.org/10.1007/s13347-022-00553-z

HARGITTAI, E. (2020): Potential biases in Big Data: Omitted voices on social media. In: *Social Science Computer Review*, 38, pp. 10–24. [online] [Retrieved October 21, 2023]. Available at: https://doi.org/10.1177/0894439318788322

HEENEY, C. (2017): An "ethical moment" in data sharing. In: *Science, Technology, & Human Values*, 42(1), pp. 3–28. [online] [Retrieved April 17, 2024]. Available at: https://doi.org/10.1177/0162243916648220

HOOKER, J. N. & KIM, T. W. N. (2018): Toward non-intuition-based machine and artificial intelligence ethics: A deontological approach based on modal logic. In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 130–136. [online] [Retrieved May 15, 2024]. Available at: https://doi.org/10.1145/3278721.3278753

HURSTHOUSE, R. & PETTIGROVE, G. (2023): *Virtue ethics*. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University. [online] [Retrieved April 13, 2024]. Available at: https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/

JOHNSON, R. & CURETON, A. (2021): Kant's moral philosophy. In: E. N. Zalta (ed.): *The*

Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University. [online] [Retrieved September 23, 2023]. Available at: https://plato.stanford.edu/entries/kant-moral/
KANT, I. (1785/2012): Groundwork of the metaphysics of morals, eds. M. J. Gregor & J. Timmermann, rev. ed. Cambridge: Cambridge University Press.
KAPTCHUK, T. J. (2003): Effect of interpretive bias on research evidence. In: *BMJ*, 326(7404), pp. 1453–1455.
KRAUT, R. (2022): Aristotle's ethics. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Fall 2022*)*. Metaphysics Research Lab, Stanford University. [online] [Retrieved May 20, 2024]. Available at: https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/
MARKHAM, A. N., TIIDENBERG, K. & HERMAN, A. (2018): Ethics as methods: Doing ethics in the era of Big Data research—Introduction. In: *Social media + society*, 4(3), 2056305118784502. [online] [Retrieved April 24, 2024]. Available at: https://doi.org/10.1177/2056305118784502
MAYBEE, J. E. (2020): Hegel's dialectics. In: E. N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. [online] [Retrieved October 12, 2023]. Available at: https://plato.stanford.edu/entries/hegel-dialectics/
MCCALLUM, A. (2005): Information extraction: Distilling structured data from unstructured text. In: *Queue*, 3(9), pp. 48–57. [online] [Retrieved October 15, 2023]. Available at: https://doi.org/10.1145/1105664.1105679
MORIARTY, J. (2021): Business ethics. In: E. N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. [online] [Retrieved June 13, 2023]. Available at: https://plato.stanford.edu/archives/fall2021/entries/ethics-business/
NICKEL, J. (2019): Human rights. In: E. N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. [online] [Retrieved October 18, 2023]. Available at: https://plato.stanford.edu/entries/rights-human/
PARK, S. (2012): Dimensions of digital media literacy and the relationship with social exclusion. In: *Media International Australia*, 142, pp. 87–100. [online] [Retrieved November 4, 2023]. Available at: https://doi.org/10.1177/1329878X1214200111
PASQUALE, F. (2015): *The black box society*. Cambridge, MT: Harvard University Press.
PRABHUMOYE, S., BOLDT, B., SALAKHUTDINOV, R. & BLACK, A. W. (2021): Case Study: Deontological ethics in NLP (arXiv:2010.04658). arXiv. [online] [Retrieved May 20, 2024]. Available at: https://doi.org/10.48550/arXiv.2010.04658
PRICE, M. & BALL, P. (2014): Big Data, selection bias, and the statistical patterns of mortality in conflict. In: *SAIS Review of International Affairs*, 34(1), pp. 9–20. [online] [Retrieved November 7, 2023]. Available at: https://doi.org/10.1353/sais.2014.0010
ROFF, H. M. (2020): Expected utilitarianism (arXiv:2008.07321). arXiv. [online] [Retrieved May 20, 2024]. Available at: https://doi.org/10.48550/arXiv.2008.07321
ROHLF, M. (2023): Immanuel Kant. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University. [online] [retrieved March 13, 2024]. Available at https://plato.stanford.edu/archives/fall2023/entries/kant/
SANDERS, N. R. (2014): *Big Data driven supply chain management: A framework for implementing analytics and turning information into intelligence.* London: Pearson Education Limited.
SCOPELLITI, I., MOREWEDGE, C. K., MCCORMICK, E., MIN, H. L., LEBRECHT, S. & KASSAM, K. S. (2015): Bias blind spot: Structure, measurement, and consequences. In: *Management Science, 61*, pp. 2468–2486. [online] [Retrieved November 9, 2023]. Available at: https://doi.org/10.1287/mnsc.2014.2096

SEARLE, J. R. (1980): Minds, brains, and programs. In: *Behavioral and Brain Sciences*, 3, pp. 417–424. [online] [Retrieved November 3, 2023]. Available at: https://doi.org/10.1017/S0140525X00005756
SHIELDS, C. (2023): Aristotle. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University. [online] [Retrieved April 13, 2024]. Available at: https://plato.stanford.edu/archives/win2023/entries/aristotle/
SINGER, P. (2017): *Applied Ethics. A Multicultural Approach: Sixth Edition*. London: Taylor and Francis.
SINNOTT-ARMSTRONG, W. (2023): Consequentialism. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University. [online] [Retrieved April 28, 2024]. Available at: https://plato.stanford.edu/archives/win2023/entries/consequentialism/
SPICHKOVA, M., VAN ZYL, J., SACHDEV, S., BHARDWAJ, A. & DESAI, N. (2019): *Easy mobile meter reading for non-smart meters: Comparison of AWS rekognition and Google cloud vision approaches.* arXiv:1910.12617 [cs].
STANDAGE, T. (2002): *The Turk: The life and times of the famous eighteenth-century chess-playing machine.* New York: Walker & Co.
STRUCTURED VS. UNSTRUCTURED DATA, (2018): *Datamation*. [online] [Retrieved April 12, 2023]. Available at: https://www.datamation.com/big-data/structured-vs-unstructured-data/
VAN MAANEN, G. (2022): AI ethics, ethics washing, and the need to politicize data ethics. In: *Digital Society*, 1(2), p. 9. [online] [Retrieved March 28, 2024]. Available at: https://doi.org/10.1007/s44206-022-00013-3
WHAT IS BIG DATA? | ORACLE. (n.d): [online] [Retrieved April 15, 2023]. Available at: https://www.oracle.com/big-data/what-is-big-data/
WILLIAMS, D. P. (2020): Fitting the description: Historical and sociotechnical elements of facial recognition and anti-black surveillance. In: *Journal of Responsible Innovation*, 7, pp. 74–83. [online] [Retrieved August 24, 2022]. Available at: https://doi.org/10.1080/23299460.2020.1831365