Ricky Li

Best Dissertation Prize Winner

MSc Political Theory 2018-9

LSE Department of Government

gov.msc@lse.ac.uk

Data Ownership, Fraud and the Tainted AI:

On the Unjust Means in the Development of

Artificial Intelligence

A dissertation submitted by Candidate to the Department of Government, the London School of Economics and Political Science, in part completion of the requirement for the MSc in Political Theory.

August 2019

Word count: 10379

Table of Contents

Abstract

As in most cases of production, a developer is the de facto owner of the Artificial Intelligence (AI) that she creates using her intellectual labour and other necessary means of production. The ownership of one's creation is justified by the prominent ideal of workmanship; the workmanship ideal claims that through the process of mixing a person's productive capacities with other justly acquired resources, that person is the legitimate owner of the output of the productive process. In addition, throughout the process of developing an AI system, a vast amount of data is required as an input, which is used to train the AI and thus improve its competency. However, if such data were not previously owned by the developer of the AI, and if the means of acquiring the data was unjust, then the AI developer's ownership of its creation is tainted and cannot be vindicated by the workmanship ideal. This dissertation is an inquiry into the injustice in the process of AI production and the plausibility of different approaches to rectify this injustice.

Introduction

Let us first consider the following cases regarding the capability of Artificial Intelligence (AI).

1) AlphaGo is an AI algorithm designed to master the strategic game of Go, through the computational technology of a ‘value network’, which trains the algorithm using ‘a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play’ (Silver et al., 2016). AlphaGo was able to defeat the human European Go champion by five games to zero.

2) ‘… half the activities people are paid to do globally could theoretically be automated using currently demonstrated [AI] technologies’ (Manyika et al., 2017, p.2).

3) According to a report by The Guardian, during the 2016 United Kingdom European referendum, the Leave. EU Campaign utilised the power of AI in targeting individuals to gain more supports. Therefore, as Arron Bank, the co-founder of the Leave.EU Campaign, claimed: ‘AI won it for leave’. (Arron Banks, cited in Cadwalladr, 2017).

While case 1 demonstrates the approach by which an AI is developed and its subsequent capability to perform the assigned task at the highest competency level, cases 2 and 3 provide a glimpse of the potential impacts a well-trained AI can have on society.

Given the enormous potential power of AI technology, various thinkers have weighed in on the debates surrounding its development. In general, there are two distinct views regarding the subject of AI: the alarming view and the optimistic view. The alarming view urges people to stay vigilant in the pursuit of progress in AI technology as its uncritical advancement will be harmful to human society. This view can have at least two forms. One is that the alarming view argues that the rapid development of AI could pose a threat to the survival of humanity. Frischmann and Selinger (2018) argue that a society with increasing applications of AI technologies could gradually turn humans into programmable machines, while Bostrom (2014) hints at the possibility of the emergence of a powerful system of superintelligence that, instead of being controlled by humans, could replace humans as the most dominant species on earth. The second form of the alarming view is cautious about the unrestricted development of AI since, according to cases 2 and 3, AI could bring an enormous amount of social resources and political power to its owners. Therefore, an unregulated AI distribution undermines the feature of reciprocity in a liberal society and renders society unstable (Cohen, 1995, p.25; Rawls, 2001, pp.123, 138).

Conversely, the optimistic view indicates the potential Pareto improvement in people's living standards in a state of affairs in which AI systems are integrated into social lives. For example, an AI-fuelled automation's ability to replace mind-numbing jobs and the possible implications of shorter working hours should be welcomed by welfare economists, as well as by philosophers who value the Aristotelian ideal of meaningful work or who perceive the distribution of free time as a demand of justice (Murphy, 1993; Rose, 2013). Moreover, proponents of the optimistic view are faced with a dilemma: on the one hand, they must accept that some forms of incentive to the developers of AI are beneficial for the acceleration of breakthroughs in AI technologies, for example tax break or the lessening of government intervention. On the other hand, they are also concerned with the inequalities that are generated by the rapid growth of AI technologies and, therefore, advocate a redistributive policy that targets the AI (Stone et al., 2016, p.9; Korinek and Stiglitz, 2018). In other words, the dilemma is a trade-off between the freedom in AI development and social equality.

It should be noted that both the alarming view and the optimistic view look at the issue of AI development from an end-state perspective; namely, the desirability of the state of affairs in which the extensive application of AI technologies becomes the norm in social life.[1] [2] While the two end-state views have occupied the current public and academic debate on AI development, I believe that the analysis of justice in the production of AI deserves greater attention. The reason for this is that the creation of AI should not be perceived simply as ipso facto just, and it is dangerous to uncritically accept the clichéd view that, since AI developers mix their intellectual labour with other necessary resources to produce intelligent machines, they are, therefore, the rightful owners of the machines. In contrast, if injustice does occur during the production of AI, then the AI developers' entitlement to their products is tainted. Subsequently, a legitimate claim to their AI creation is limited or even extinguished (Goodin, 2013, p.487).

Moreover, taking a stance on the topic of justice in the creation of AI helps us to advance our understandings on the debates regarding the two end-state views. For the first form of the alarming view, although Bostrom (2014) rightly proposes that we should carefully analyse the true potential of AI before actually beginning the AI development process, his suggestion fails to impose enforceable obligations to relevant agents. Therefore, given the fetish for technology and the financial incentives in the real world, it is no surprise that Bostrom (2014) warns that 'some little idiot is bound to press the ignite button just to see what happens [when AI is pushed to its limits]' (Bostrom, 2014, p.259). However, the unjust ownership of AI provides a justification for the government to intervene in the development of AI. Furthermore, depending on the nature of the wrongdoings in producing AI, one solution to rectify such injustice is to open AI technologies to public ownership, which would ensure that the course of AI

[1] The concept of the 'end-state' principle comes from Nozick (1974, p.155). For Nozick, an end-state principle differs from a historical principle as, while the latter vindicates the rightness of a state of affairs by looking mainly at the process of how it has come about, the former imposes extra criteria on the features of the state of affairs in judging its rightness (Van Der Veen and Van Parijs, 1985, p.71).

[2] It should further be noted that the alarming view and the optimistic view are not incompatible; it is theoretically possible to be intrigued by the Pareto improvement that is brought by the technologies of AI while also arguing for tighter regulations to constraint the alarming features that could be brought by it.

development is decided by democratic process. The plausibility of this model of AI socialism will be discussed briefly at the end of the dissertation. Regarding the second form of the alarming view, the liberal argument of resource redistribution through taxation of the owners of the AI to preserve the ideal of equality may be unconvincing for non-liberal scholars. However, as this dissertation attempts to demonstrate, most political theorists can at least agree that an obvious injustice occurs during the process of AI building, and also that steps should be taken to rectify this injustice. Therefore, I believe that the argument in this dissertation, if successful, should appeal to a wide range of audiences in the theoretical spectrum.

Finally, for the optimistic view, the argument in this dissertation dismisses the trade-off between the pace of AI development and social equality; if wrongdoings occur during the production of AI, then AI becomes an inextricable joint product that is created through the productive capacities of the AI developers and contributions from the victims of the wrongdoings. According to Nozick (1974, pp.188–189), incentives are only required when the producer's contribution to the final output can be clearly identified, and this is not the case in the unjust productive process of AI.

This dissertation conducts an inquiry into justice in the production of AI. By referring to a widely applied principle of entitlement in Western political philosophy, i.e. the workmanship ideal of ownership, this dissertation argues that the current practice of AI development involves fraudulent actions and is, thus, unjust. The fraud in the development of AI occurs when AI developers try, through deception, to acquire a crucial resource as the input to produce the AI system – that is, information in the form of data. Specifically, the dissertation begins at Part I by introducing the workmanship ideal of ownership (section I-i) and applies this to the context of AI development in constructing the formal version of the argument for the injustice in AI development, known as the taintedness argument of AI ownership (section I-ii). Subsequently, Part II breaks down the taintedness argument by first focusing on the ownership of different types of data and tries to determine the rightful owners of the data (section II-i). This is followed by the argument that the action that AI developers take to make people to transfer data or provide the labour of data reporting is fraudulent in nature and is, therefore, unjust (section II-ii). Part III begins by analysing one proposal to rectify the injustice in AI production, i.e. the creation of a data market (Posner and Weyl, 2018). The dissertation argues that the proposal of a data labour market cannot eradicate the unjust nature of data production because the mechanism of market exchange for data transaction wrongly denies the legitimate claims of the data providers. Finally, the dissertation will end with a brief analysis of an alternative model to rectify the injustice in AI production, i.e. an AI socialist system. Overall, this dissertation does not aim to provide an all-things-considered theory of justice with a background structure that regulates the distribution of AI as a property.[3] Instead, it will demonstrate that a widely recognised injustice has occurred in the current practice of AI production, and will suggest that some measures should be taken to rectify this injustice.

[3] However, this is a worthwhile project that the holders of the end-state views should be interested in. Moreover, as I have discussed, my argument will be helpful for the holders of the end-state views to this end.

Part I. The Workmanship Ideal and the Taintedness Argument

I-i. The Workmanship Ideal and the Theory of Justice

The substantiality of the claim that the means to create an AI is unjust requires both a normative principle of justice as well as empirical evidence that comply with this principle. This part of the dissertation begins by identifying the former before applying it to the context of AI development.

One prominent idea that has been incorporated by many Western political thinkers in the construction of their theories of property rights is the workmanship ideal of ownership.[4]

The workmanship ideal of ownership: if person *A* mixes *A's* productive capacities with resources that are justly acquired, then *A* is the legitimate owner of the output of this productive process (Shapiro, 1991, p.47).

I believe that the workmanship ideal provides a strong prima facie reason for assigning ownership of the outputs of a productive process to its producer. Moreover, according to Cohen (1995, p.67), the Nozickian argument against the liberal equality-preserving redistribution of property is derived from a strong notion of self-ownership that is also embedded in the workmanship ideal (Van Der Veen and Van Parijs, 1985, n.5).

To reformulate Nozick's principle of justice in transfer with the workmanship ideal:

[4] However, as will be shown later, different theorists also have different levels of acceptance of this ideal. See Shapiro (1991) for detail on how different prominent Western political theories have embraced the ideal to different degrees.

Person *A's* strong self-ownership entails the strong ownership of both *A's* labour and the subsequent products of the labour. Therefore, the logical conclusion is that *A* has a strong right to freely contract out *A's* labour or to dispose of the products that arises from *A's* labour.

However, while the workmanship ideal has a place in most Western political theories, unlike Nozick, not all theorists embrace for a full-blown version of the ideal. For example, while Cohen (1995, p.31) accepts that gift-giving and some degrees of market transaction are compatible with a socialist theory of justice, he rejects the unrestricted voluntary transaction of people's private property as it is vindicated by the full-blown acceptance of the workmanship ideal, since this could potentially lead to political domination in a society that undermines the socialist concept of justice. Moreover, luck egalitarians would also reject the full implication of the workmanship ideal as they believe that people's productive capacities are arbitrarily decided by factors such as upbringing and native talents, and that these features of arbitrariness weaken the normative force of the workmanship ideal in determining a person's entitlement (Shapiro, 1991, pp.56–58). Therefore, one reason of disagreement between theorists on a person's entitlement to a particular property is the result of the different versions of the just background structure that the theorists hold, and thus the differences in the degree of the workmanship ideal that they are prepared to embrace (Van Der Veen and Van Parijs, 1985, n.5).

Does the analysis thus far imply that we cannot talk about justice in property entitlement and, therefore, justice in the ownership of AI without committing to a specific theoretical framework of justice? I believe that it does not, necessarily. We can consider the following reformulation of the workmanship ideal.

The negative thesis of the workmanship ideal: if person *A* mixes *A's* productive capacities with resources that are *unjustly* acquired, then *A* is *not* the legitimate owner of the output that arises from this productive process.

It should be noted that this negative thesis does not need to rely on a specific theoretical framework; this is justified by Sen's (2006) argument that some just-enhancing actions can be carried out without having to subscribe to a full theory of justice. Examples of these just-enhancing actions involve slave abolition and the prevention of arbitrary incarceration, as the rightness of these actions can be approved by most, if not all, political theorists. Similarly, despite his disagreement with other theorists on the subject of personal entitlement, Nozick approves the widely-accepted proposition that injustice can occur when 'some people steal from others, or defraud them …' (Nozick, 1974, p.152). Thus, I believe that there are certain types of wrongness in production that can be recognised by most theorists who, to some degree, have subscribed to the workmanship ideal, and I will argue that the injustice that is performed by the producers of AI during the productive process belongs to the wrongness of this type. In other words, the specific injustice in the production of AI is *sufficient* to undermine the AI producers' legitimate ownership of the final product regardless of the background theory of justice.

I-ii. AI and the Taintedness Argument

In order to identify the specific injustice in AI production, a brief account of the fundamental concepts of AI is helpful.

Research in the field of AI can be defined as studies on the properties of intelligence (Stone et al., 2016, p.13). Therefore, the task for AI researchers and developers is to 'build and study systems that exhibit intelligence' (Simon, 1995, p.101).

Initially, the system of AI was reliant on the pre-programmed formal logical rules used to process the given inputs to produce the final intelligent output (Posner and Weyl, 2018, p.212). However, this method of AI development has been replaced by the more effective approach of machine learning (ML). The computational technique of ML powers the AI system with the formulation of an algorithm that can perform human-like sophisticated operations on the given inputs, known as data mining (Stone et al., 2016). One widely applied form of data mining in ML is known as supervised learning. The process of supervised learning involves the developers providing the algorithm with a set of inputs with different labels. The intelligent algorithm then attempts to perform data mining in the given inputs and learns to make future intelligent predictions. For a simplified example, by providing the ML algorithm with a collection of emails, each with the labels of ‘spam’ or ‘non-spam’, the ML process can then identify the characteristic pattern of emails within a specific label group. Therefore, by performing data mining on the input emails, the AI system is able to predict the labels of future new emails – that is, the ability to distinguish whether or not the email is spam. Moreover, the ML algorithm’s capability of identifying features in the given data resembles the more complex aspect of epistemic cognitive ability of the human brain. This can be demonstrated by another form of data mining, known as unsupervised learning. In the process of unsupervised learning, the intelligent system is trained to make prediction through identifying the hidden characteristic pattern from unlabelled, raw and messy inputs (Posner and Weyl, 2018, p.214). To demonstrate the AI’s potential capability in unsupervised data mining, let us consider the AI application of object recognition. Starting from a collection of random imagines whose labels are unknown to the AI. The AI is first taught to analyse the fundamental relevant features of the input images, such as the pixel patterns and the number of lines in the images. This allows the AI to determine certain more complex features within the images; for example, whether the image is an animal or a vase. Next, the AI uses the previously obtained features to conduct pattern mining in identifying even more complex features; for example, the types of animal in the image. This process is repeated until the AI is able to make intelligent predictions on the specific objects in the image (Salian, 2018; Posner and Weyl, 2018, p.216).

To generalise, for an ML algorithm, the input is constituted by a variety of pertinent information collected in the form of data and, according to the above depiction of the concept of ML, data as an input sets the whole process of AI development in motion. Moreover, the increase in the quantity and quality of data also plays an important role in training the AI system, which enhances its competency in performing the assigned task (Posner and Weyl, 2018, pp.216–220). In the example of object recognition, when insufficient data is supplied to the ML algorithm, the AI may misconceive certain important features within the images, subsequently mistaking one type of object for another. Therefore, it is unsurprising that the technologies and applications of AI have grown rapidly alongside the breakthroughs in big data technologies such as cloud computing resources and advancements in data-gathering methods (Stone et al., 2016, p.14).

Given the relationship between data and AI, one possibility for injustice to occur in AI development is when the data used in the process of ML belongs to people other than the AI developer and, moreover, when the AI developer's actions in acquiring these data involve wrongness. If this kind of injustice does occur, then, according to the negative thesis of the workmanship ideal, the unjust means in the productive process are sufficient to taint the ownership of the output for the AI developer. Specifically, the formalised version of this taintedness argument is presented as follows:

P1. Person *A* has ownership of *X*.

P2. Person *B's* action *C* that results in *A's* surrender of *A's* ownership of *X* is unjust.

P3. Because *C* is unjust, *B's* ownership of *Y* is tainted, where *X* is the input of *Y*.

C. Steps should be taken to rectify the taintedness of the ownership of *Y*.

While *P3* is supported by the negative thesis of the workmanship ideal and *C* follows logically from *P3*, the plausibility of the taintedness argument depends on the soundness of *P1* and *P2*.

To elaborate, we can consider *P1* in the context of AI development. If we fail to establish the ownership of data for person *A*, then this will imply that the claim in *P2* cannot be sustained. This is because person *A* cannot accuse person *B* of violating *A's* right to an object that does not belong to *A*.[5] Moreover, even if it has been proved that the ownership of data does not initially belong to the AI developer, the taintedness argument can still be rendered implausible if *P2* is false – this can be proved by demonstrating that action *C*, the process of data transaction from its original owners to the AI developers, is just.

The next part of the dissertation will be used to consolidate the taintedness argument by first looking at the issue of data ownership before turning to identify injustice in the data transaction.

[5] Except perhaps in the case of initial resource acquisition: if the worldly resources are initially unowned, then, if person *B* acquires *X*, which represents a large portion of the worldly resources, and leaves person *A* with almost nothing, according to Otsuka (2003, p.23), it still seems that some wrongness has been committed against *A*. This dissertation will not analyse the injustice in data acquisition from this perspective.

Part II. Data Ownership and Fraud in the Data Transaction

II-i. The Rightful Owners of Different Types of Data

Let us first consider the first premise of the taintedness argument.

P1. Person *A* has ownership of *X*.

Different types of AI system are built to perform different tasks that incorporate a feature of intelligence (Simon, 1995, p.47). As a result, the types of data as input for the ML process also vary according to specific intelligent tasks. Therefore, to discern the rightful owners of data used to train the AI, it is beneficial to first distinguish different types of data that are of interest to AI developers. Table 1 categorises four different types of data using a two-dimensional distinction, i.e. the labour/non-labour distinction and the marginally/aggregately valuable distinction.

Table 1. Examples of different types of Valuable data for AI training

Nature/Improvement	**Marginally Valuable**	**Aggregately Valuable**
Labour	Expert human moves in Go	General feedbacks and judgements on the output of AI Language inputs in training the natural language processing system
Non-labour	–	Images (for object/facial-recognition system)

To elaborate, the labour/non-labour distinction attempts to highlight the origins of the data. Some data, such as information on how to win in a strategic game, involves the deployment of a person's intellectual labour. It should be clarified that the labour category includes both data that is produced by labour alone, as in the case of contemplating a strategic game move, and data that is produced by mixing labour with other external resources, such as testing a chemistry theory in a laboratory setting, leading to new scientific findings. In this dissertation, the data that is identified as valuable in AI training in the labour category belongs mostly to the pure-labour data. Alternatively, the non-labour category includes data that is not produced by any labour; for example, information about a person's height or body temperature.[6] The marginally /aggregately valued distinction is introduced due to the different nature of the tasks performed by AI. Some intelligent systems are created to perform tasks that attempt to make predictions based on statistical distributive models on general populations. For example, natural language processing is a subfield of AI that works on the application of speech recognition or language translation. The intelligent systems assigned to these tasks require large collections of relevant language inputs that can represent specific language systems in order to perform accurately. Therefore, the information contained within a single piece of relevant input is aggregately valuable, as this data *alone* does not lead to intelligent output from the system (Stone et al., 2016, p.32). Alternatively, other systems of AI are given the goal of performing strategic tasks such as playing a game of chess, or Go. For such AI, as suggested by case 1 in the introduction, by inputting strategic moves from experts of the game, the AI's performance can improve rapidly in terms of the ability to win games (Silver et al., 2016, pp.484–485). Since data such as chess moves from experts can independently improve the output of the AI, it is, therefore, marginally valued.

It should be noted that the marginal/aggregate distinction does not have an impact on the analysis of the ownership of the data since a person's entitlement to an object does not depend

[6] It might be argued that this information is also determined by certain labour; for example, my effort in rope-skipping affects my height measurement in the future. However, this information is not created by labour in a sense that I will have a height measurement *simpliciter*, regardless of my action of rope-skipping.

on how the object is valued. Therefore, the remainder of this section will investigate the ownership problem based on the labour/non-labour distinction.[7]

Ownership of Labour-created Data

According to the distinctions above, labour-created data for AI training can be either marginally valued or aggregately valued; the former can include data such as chess or Go moves from human experts, while the latter can include data in the form of human feedback used to improve the accuracy of the performance of an AI natural language processing system.

Otsuka (2003, p.19) believes that there is a strong reason to assign ownership of productive output to its producer if:

> … her means of production consist of nothing more than her mind and parts of her body, and the fruits of her labour consist of nothing more than parts of her body that have been transformed into items [the productive output] …

By avoiding the controversial problem of the ownership of external resources used in production, Otsuka's argument can draw support from the less problematic part of the workmanship ideal, which condemns the wrongness of denying people what they produce with purely the resource of their bodies. Moreover, while Otsuka (2003, p.19) presupposes 'a full right of self-ownership' when making this argument, he believes that, even for scholars who subscribe to a weaker version of the right of self-ownership, it is difficult to justify the policy of radical taxation on purely labour-created output. I would like to add to this argument by pointing out that the normative force for condemning the wrongness of denying producers their purely labour-created output becomes even stronger if: a) the possession of the product does not undermine the welfare of other people; and b) the inability of other people to produce the

[7] This dissertation can't address all sorts of data used in AI training. Therefore, mainly the examples of each category of data as shown in Table 1 are used in illuminating the ongoing arguments.

similar output will *not* undermine their lives in a significant way.[8] Violation of criterion a) can be illustrated by the Nozickian Wilt Chamberlain example, where Chamberlain's ability to play basketball allows him to accumulate an enormous amount of social resources which is the sine qua non of political domination(Cohen, 1995, p.25). Criterion b) is unsatisfied in Otsuka's (2003, p.18) example of an able weaver and an unable weaver, whereby the former has the ability to produce cloth that can prevent death while the latter lacks such an ability, which, therefore, provides grounds for redistribution.

Let us apply the argument in the context of labour-created data. In both cases of aggregately and marginally valued data, a person exercises his mental capacity alone in producing the new valuable information. Moreover, information such as judgements on AI performance or Go knowledge satisfy criterion b), as the inability to produce the information does not undermine a person's chance to lead a meaningful life. In terms of criterion a), while the possession of a personal judgement or linguistic ability which can enhance an AI's performance will generally be harmless to other members of society, an argument can be made that, similar to Chamberlin's basketball skills, knowledge of how to win a game of chess could potentially help a person to accumulate a substantial amount of resources in a chess-fanatic society. While I acknowledge that a society will always favour some skills over others, a society can address criteria a) without having to violate the right of self-ownership. As Walzer (1983, xiv) suggests, this can be achieved by making sure that 'no social good…can serve as a means of domination'. Even for people who think that Walzer's proposal might be too idealistic, the empirical observation that knowledge that is valued by the AI trainers usually do not lead to the original holders of the knowledge to accumulate an enormous amount of political influence, can at least provide an *ad hoc* justification for respecting the ownership of this knowledge.

[8] According to Sen's (1999) capability account of freedom and the Rawlsian(1971) conception of primary good, there are certain personal capabilities and resources that are of significant importance to every individual's life.

Therefore, according to the above argument, there is a strong reason for allowing a person to be the rightful owner of her purely labour-created information that is used in AI development. This implies that she has the right to withhold the information or to choose the terms for sharing the information. Moreover, she should not be forced to share this information against her will.

However, critics might suggest that the Otsukian argument of ownership could seem convincing in the case of tangible properties, but not in the context of the intangible property of data, and the ownership right of intangible data that is produced by a person's intellectual capacity is less stringent. This is due to the fact that intangible information is non-rivalrous, meaning that it can be used by other people without the need to reduce the quality and availability of the information for its original owners.

I believe this critique is supported by two distinct but connected justifications. According to the first justification, the claim is that, once the labour-created information is exposed in public, the information producer effectively imposes the knowledge on others. Moreover, since other people can independently access the information and the original producers cannot practically force others to forget the information, there is no reason to exclude others from the information (Moore, 2012, p.1092). For example, when a Go expert plays her moves in a public tournament, the moves are mentally recorded by observers, and it becomes difficult practically to exclude others from accessing her strategy. However, this practical difficulty in excluding access to information once it has been exposed to the public does not itself suggest that it is morally right for everyone to acquire this information (Himma, 2005, p.8). For example, let us assume that the vicious craft of dark magic from the world of *Harry Potter* has been exposed to the public and, hence, it is now known to everyone; this does not suggest that it is morally unproblematic to acquire this knowledge or to cast this evil spell. The point is that a normative reason needs to be supplied to justify why our acquisition of the labour-created data produced by others is morally permissible. Moreover, as the next section will argue, AI developers who try to access data are not always as innocent as depicted in the case of the Go game observer since information is sometimes intentionally extracted from its original producers for the benefit of AI developers. Overall, it is implausible to claim that labour-created data should be made free for use by everyone simply because everyone can access it; it is also wrong to assume that this type of data will always automatically be exposed to the public. The first justification regarding the empirical availability of data fails to undermine a person's right to his labour-created data.

The second justification then tries to provide a normative reason for justifying the liberation of the labour-created information. According to this justification, data should be liberated because it is good for someone but worse for no one. For example, if I have discovered the secret to living a happier life, I will not be at a disadvantage if I accidentally give it away while chatting to someone in a conversation in public. Moreover, the person who receives the knowledge from me will gain from my discovery. In response to this justification, I would like to point out that, when using labour-created data to train AI, the original data creator does lose out from the liberation of their data. As suggested by case 2 in the introduction, every time data is collected, this contributes to the emergence of a powerful, fully trained AI, which could impose a heavy cost on the data providers by disrupting the structure of the job market. An optimistic AI fanatic might then argue that the costs the AI imposes on people are outweighed by the benefits it provides. If this optimistic view is made on an individual level, we can easily reject it by identifying people whose life plan will be heavily disrupted by the rapid growth of AI; for example, a data provider who dreams of becoming a truck driver but, after years of training, finds herself in a world of driverless transportation systems. Alternatively, if the optimistic view is made on an aggregate level regarding net social wellbeing, it neglects the wellbeing of the people who suffer as a result of AI development and fails to recognise the moral significance of individual human life and, thus, fails to recognise the separateness of persons

(Rawls,1971,pp.26-27). Overall, the normative argument that people's labour-created data should be liberated since it is beneficial for all members of society is implausible. Together, the argument so far suggests that people should have strong ownership of their labour-created data that is used in AI training.

Ownership of Non-labour-created Data

Non-labour-created data can also be valuable in training AI. Let us recall the ability of AI to extracting substantial information from raw input through the process of ML to make smart predictions; this can include personal information such as income, height and health status, all of which can be useful for training AI to perform intelligent tasks. However, our ownership regarding this type of personal data is more difficult to establish. As Moore (1998) notes:

> … consider someone walking in a public park. There is almost no limit to the kinds of information that can be acquired from this public display. One's image, height, weight, eye colour … are all readily available … it would seem that all of one's genetic profile *is not private information* (Moore, 1998, p.372, emphasis added).

Moreover, Moore (1998, p.372) is also aware that 'what is publicly available information is dependent upon technology'. For example, data such as a person's heart rate is becoming easy to obtain with inventions such as the smart watch.[9]

Regarding non-labour-created data, I agree that it is difficult to establish ownership of such information. This is not because the data is publicly observable; as argued earlier, the ability to

[9] Moore (1998, p.372) draws a distinction between the violation of property rights and the violation of the right of privacy; he argues that the collection of non-labour personal data is wrong for the latter reason. In order to avoid conflation between different arguments, this dissertation focuses only on the wrongness of data collection from an ownership perspective.

access the data does not suggest that it is just to do so. However, privacy reasons aside, it is difficult to make a normative claim that a person owns the measurement of his height or his body temperature.

I would argue that a connection can still be found between our non-labour-created personal data and our right to property. This connection is justified by the empirical observation that AI developers usually require an enormous amount of personal data for AI training and, for reasons of efficiency and accuracy, AI developers are usually reliant on people to report such personal data. For this reason, the success of AI training is dependent on people's actions in reporting relevant information about themselves; therefore, even if a person does not have exclusive ownership of non-labour-created personal data, he certainly has ownership of his *labour of reporting* such data to AI developers, which can also be vindicated by the workmanship ideal (Posner and Weyl, 2018). I am aware that this argument has a deficiency, as it implies that, if AI developers decide to take up the task of data collection themselves, for example, by setting up a biological detector in a public place, or collecting users' biological features through smart gadgets, then we can have no complaint *from a property right perspective* against the data-collection activities.

II-ii Fraud in Data Transaction

In the last section of the dissertation, it was argued that people have ownership of data that is created purely by their labour, and they also have ownership of their labour in reporting non-labour-created data to AI developers. Overall, *P1* of the taintedness argument is justified as some forms of ownership have been established that are related to data that is used in the process of AI training that does not initially belong to AI developers. This section of the dissertation analyses the second premise of the taintedness argument.

P2. Person *B's* action *C* that results in *A's* surrender of their ownership of *X* is unjust.[10]

The wrongness of data collection by AI developers has been indicated by Posner and Weyl (2018, pp.220–223). Posner and Weyl rightly identify that there is an oligopoly model of competition in the AI industry, which is dominated by leading technology firms such as Facebook and Amazon. The oligopoly structure of the AI industry is unsurprising if we consider that these leading firms provide the most essential services in the digital community such as platforms for social networking and e-commerce; through their daily interactions with users, these firms can extract numerous valuable data to conduct AI training. Moreover, these firms are also capable of extracting specific data of interest by psychologically nudging their users when using their online platforms. Overall, the privileged position of these firms in accessing clients' data allows them to stay ahead in the development of AI technologies. If we take the example of Facebook, Posner and Weyl (2018, pp.220–221) highlight that the firm benefits from the millions of photos that are uploaded by its users, which can be valuable in training the AI system responsible for object recognition. Moreover, by creating the 'label' and 'hashtag' functions on its social networking platform, Facebook also nudges its users to provide more refined data that can improve the efficiency of the AI learning process.

According to the depiction above, this dissertation recognises the wrongness that is committed by the AI developers as the unethical action of fraud. The notion of fraud can be broken down into three criteria:

1. Transaction takes place between person *A* and person *B*.

[10] Following the findings from the previous section, *X* can be either a person's ownership to her entitled data or the ownership of her labour of data reporting.

2. *B* provides *A* with an offer that involves information that is intended to misguide *A*'s judgement about the 'character and/or the economic value' (Cohen, 1995, p.39) of the transaction.

3. *A* accepts *B*'s offer.

The wrongness of fraud can be condemned on various grounds. Since the act of fraud uses the deceived victims as a means of achieving the fraudster's goal, the fraudulent act should be prohibited by foundationalist political theorists who subscribe to the widely accepted Kantian morality of 'Kingdom of Ends' (Kant, 2002). Moreover, for constructivist theorists such as Rawls(1971) and Scanlon(1998), our strong intuition against fraud also provides a strong justification to identify the fraudulent acts as unjust.[11] For the current argument, it is sufficient to note that theorists who subscribe to the workmanship ideal will condemn the illegitimate acquisition of property from its rightful owners through fraud. Moreover, as Child (1994, p.723) shows, a theory that is unable to address the wrongness in fraud makes the theory deficient. Since the injustice in fraud is widely condemned, the more interesting task is to demonstrate that the actions of AI developers satisfy the definition of fraud. Let us consider the following running example as an illustration of fraud in a transaction.

Landslide

A group of scientists go on an expedition as they believe there is an extremely valuable material located at the peak of a mountain that they can use to produce a profitable high-tech laser gun.[12] The mountain is located in a rainforest and there is a village at the foot of the mountain. Upon reaching the foot of the mountain, the scientists find that there is no readily available path to lead them to the peak of the mountain. However, they discover that there is one square meter

[11] The foundational and constructive approaches are two of the main models adopted by philosophers in moral justification (Dworkin, 1973, pp.509–511).

[12] Let us assume the material is left behind by the ancestors of the scientists and, therefore, belongs to the scientists, so to avoid the complication of the distribution of unowned worldly resources.

of land located just outside village, which we will denote as *L*. The scientists realise that whenever someone walks on *L*, it slightly alters the geological structure of the mountain because of the trampling activities. Moreover, if the trampling activities take place in a sufficiently large quantity – at the *n*th time – a landslide will occur that will create a path to the peak of the mountain. The group then builds a carnival circus next to *L*, hoping that more villagers will walk on *L*.

The Landslide example demonstrates a clear case of fraud. In the fraudulent transaction that takes place between the scientists and the villagers, the scientists' offer involves the misguidance of the villagers' judgment. Specifically, the villagers are offered entertainment in the form of a circus performance. The villagers then accept the offer without realising that they are giving up ownership of their labour, which is being used as a means of helping the scientists to produce the profitable laser gun.

The practice of AI developers in collecting data resembles the Landslide example. Let us recall the findings thus far: people have ownership of their labour-created data, and they also have ownership of their labour of reporting the non-labour-created data to the AI developers. Moreover, users of social digital platforms give up this information to technology firms such as Facebook and Google while enjoying the 'free' services the firms provide. The transferral of users' legitimately owned data to the AI developers occurs when the users are deceived into helping Google's AI system to perform text-recognition tasks (labour-created data) while believing that they are completing a word-distinguishing puzzle for security purposes (Posner and Weyl, 2018, pp.235–236). Fraud also occurs when users upload their selfies to the picture-processing application FaceApp (labour of reporting non-labour-created data), where the uploaded images can be useful in training the AI system to perform image recognition (Brewster,2019). Therefore, the analogy of the Landslide example and the method of data collection in AI development suggests that, as in the former, the latter practice is also fraudulent and thus unjust.[13]

Sceptics of my analysis might resist the above conclusion by providing a value-judgemental argument; the argument attempts to show that the transaction between the data owners and the AI developers is not fraudulent by making one of the two following value-judgemental claims.

Claim 1: The data transaction is not fraudulent since, unlike the Landslide example, the profitability of the AI is undetermined. The development of AI may fail as a scientific project, or the mature AI system may fail to be profitable in the market. Therefore, given the expected value of the profitability of the AI system, and considering the risk of failure taken by its developers, the users receive a fair share from the data they have provided in the form of free usage of online services.

Claim 2: The data transaction is not fraudulent even if, like the Landslide example, the AI system is destined to be profitable. This is because, unlike the Landslide example, the success of an AI system is mainly due to the skills and efforts of the AI programmers and entrepreneurs, while the contribution from the data providers is relatively small and they have received proper compensation in the form of the free usage of online services (Posner and Weyl, 2018, p.224-226).

Both of the above claims impose a judgement on the value of the data that is used in AI training. Moreover, they both attempt to argue that, through the unrestricted consumption of online

[13] The marker of this dissertation raised the interesting point that the users' action of making data publicly available can be interpreted as a form of tacit consent for the general public, including the AI developers, to acquire such information. An obvious example is the uploading of one's pictures into Faceapp or Facebook. However, I believe the normative force of using such consent to validate the justice in data transaction is weak for two reasons. First, as it will be shown later, if the structure of the contract is fraudulent then the consent on such unfair terms also loses its moral significance. Second, it is unclear that what exact terms a person has consented to when she chooses to share her information. However, it is uncontroversial to assume that she would not consent the AI developers to use her data in building a dominant AI system that could undermine her well-beings.

services, the data providers receive a fair reward for their contributions and, therefore, the transaction of data is not fraudulent, but is fair and just. I think that the value-judgement argument is implausible for the following reasons.

First, the act of fraud does not depend on the upshot of the transaction. For example, in the Landslide example, after the success of their plot, the scientists reach the peak of the mountain only to find that the valuable material has been destroyed during the landslide and, therefore, they have provided the free circus for the villagers for nothing. However, despite this outcome, the act of violating the villager's ownership of labour to achieve a goal is still unjust. In other words, if we distinguish just steps from just situations without presupposing that the former automatically entails the latter, then it is possible to have a just situation in which everyone is legitimately entitled to their holdings, which are the outcome of unjust steps (Cohen,1995, p.43). Specifically, in the new Landslide case, the unjust step of fraud is neutralised by the natural event of the Landslide, since the failure to produce the laser gun means that the scientists do not have the opportunity to possess the would-be unjust property.

However, using the above reason *alone* to discredit the value-judgemental claims is uncharitable, since the claims do not attempt to use the possible final outcomes per se to justify that the transaction is not fraud. Instead, the claim is that, whatever the outcome, it follows from procedures that warrant a fair distribution of the joint productive venture, which acknowledges the contribution from all relevant parties; therefore, the data providers cannot reasonably reject the transaction.[14] However, I believe that the argument that the final distribution is fair, and that, therefore, the transaction is just, moves too quickly. If the contributors did not enter the productive process voluntarily but as a result of the unjust actions of others, then we cannot discuss the fairness of the outcome before rectifying the injustice in

[14] The concept of reasonableness comes from Scanlon (1982, pp.110-112). For Scanlon, a contractual agreement is justified if no relevant parties can reasonably reject the terms of the agreement.

the production. I believe that this follows from the rationale of Nozick's (1974, p.95) argument against the principle of fairness; Nozick believes that we cannot 'just act so as to give people benefits and then demand (or seize) payment'. I believe this Nozickian argument becomes more convincing if applied in the context of private transaction where no common ownership is involved; AI developers cannot illegitimately acquire data and labour from their original owners and compensate them with free services while claiming that the owners have received their fair share of contributions.[15]

To clarify, the fraudulent problem with the data transaction is distinct from the problem that the payment to the data providers fails to reflect their contribution to the productive process. Rather, it is the intention to trick the data owners into participating in the production of AI and into giving up their rightfully owned properties in the process without properly declaring the true intention of the transaction that renders the data transaction fraudulent and, therefore, unjust. As a result, instead of treating the data providers as stakeholders in the productive process, the AI developers unilaterally impose a price on the provided data which exacerbate the injustice in data transaction.[16]

If we combine this finding with the finding on data ownership from the last section, this dissertation, thus far, has demonstrated that the ownership of data used to train AI does not initially belong to the AI developers. Moreover, the approach the AI developers use to acquire such data is unjust. Together, the validation of *P1* and *P2* of the taintedness argument in relation to AI brings us to the third premise which is supported by the negative thesis of the

[15] A difference between Nozick's (1974, pp. 94-95) examples and the case of data collection is that, in Nozick's case, the person is being forced to contribute to the productive process, while in the data case, the person is being tricked into participating in the productive process. However, I believe that Nozick's argument can be adequately applied in the case of fraud.

[16] It will be shown in the next section that the requirement of treating the data providers as stakeholders is also unsatisfied in the system of the data market.

workmanship ideal and, hence, the conclusion that we should abandon the current practice of AI development.

P3. Because *C* is unjust, *B's* ownership of *Y* is tainted, where *X* is the input of *Y*.

C. Steps should be taken to rectify the taintedness of the ownership of *Y*.

The next part of the dissertation analyses the plausibility of proposals that aim to rectify the injustice in the AI production process.

Part III. Critique of the Data Market; the Proposal of AI Socialism

The previous part of the dissertation analysed the validity of the taintedness argument in relation to AI and justified the claim that AI developers' entitlement to the AI systems they have created is tainted, given the fraudulent actions involved in the data collection while building the AI. Subsequently, actions should be taken to rectify this wrongness. Moreover, before certain forms of rectificatory action take place, a person's ownership of a tainted property is void, meaning that the person cannot make a claim on the property (Goodin, 2013, p.487).

In the context of AI as a tainted property, Posner and Weyl (2018) propose one form of rectification, i.e. the creation of a data market.[17]. Let us recall the value-judgemental argument from Part II-ii, in which Posner and Weyl's proposal is a challenge to the value-judgemental argument over the fair reward for the data providers. Specifically, Posner and Weyl use the

[17] The approach of the data market is forward-looking, aiming to prevent the unjust collection of data in the future. However, Posner and Weyl must also provide a solution to rectify the past wrongness that has already been committed by the AI developers.

term ‘technofeudalism’ to criticise the underpayment to the data providers by the AI developers (Posner and Weyl, 2018, p.231). To advance their argument, Posner and Weyl (2018) are faced with the challenge of determining the fair market price of data and the labour of reporting data; therefore, a substantial part of their proposal addresses this issue. For example, they attempt, rightly, to show that information should be valued at a higher price since it is both marginally valuable and aggregately valuable at the same time (Posner and Weyl, 2018, p.225–230); they are also uneasy regarding the oligopoly structure of the AI industry as it could suppress the price of data. Therefore, they have adopted a socialist argument of establishing a labour union in the data market (Posner and Weyl, 2018, pp.234, 239–243). Moreover, they envision a technological system that can track and measure the continuous value that is generated by people’s data and the labour of data provision (Posner and Weyl, 2018, p.244).

While I can agree with Posner and Weyl that the current practice of data transaction denies the data providers with their fair share of income, I argue that the proposal of the data market cannot rectify the injustice in the practice of the data transaction. My claim will now be elaborated.

In Part II-ii, I argued that it is inadequate to discuss fairness in the distribution of the output in AI production without first resolving the injustice in the data transaction. Moreover, I identified the injustice in data collection as the fraudulent acts of AI developers in misguiding people to transfer their rightfully entitled data and labour in a transaction. However, people who find the proposal of the data market appealing could point out that, by engaging in the market, people are aware of the transaction and are, therefore, no longer subject to the AI developers’ misguidance. Moreover, people in the data market act as different data merchants who can autonomously and rationally decide whether to engage in the data transaction. As a result, proponents of the data market can argue that fraud is eliminated by the market system and the data market simultaneously rectifies the problem of injustice in the data transaction and determines the right price for the data. For this reason, it can be concluded that my concerns have been addressed by the data market system.

Proponents of the data market may try to demonstrate their argument further with a non-fraud version of the Landslide example. In the non-fraud Landslide example, the scientists honestly tell the villagers about their actual project and the villagers agree to contribute to the production of the profitable laser gun. It could be argued that, as in the non-fraud version of the Landslide example, the data market serves as the platform for AI developers to honestly express their intention while the data providers decide whether to participate in the developers' project base on the market price. Therefore, if fraud is absent in the non-fraud Landslide example, then it should also be absent in the data market.

However, I believe that the data market is disanalogous to the non-fraud Landslide example for the following reason. In the non-fraud Landslide example, the villagers provide 'the muscle' by creating the path towards the peak of the mountain, and the scientists deploy their intelligence in processing the material. Together, the laser gun is a product that has been jointly created using the labour from both the villagers and the scientists. Moreover, as Attas (2004, p.539) highlights, one feature of a joint product is that there is more than one legitimate claimant and, moreover, no single claimant can transfer, destroy or exchange the joint product without the unanimous agreement of all other claimants. As a result, before an agreement is reached between the villagers and the scientists, the ownership of the laser gun is inextricably shared by the two groups. In the data market, however, AI developers always have the sole claim over the ownership of AI as well as the product produced using AI; for the data providers, who have also contributed to the productive process, their claim over these products is denied.

Proponents of the data market may respond by claiming that I am wrong to suggest that the data providers should have a claim over the AI output as they have already received payment for their data or the labour of providing data. As Attas (2004) notes:

> He … has … a right to exchange his product with others. If [the product] then becomes a part of a larger or more complex product, this is no concern of his, and he has certainly no further right to the final product of any part of it … (Attas, 2004, p.551).

However, Attas (2004, p.552) usefully draws a distinction between a primitive exchange system and a market exchange system. In the former system, people engage in production for the use-value of the output, but not the exchange-value. The latter system resembles the economy system in modern society, where people specialise in certain productive tasks with the goal of exchanging their products with others. Following Attas's rationale, it is plausible to claim that, in a primitive exchange system where the data creators engage in the process of data production with the pure goal of consuming the data themselves, and they are later involved in an unanticipated transaction with the AI developers as they would like to consume some online services that the AI developers provide, then they have legitimately transfer their ownership of data and labour to the AI developers. Therefore, they should have no claim over the final AI output even though it is derived from their data.[18]

In the actual data market, unlike the primitive exchange system, one group of market participants – the AI developers – has a clear intention in benefiting from AI as a final output, and they realise that they can only achieve their goal with cooperation from another group of market participants – the data providers. Meanwhile, the data provider would not engage in the production of data, for example helping to train the AI or reporting their data to the AI developers, had they not known that their product of data or labour can help to develop the final AI product.[19] Given this nature of AI development, both groups of market participants produce with the intention to exchange with each other in building the AI as the final product,

[18] The transaction is unanticipated since the original goal of production was not for the sake of exchange.
[19] In the case of data such as strategic game moves, the initial knowledge might be created for its use-value. But the emergence of AI also adds the layer of exchange to this knowledge.

as oppose to the case of a primitive exchange system. Therefore, the joint ownership problem emerges. The emergence of the joint ownership problem means that we cannot simply dismiss claims over the final joint product from any of the relevant contributors. Moreover, since the market is a platform on which all labours and products are aggregated for the emergence of the final output, in this case, ‘the [market]exchange system raises the problem of [the] division of joint production’ (Attas, 2004, p.552), rather than providing a solution for the distribution. Overall, if this argument is sound, then the data market is unlike the non-fraud case, since, in the non-fraud Landslide example, all contributors should have claims over the collective output, while the data market arbitrability awards the ownership of AI to the AI developers.

Therefore, the data market is unjust since, for the data providers, even if they engage in AI training process to improve the AI as the final output, they are denied their claim over the AI by the market, let alone having the ability to engage in an agreement with AI developers regarding how to distribute the benefits of this output. Instead, the data providers have to agree to follow the market system, without any feasible alternatives.

This feature becomes more worrying by considering the empirical observations relating to AI’s capabilities as discussed in this dissertation. It is uncontroversial to predict that, without entitlement to AI, the data providers are likely to lose out in the future, given the ‘probabilities of all significantly different possible outcomes of the market transaction’ (Cohen, 1995, p.48). At the same time, AI developers will gain from market activities. Therefore, following Cohen’s (1995, pp.48–53) analysis, if the benefits of engaging in market transactions are asymmetrically biased towards AI developers, and data providers are forced to accept their odds without the opportunity to propose new arrangements on sharing the benefits of the final output, then the data market is deeply unfair towards data providers.

However, it should be clarified that the above argument does not need to rely on the judgement on the upshot of the data transaction in the market for data providers. Rather, injustice in the data market is *partly* similar to the case of the current approach of data collection by Facebook and Google, which is the refusal to acknowledge the data providers as contributors to the AI development process and, therefore, the failure to assign them entitlement to the products that they jointly produce.[20] It may be the case that, although unlikely, data providers and AI developers agree on a distribution of AI that resembles the current distribution, in which data providers are compensated with free usage of online services. However, the point is that the lack of such opportunity to engage in the negotiation on the distribution of the ownership of AI renders the data market proposal, as well as the current fraudulent practice by AI developers, unjust.[21]

The argument thus far may lead us to consider alternative systems of AI development that assign claims over AI to all contributors in the productive process – an AI socialist system may be suitable for this end. While a detailed depiction of the AI socialist system is beyond the scope of this dissertation, I believe that it can take the form of the state, on behalf of its citizens, becoming the monopoly owner of AI technology and data for AI training. The state can lease these resources to democratic worker cooperatives, which will focus on different areas of AI research and production (Thomas, 2016, p.225). Moreover, cooperatives can trade in a capitalist market and receive profits from their production. Part of the profits made from such trade will be paid back to the state for leasing the data and technology; such funds will subsequently be redirected to the citizens qua data providers. The residual profits will be

[20] See footnote 14 above.

[21] To elucidate my position. I believe there are two types of wrongness in the current AI transaction, first, the intention to trick the data providers to participate in the AI production process and second the refusal to assign entitlement of AI to data providers, which is built using their labour or data. The proposal of the data market can be perceived as a massive information campaign in improving the awareness of the values of data, therefore it is free of the first type of wrongness. However, since the capital market still arbitrarily assigns full entitlement of AI to AI producers, the data market proposal still subjects to the charge of the second type of wrongness. After all, I believe, to rectify the injustice in AI development, it is necessary to raise public awareness. However, this approach alone is insufficient, as structural change for a fair system of AI distribution is required for this end.

divided among the workers according to their different productive contributions; therefore, an AI programmer will receive more financial rewards compared to an ordinary data provider.

While both the data market and the AI socialist system are efficient at helping AI to progress, the socialist system is more advantageous since it allows each contributor to have a claim over the joint products of AI. To elaborate, since it is unproblematic to assume that the majority of citizens have provided or will provide data that can contribute to the development of AI, by converting AI technology into public-owned property, each citizen qua data providers will have the opportunity to negotiate how the benefits of AI are shared and the direction of AI development through democratic means.

Conclusion

To conclude, this dissertation has focused on justice in the process of AI development. It has attempted to demonstrate that the current practice of AI training by developers such as Facebook and Google is unjust since the developers misguide their users into giving up their entitled data, or to unknowingly providing the labour of data reporting in the process of AI development. Moreover, by analysing the system of the data market, which is proposed as a means of rectifying the unjust practice of data collection, the dissertation argues that the data market system fails to treat data providers justly, since the market imposes unfavourable terms on them and denies them a claim over the products of AI that they jointly produce with others. On the other hand, the brief account of the AI socialist system proposed at the end of the dissertation allows each contributor to AI to have a claim on the AI through the democratic process.

Overall, this dissertation does not try to depict an ideal theory that addresses justice in an AI-integrated society.[22] Afterall, a future with AI can indeed be unpredictable, as shown in the alarming and optimistic views in the Introduction. Moreover, I believe that political philosophers are not necessarily in the best position to provide guidance on the direction of AI development compared to scientists and computer programmers, who have better knowledge of the current technological constraints. However, the least we can do as political philosophers is to remind the world about the ongoing wrongness that accompanies technological advancement and, hence, to steer the wheel of human progress back to a just course.

[22] The concept of ideal theory is provided by Rawls (1971, pp.8-9).

Bibliography

Attas, D., 2004. Who Owns the Production? *The Philosophical Quarterly*, 54(217), 537-556.

Bostrom, N., 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Brewster, T., 2019. Who's the Face Behind FaceApp? Meet the Rich Russian Who Built the Wildly Viral App. *Forbes*.

Available at:

https://www.forbes.com/sites/thomasbrewster/2019/07/24/meet-the-face-behind-faceapp-rich-russian-yaroslav-goncharov/#5771fe9d255a [Accessed: Aug 2019].

Cadwalladr, C., 2017. Revealed: How US Billionaire Helped to Back Brexit. *The Guardian*.

Available at:

https://www.theguardian.com/politics/2017/feb/26/us-billionaire-mercer-helped-back-brexit

[Accessed: July 2019]

Child, J, W., 1994. Can Libertarianism Sustain a Fraud Standard? *Ethics*, 104(4), 722-738.

Cohen, G, A., 1995. *Self-ownership, Freedom, and Equality*. Cambridge: Cambridge University Press.

Dworkin, R.,1973. The Original Position. *The University of Chicago Law Review*, 40, 500-533.

Frischmann, B & Selinger, E., 2018. *Re-Engineering Humanity*. Cambridge: Cambridge University Press.

Goodin, R, E., 2013. Disgorging the Fruits of Historical Wrongdoing. *The American Political Science Review*, 107(3), 478-491.

Himma, K, E., 2005. Information and Intellectual Property Protection: Evaluating the Claim that Information Should be Free. *APA Newsletter on Philosophy and Law*, 4, 3-9.

Kant, I., 2002. Groundwork of the Metaphysic of Morals. In L. Pasternack ed., *Immanuel Kant: Groundwork of the Metaphysic of Morals in focus*. New York: Routledge. pp.17-97.

Korinek, A & Stiglitz, E, J., 2018. Artificial Intelligence and Its Implications for Income Distribution and Unemployment, NBER Chapters, in *The Economics of Artificial Intelligence: An Agenda*, pp. 349-390. National Bureau of Economics Research, Inc.

Manyika, J *et al*., 2017. Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation. *McKinsey Global Institute*.

Available at:

https://www.mckinsey.com/~/media/mckinsey/featured%20insights/Future%20of%20Organizations/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Report-December-6-2017.ashx [Accessed: July 2019]

Moore, A, D., 2012. A Lockean Theory of Intellectual Property Revisited. *San Diego Law Review*, 49, 1069-1104.

- 1998. Intangible Property: Privacy, Power, and Information Control. *American Philosophy Quarterly*, 35(4), 365-378.

Murphy, J, B., 1993. The Moral Economy of Labour: Aristotelian Themes in Economic Theory. New Haven: Yale University Press.

Nozick, R., 1974. *Anarchy, State, And Utopia*. Oxford: Blackwell.

Otsuka, M., 2003. *Libertarianism Without Inequality*. Oxford: Clarendon Press.

Posner, E, A and Weyl, E, G., 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton: Princeton University Press.

Rawls, J., 2001. *Justice as Fairness: A Restatement*, E. Kelly ed., Cambridge, Massachusetts: Harvard University Press.

- 1971. *A Theory of Justice*. Cambridge Massachusetts: Harvard University Press.

Rose, J, L., 2014. Money Does Not Guarantee Time: Discretionary Time as a Distinct Object of Distributive Justice. *The Journal of Political Philosophy*, 22(4), 438-457.

Salian, I., 2018. SuperVize Me: What's the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning? *Nvidia*.

Available at:

https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/ [Accessed: July 2019].

Scanlon, T, M.,1998. *What We Owe to Each Other*. Cambridge Massachusetts: The Belknap Press of Harvard University Press.

- 1982. Contractualism and Utilitarianism. In A, Sen and B, Williams eds., *Utilitarianism and Beyond*. Cambridge: Cambridge University Press. pp. 103-128.

Sen, A., 2006. What Do We Want from a Theory of Justice? *The Journal of Philosophy,* 103(5), 215-238.

- 1999. *Development as Freedom*. Oxford: Oxford University Press.

Shapiro, I., 1991. Resources, Capacities, And Ownership: The Workmanship Ideal and Distributive Justice. *Political Theory*, 19(1), 47-72.

Silver, D *et al*., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484-489.

Simon, H, A., 1995. Artificial Intelligence: An Empirical Science. *Artificial Intelligence*, 77, 95-127.

Stone, P *et al*., 2016. Artificial Intelligence and Life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*. Stanford: Stanford University.

Available at:

https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf

[Accessed: June 2019].

Thomas, A., 2016, *Republic of Equals: Predistribution and Property-Owning Democracy*, Oxford: Oxford University Press.

Van Der Veen, R, J and Van Parijs, P., 1985. Entitlement Theories of Justice: From Nozick to Roemer and Beyond. *Economics and Philosophy*, 1, 69-81.

Walzer, M., 1983. *Spheres of Justice: A Defence of Pluralism and Equality*. New York: Basic Books, Inc.