Onward for the freedom of others

Marching beyond the AI Ethics

Petros Terzis[†]
Centre for Information Rights
University of Winchester
Winchester, UK
petros.terzis@winchester.ac.uk

ABSTRACT

The debate on the ethics of Artificial Intelligence brought together different stakeholders including but not limited to academics, policymakers, CEOs, activists, workers' representatives, lobbyists, journalists, and 'moral machines'. Prominent political institutions crafted principles for the 'ethical being' of the AI companies while tech giants were documenting ethics in a series of self-written guidelines. In parallel, a large community started to flourish, focusing on how to technically embed ethical parameters into algorithmic systems. Founded upon the philosophical work of Simone de Beauvoir and Jean-Paul Sartre, this paper explores the philosophical antinomies of the 'AI Ethics' debate as well as the conceptual disorientation of the 'fairness discussion'. By bringing the philosophy of existentialism to the dialogue, this paper attempts to challenge the dialectical monopoly of utilitarianism and sheds fresh light on the -already glaring- AI arena. Why is 'the AI Ethics guidelines' a futile battle doomed to dangerous abstraction? How this battle can harm our sense of collective freedom? Which is the uncomfortable reality that remains obscured by the smoke-gas of the 'AI Ethics' discussion? And eventually, what's the alternative? There seems to be a different pathway for discussing and implementing ethics; A pathway that sets the freedom of others at the epicenter of the battle for a sustainable and open to all future.

CCS CONCEPTS

Computing methodologies~Philosophical/theoretical foundations of artificial intelligence • Social and professional topics~Codes of ethics

KEYWORDS

Ethics, artificial intelligence, algorithms, existentialism, philosophy

ACM Reference format:

Petros Terzis. 2020. Onward for the freedom of others: Marching beyond the AI Ethics. In *Proceedings of ACM Conference on Fairness, Accountability, and Transparency conference (ACM-FAT* 2020). ACM, New York, NY, USA, 10 pages.*

1 Introduction

The 'Ethics of AI' has conquered industry news and academic literature at least for the last 15 years. Promising endeavors by public institutions were cemented in pages of Guiding Principles for Ethical Artificial Intelligence and at the same time companies tried to take the lead in promoting ethical practices by crafting strategies or by setting up ethical boards. Within the same realm, a distinct community emerged focusing on how to pragmatically and technically embed our ethical commitments and values into the systems we build.

Scholars and journalists have criticized the bigoted emphasis of industry leaders, academia and policymakers on the production of 'AI Ethics', guidelines and policies. In the AI Now Institute's Report [4] it was stated that: 'ethical codes meant to steer the AI field should be accompanied by strong oversight and accountability mechanisms', while Julia Powles and Helen Nissenbaum in a widely cited blog post [17] warn that too much attention and resources in 'fixing' AI eventually leaves no room for discussing whether AI should be used at the first place. Hence, the prioritization of external and non-self-regulatory mechanisms along with the necessity of a broader vision on the social utility of AI, are the principal arguments on the insufficiency 'ethics' and the dearth of the fairness discussion.

This paper echoes some of these critics and attempts to provide a new perspective into the analysis of the current state of affairs in the field of AI and Ethics. Firstly, it aims to challenge the dialogical monopoly of the 'ethical guidelines' and suggests a different approach for the fairness discussion by confronting the conceptual grounds and the philosophical foundations on which these two research areas have been built. Secondly, this paper contributes to the 'ethics debate' by recommending a different philosophical perspective to ground and promote the discussion on and application of ethics in the real world. This new perspective is rooted in the ideas of existentialism as expressed

[†]PhD Student, Centre for Information Rights, University of Winchester

FAT* '20, January 27–30, 2020, Barcelona, Spain

ACM ISBN 978-1-4503-6936-7/20/01…$15.00
https://doi.org/10.1145/3351095.3373152

in the work of Simone de Beauvoir and Jean-Paul Sartre. References to other representatives of this philosophical realm are found throughout the paper.

Emanating from this philosophical foundation, some of the issues to be explored in this article are: How did we end up with so much abstraction? What are some of the facts we do not see every time we –in good faith- contribute to the AI Ethics debate? Which is the most profound and uncomfortable reality that is obscured by the smoke-gas of the 'AI ethics' discussion? Why allocating our resources in 'de-biasing' algorithms can be proven a Sisyphean task? And finally, what's the alternative for building ethical systems authentically loyal to social good?

Beauvoir and Sartre, albeit their sometimes different approaches, were prominent figures of existentialism. Their work could be perceived as supplementary of one another. Sartre deals with the nature of existentialism; the ingredients of what it means to *live authentically.* Beauvoir, on the other hand, approached the ethics of existentialism; what it means to live authentically and act in the world around you as a good human being. As existentialists, neither of them accepted the universality of any commonly shared values nor acknowledged any principle captured as 'given', somewhere in the abstract world. And this distinct characteristic differentiates them from Nietzsche whose ideas inspired them both. Nietzsche's writing for the 'death of God' in *Zarathustra*, challenged the 'free spirits' and called them to assume the responsibility of crafting a divine teleology to create 'life-affirming moral and life-enhancing aesthetic values'[5]. Now that God is dead, human-beings are alone. In his own words: 'Today [...] when only the herd animal is honored [...] the concept of 'greatness' entails being noble, wanting to be oneself, being capable of being different, standing alone and having to live independently'. Contrary to this interpretation that nurtures lone individuals who cultivate their ethics in solitude before embarking on the task of imposing them on others, Beauvoir and Sartre reformulate the discussion by bringing the freedom of the self and the freedom of others in the epicenter. For them, ethical principles and guidelines, either crafted by those in solitude or given by God, are '*tabula rasa*' unless they recognize the predominance of the freedom of the individuals and the burden of responsibility that they carry –admittedly- with anguish and risk, when they offer themselves in the endeavor to lead an ethical life.

2 An existentialist critique on ethics

2.1 The freedom of the subject

'Your existence precedes your essence' is perhaps the most popular Sartrean axiom[14]. Coupled with his remarkable lecture entitled 'Is Existentialism a Humanism?'[15], these elements constitute the essence of his philosophy. 'In choosing anything at all, I, first of all, choose freedom' [15] he writes.

Sartre advocated that there is no prescribed reason dictating our being and that there is no recipe on how to pursue our life as good and decent human beings. Contrariwise, the argument continues, the -sometimes inconvenient- truth is that we are all free. We are free to decide what to do and what to become. Free to choose whether to give or take, play or leave, call or pass, comply or disobey. If we decide to conform to certain norms or follow specific guidelines it is us who freely choose to do so. And if we act as people expect from us to act this is again a result of our free choice as free individuals. Nothing precedes our choices, neither the fictional state that we are used in calling 'character', nor some abstractly defined norms crafted by someone's God.

Existentialism is by no means founded upon the demonstrations of the non-existence of God. It declares, Sartre writes, 'that even if God existed that would make no difference from its point of view. Not that we believe God does exist, but we think that the real problem is not that of His existence; what [man] needs is to find himself again and to understand that nothing can save him from himself, not even a valid proof of the existence of God. In this sense existentialism is optimistic'. By 'God', the reader should define any dogma, doctrine or theory that comes with an objective, impossible to appeal context as a given truth. In that perspective, objectivity, scientific justification and rational reasoning are becoming an inherently limited process that inhibits our journey towards disclosing our true selves. 'When I deliberate... the chips are down' Sartre explains in that logical argumentation is only a small piece of the puzzle of our subjectivity for it cannot disclose the deeper emotional instincts that guide our choices and subsequently, our lives.

Such a disclosure, Beauvoir adds in the 'Ethics of Ambiguity' [2] 'implies a perpetual tension to keep being at a certain distance, to tear oneself from the world, and to assert oneself as a freedom. To wish for the disclosure of the world and to assert oneself as freedom are one and the same movement'. Freedom, thus, becomes the foundational condition of our existence and those who want freedom must want it *categorically*, *unequivocally* and *universally.* Linked with and emanating from our existence, our freedom has nothing ready-made in it. It is not illuminated by an abstract sphere of pure morality nor does it emerge from uncontested principles set by others or God. Our freedom consists of the values it establishes and through which it is fulfilled. Beauvoir summarizes it as follows: 'To will oneself moral and to will oneself free are one and the same decision'[2].

The same goes, an existentialist would continue, for any attempt aiming at crystallizing value-laden principles in paper. In that sense, the 'AI Ethics' discussion is a genuine fallacy for it does not address one of the most fundamental issues in the field of advanced computational technology: the freedom and the subjectivity of all the agents involved, be it the CEO of a tech-giant, the project manager, the business analyst, the developer or the micro-worker. On the contrary, all the institutions involved in the deliberations for crafting these principles, albeit their distinct point of departure, are trying to cement principles and guidelines which will end up as 'given' to the world. Hence, by ignoring the limits of their subjective choices, they pretend that the irrefutable value of their products (or objects) is somehow rooted and certified in them. To exacerbate the issue, the lack of reference to the subjectivity of the individuals present and acting in the arena of AI, is supplemented by a trend of carpeting this subjectivity under a fictional personification that treats the already elusive 'Artificial Intelligence' as a distinct, separately recognizable agent. Indeed, most of us, do not talk about the 'Ethics of the AI Companies' let alone the 'Ethics of the AI's CEOs and/or developers'. Instead, we are building an additional layer of protection for the latter's realization of freedom of choice and responsibility by inventing imaginary subjectivities. With our project in itself being so abstract and elusive how do we expect the commonly agreed values to become something more than a potential hindrance from our acceptance of freedom and responsibility?

Similarly, although the discussion on machine learning techniques and algorithmic design are useful as they provide developers with something to contemplate and reflect upon, the scientific community should not treat these techniques as a panacea for building fair applications. We should always bear in mind, an existentialist would argue, that being 'fair' and 'ethical' are not fixed elements granted to a subject with certification-based procedures and compliance exercises, but instead, they consist deliberate choices according to which the individual shall lead its life the way it chooses to do so. In this context, machine learning techniques might indeed provide effective tools to help developers reflect on their design choices. However, there exists the risk of using these techniques as a 'ticket' to evade the burden of responsibility, a phenomenon that Sartre referred to as 'bad faith'.

2.2 Bad faith

To understand the Sartrean concept of 'bad faith', we first need to introduce ourselves to the duality of our existence as described by Sartre and to further dichotomize our being into our *being-in-itself* and our *being-for-itself.* The latter denotes 'how things are' or 'how things are given'. It is the objective sphere of *facticity*, the 'given world' within which facts like our physical characteristics, our talents, our neighbors, our past or our scheduled trajectory, solidify our current situation. On the other hand, *being-in-itself* is procedural. It is a transition, a journey that ends when we realize our thesis in this given world. It is the sphere of transcendence that is fulfilled by the time individuals raise their consciousness to the understanding of their subjectivity within a particular given situation. This transcendence is eventually our realization of the freedom we enjoy and the responsibility we shoulder. Everything that stands between these two spheres is corrosive of our transcendence.

'Bad faith' is in this context a wall intentionally built by ourselves that separates our duality and inhibits our transcendence. It is a way of willfully blinding ourselves from seeing the space within which we have to raise our consciousness for realizing our freedom; a way of running away from the conceptualization of our responsibility. 'Bad faith' stands for everything that constantly leads us to the passive and comfortable acceptance that 'it is what it is'. Sartre aptly describes this by witnessing a waiter in a café going from one table to another, carefully holding his tray and leaning forward politely when listening to customers. The waiter is in bad faith when he accepts these attitudes and his 'waiter' identity as something natural from which he cannot escape.

To land it in our argumentation, an existence in 'bad faith' refers to the individuals, be it industry leaders, lobbyists, or developers, who hide themselves behind the curtains of the others' expectations or the disguise of the corporate environment in order to evade the anguish and the risk of realizing the burden of responsibility that they pragmatically shoulder. A lobbyist in a suit who strolls down the hallway of the European Parliament, with his iPhone always at hand, messaging automatically without even looking at the intern upon whom he stumbles to receive a coffee stamp on his shirt, is no different than the Sartrean waiter in the café who remains restricted to other people's expectations and thus, in bad faith. Similarly, companies' CEOs who preach general ethical principles while at the same time follow the 'business necessity' of competing with other similar preachers in suspicious use of AI technology because 'we need to move fast', circumvent from standing up to the realization of their responsibility and they are doing so by using the fictional concept of competition as a pretext. As a result, generally expressed norms, abstract formalities, and illusive statements are thus being transformed into tickets for escaping the transcendence beyond our given situation and, eventually, lead us to accept that 'this is just the way things are'. Thus, life in bad faith delays our transcendence and the realization of our freedom thereby perpetuating our pernicious swirling in the cage of our given world.

2.3 Ethics are inherently ambiguous and human-beings inherently biased.

Have you ever found yourself in a challenging conversation with a group of friends, where one of the group calmingly and persuasively says: 'Look, you know I am objective and impartial on this, right?'. Well, if Beauvoir was among your group of friends that day, she would probably interrupt your friend only to say 'No, you are not, but please go on'.

Because Beauvoir believed that objectivity is inhuman and every effort to claim otherwise is nothing but a biased assertion beclouded by our denial to realize our subjectivity. According to Beauvoir: '[the objective man] does not have to choose between the highway and the native, between America and Russia, between production and freedom. He understands, dominates, and rejects, in the name of total truth, the necessarily partial truths which every human engagement discloses'[...]. Lost in the illusion of objectivity, the 'impartial man' denies the subjectivity of his judgment rendering himself a 'shameful servant of a cause to which he has not chosen to rally'. By supporting the objective truth, the argument continues, he places his 'object' as the only truth.

However, Beauvoir did not consider our inescapable bias to be an inferior characteristic of our nature as human beings. On the contrary, she believed that the conceptualization of our biases within a world full of biased people is the first step for our transcendence towards our subjectivity as free individuals. Whereas, declining our biases and preaching our objectivity is a characteristic of a self-imposed tyranny that will eventually seek to impose itself on the outside world.

Today, the utilitarian argument that perceives technology as something intrinsically neutral has become 'common sense'. Of course, there is nothing wrong with the commonly used sophism that regards technology as a tool that can be used both in good and bad ways. However, it is by no means self-evident that, due to this indisputable fact, the creators of a certain tool are themselves neutral. Of course, a gun can be used both by an individual who targets innocent people and by a police officer who shoots the killer in the leg. However, the ontological existence of both possibilities does not equate with neutrality. For gun manufacturers would fool themselves if they camouflaged their deliberate choices under the argument of the ex-post transformation of the objects they put in the market. Because, by manufacturing guns, apart from getting different pieces of metal together, they simultaneously make the value-laden statement that, sometimes, killing people is good[1].

Similarly, a research scientist that tries to design technical parameters for fair and transparent use of facial recognition by the police, implicitly acknowledges that sometimes mass surveillance of our biometrical characteristics in public places is good. To deny this subtle but emphatic articulation of our responsibility under the disguise of a hypothetical and unsubstantiated notion of ‘neutrality’ would be bad faith according to Sartre; Beauvoir would perhaps use the word ‘tyranny’.

Irrespective of our acceptance or not of the existentialism as philosophical background, Beauvoir’s ideas on ethics and human biases, as expressed mainly in her essay ‘The Ethics of Ambiguity’[2], are penetrating for they can stimulate our thinking towards both the current state of affairs in the technical realm of machine learning as well as the general ‘AI ethics’ debate. It’s worth mentioning at this point that Beauvoir, talking about the ‘Ethics of Ambiguity’ twenty years or so after its publication, was explicitly angry with her-then-self. In her autobiography ‘The Force of the Circumstance’[3], she confesses: ‘My descriptions of the nihilist, the adventurer, the esthete, obviously influenced by those of Hegel, are even more arbitrary and abstract than his, since they are not even linked together by a historical development; [...](emphasis added) I was in error when I thought I could define a morality independent of a social context’. This paper echoes this later self-review of hers; Morality is explained as a concept inextricably linked with the social context from which it is engineered and to which it contributes. In that sense, existentialism –or at least our existentialism- is non-individualistic.

2.3.1 Algorithmic de-biasing. First of all, Beauvoir’s argument around the unavoidability of human biases touches upon the research efforts to de-bias algorithmic systems and datasets. To accept the innate biases of human beings is to accept that everything we are dealing with is an articulation of biased subjectivities and that those who are called to de-bias the systems are themselves biased. The reference in our inextricable ambiguity, however, is not an argument to diminish the value of our efforts. Contrariwise, the moment we realize the innate biases in every human-being, ourselves included, is the moment we fully conceptualize two vital elements for our endeavor.

The first one is a warning, highlighting that the efforts to de-bias a system do not emanate from our tendency to be objective; by building or calling for measures for fairness, transparency or accountability, humans do not attempt to render algorithms objective. What they aim, is the building of systems that are biased towards what they perceive as optimal. For example, by presenting a model for fairness at a global conference, a researcher does not claim that she wants to build fair systems. What she declares, is that she wants systems to be fair according to her subjective interpretation of the word 'fair'. This may not be a problem in certain conferences where what is optimal for me might as well coincide with what is optimal for my next-seating participant. But things will become more complicated when our 'optimalities' stumble upon other people's ones. What is 'fair' or 'transparent' for me, might not be the same with what a privacy advocate or a CEO of an AI company considers as such.

The second assertion that derives from our acceptance of the unavoidability of human biases, is even more telling. For accepting that bias is inherent in every human being and dataset, which are also inherently and inextricably biased as they have been based on human decisions, generates a thorny question: Are we content with using machine learning systems to produce seemingly objective but inherently subjective outcomes? Are we happy by surrendering all the elements that we cannot change for they constitute our 'given world' (our height, our neighbors, our date of birth, our skin, our friends, our past calls, our past convictions, etc) to a system that will use biased assertions in an attempt to lodge other people's subjectivities and expectations to our sphere of 'given'? For example, if I know that all police data are -and will always be biased- and if I do not believe that we will ever reach a point where police officers will make unbiased decisions, why should I even start the discussion of de-biasing them? And if I believe that my facial characteristics, which are part of the things I cannot change should not be a basis for other people’s subjective and biased assertions about my subjectivity, wouldn’t de-biasing be a challenge inherently futile? Why should I speak the same language with people whose ends are elusive and take part in endeavors whose goals are unattainable? Beauvoir writes: ‘I was — like Sartre — insufficiently liberated from the ideologies of my class; at the very moment I was rejecting them, I was still using their language to do so. That language has become hateful to me because, as I now know, to look for the reasons why one should not stamp on a man's face is to accept stamping on it’[3]. Ultimately, does this negation render every attempt to deal with data meaningless?

[1] This example was introduced by Matthew Beard, Associate Lecturer at the University of Notre Dame Australia and a Fellow at the Ethics Centre, during his keynote speech at the Ethics of Data Science Conference 2019, University of Sydney. I hope I transferred it accurately.

By no means, I would answer. For it seems that there is a qualitative difference between a system that estimates the risk of breast cancer and another that predicts the likelihood of re-offending. In the first case, the subject is dealing with a system that estimates the possibility of an event that will be (or not) part of its future 'given world' (of the objects that cannot be altered); this system eventually delivers an X percentage that could be expressed as follows: 'Today your *sphere of all the things you cannot alter,* changes (because an X% of breast cancer has been added) due to the fact that your future factual situation (your status as being diagnosed or not with breast cancer) is likely to change'. Whereas, in the case of a system that predicts the likelihood of re-offending, the subject is confronting with an event that is likely (or not) to occur in its future sphere of subjectivity (its free choice to re-offend) and which, although produced by a system based on inherently biased subjects and despite the event's linkage with its future being, is injected into its sphere of present 'given world' like a prophecy that its doomed to happen. The system, in this case, says: 'Today your *sphere of all the things you cannot alter*, changes (because an X% of re-offending has been added) because your future subjectivity (whether you will freely choose to re-offend) is likely to change'. And granted that we are all biased, this process of objectification of other people's subjectivities in a subject's sphere of the 'given', must seem alarming for it obscures the subject's present situation by interfering with its future transcendence. Hence, if we accept the unavoidability and omnipresence of bias within human beings, dealing with de-biasing systems that, although emanating from things we cannot change, end up producing subjective assertions about our future subjectivities, engenders an existential anomaly.

2.3.2 The hypocrisy of the ethical person. Ambiguity lies at the heart of our existence and this is by no means a reason to reject it. To disclose our truth is to transcend to our subjectivity, renounce the very possibility of objectivity and embrace the challenge of seeking our biases. Think of it as a mindful process towards self-awareness. An honest discussion with ourselves which repudiates the 'non-judgmental' stance that traditional mindfulness promotes. This mindful journey towards the realization of our biases demands the confrontation of our freedom; It is a ladder towards our transcendence and, as such, it is a process paved in a road of anguish and risk; Individuals that walk this ladder, understand that being ethical -or whatever else you claim to be- 'is not a matter of being right in the eyes of a God, but of being right in their own eyes' as Beauvoir says[...]. By rejecting the illusion of a refuge built to host our consciousness and guarantee our existence before our transcendence, individuals will also refuse to believe in ready-made and unconditioned values whose aim is to objectify and fossilize someone else's object as truth.

Had Beauvoir lived the dawn of the AI era, she would be rather skeptical about those people or groups of people who hold themselves to the world as 'devoted to our well-being' or others who 'value our privacy'. This is because Beauvoir did not believe in anyone's absolute aim. On the contrary, she constantly warns her reader about those people that insist on the objectivity of their aims. She believes that this category is an example of people who, caged in their 'given world' (*facticity*), have avoided the anguish and risk of their transcendence. By doing so they became a kind of tyrant, for tyrants, as well, rest in the absolute of their aims and usually preach their objectivity as something emerging from other abstract objectivities such as God, the legacy of a nation or someone's character. Conversely, the free man does not look for a priori certified values, thoughts, and aims; the free individual, Beauvoir continues, will keep asking themselves 'Am I really working for the liberation of [men]?'.

In this context, it is illusionary to believe that sailing in the high seas of ethics is a journey destined to reach a certain port where everything is ethereal. Even worse, in case we have a captain on board who keeps turning the rudder towards this hypothetically ethereal destination, then we are doomed to paddle endlessly and vainly. Seeking ethics is a journey whose ends are neither predefined nor even defined. The honest and good captain would draw a map by putting destinations in parentheses because to name a destination as 'optimal' is like falling into a naïve, uncontested abstraction. Beauvoir describes this aptly: 'We don't ask the physicist, "Which hypotheses are true?" Nor the artist, "By what procedures does one produce a work whose beauty is guaranteed?" Ethics does not furnish recipes any more than do science and art. One can merely propose methods'.

As a result, demonstrating your moral goal or building roadmaps with bullet-points on how you plan to become ethical is a hollow attempt neither worth your time nor the public's attention. The only viable solution would be to formulate clear and persuasive techniques for safeguarding and deepening your commitment to being ethical, be open about them and test them repeatedly.

2.3 Agreeing on ethics for all is an illusion

The following remark, influenced by the impact Marxism had on Beauvoir, might seem rather unsettling to the reader but it would be disrespectful to omit it for it lies in the heart of Beauvoir's ideas on ethics. As she explains: '[man] is not alone in the world; different [men] have different forms of well-being; to work for some [men] is often to work against others; one cannot settle on this peaceful solution: to want the well-being of all [men]. We must define our well-being. The error of Kantian morals is to have pretended to do without one's presence in the world; in this manner it only ends in abstract formulas; the respect of a human being, in general, is not enough to guide us; for we have to deal with separate, opposing individuals: the human being exists in his entirety in both the victim and the executioner; are we to let the victim perish or kill the executioner?'

Beauvoir believed that oppression divides the world into two clans: from the one side, there are those who present themselves as defenders of certain values with an authority deriving from a natural, undisputable source of morality, be it God, the nation or something else. On the other side, there are those who 'mark time hopelessly to merely support the collectivity' whose life ends up being a mechanical repetition. The oppressor gets stronger by keeping the oppressed imprisoned in their 'given world' and presents this anti-transcendence as something natural against which any revolt is purposeless since nature does not err. We do not necessarily need to adopt the argument that divides people between oppressors and oppressed or between victims and executioners to appreciate the relevance of Beauvoir's thoughts in our discussion. For there always lies the danger of subjects leaping into the immanence of their 'given world' guided by sophisms about principles that other 'serious people'[2] are tasked to safeguard. These 'serious people' are well aware of this sophism and instead of constantly striving for freedom and responsibility by asking themselves 'am I doing the right thing?', they prefer presenting themselves as the defenders of certain values. Beauvoir [2] describes them as follows: 'It is not in their own name that they are fighting, but rather in the name of civilization, of institutions, of monuments, and of virtues which realize objectively the situation which they intend to maintain; [...] they defend a past which has assumed the icy dignity of being against an uncertain future whose values have not yet been won'.

An 'AI Ethics' roundtable usually includes academics, government officials, policymakers, researchers, lobbyists, workers' representatives, human rights advocates, and journalists. Every participant has her preconceptions of what it means to be ethical. Objectifying these diverse approaches to 'set them in stone' can only be achieved through abstraction. This agreement on abstraction is what Beauvoir considers to be illusionary, for you cannot agree on a common basis when your points of departure are diametrically distinct; as for example when you embark on this challenge by having 'red-lines' while at the same time the industry leaders have only 'critical concerns'[11]. Furthermore, this abstraction is dangerous in itself if it is to be perceived as a natural procedure that is inspired by the ethical character of those that set it up. Because by signing a report for 'ethical AI', one implicitly reaffirms the morality of all the signatories. In the end, we are left with an illusion of agreement, endowed with a sense of a *natural* consensus on what is ethical and as such, it cannot be contested. After all, who could disagree with the phrase: 'Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases [13]'?

3 Towards a transcendence to authentic ethics

By viewing the 'AI ethics' discussion through the philosophical lenses of Beauvoir and Sartre, I by no means aspire to condemn the relevant initiatives undertaken in good faith by prestigious democratic institutions. Although I accept such a potential result as a direct consequence of my reference to the above school of thought, I am afraid that the primary message of this article will be ignored if it is to be perceived as a mere critic of the current status-quo. This article aims to illuminate the 'AI Ethics' arena from a different angle and challenge the philosophical monopoly of utilitarianism that pervades the 'given world' of technology. Its goal is to help the reader realize the 'given's' potential to annihilate human freedom if it is to be perceived as a one-and-only solution and, finally, to understand the process towards surpassing this deceptive sense of inevitability. This article aspires to model an intellectual arsenal to challenge the game of ethics as it set today, and calls the reader to avoid it, or accept it and follow it as long as she/he freely and consciously chooses to do so.

An alternative framework for building ethical AI would be founded on the subjectivity of the individuals involved in the development process. This framework, free from 'objectively' ethical criteria, will take the form of a reflective process during which the individuals will obtain a clear understanding of their freedom and responsibility; At the end, they will eventually be judged according to their choices. Similar alternative frameworks based on relevant philosophical grounds have also been proposed for business management [9] and architectural design [1].

The following section tries to shape the landscape for this alternative framework.

4 A proposal

Despite our focus on the subjectivity of the individuals, it would be delusive and contradictory to our non-individualistic view on freedom, to assert that individuals alone and by themselves can achieve their morality and that eventually someday all humans will be moral and free. Freedom and morality are meaningless unless they are to be enjoyed among people who are free and moral. Hence, the ethics of the individuals in the AI world is influenced by the architects of the ecosystem within which they enact, in our case the corporations. Without their contribution, any individual attempt to 'live ethically' will be heroic but eternally insufficient. As a result, the proposed framework is divided into two parts in order firstly to suggest *quasi* institutional changes for the corporate environment and, secondly, a roadmap for those engaged in this world to conceptualize their authentic selves.

4.1 The ethics of the business models

Beauvoir writes: 'As we have seen, my freedom, in order to fulfill itself, requires that it emerge into an open future: it is other men who open the future to me, it is they who, setting up the world of tomorrow, define my future; but if, instead of allowing me to participate in this constructive movement, they oblige me to consume my transcendence in vain, if they keep me below the level which they have conquered and on the basis of which new conquests will be achieved then they are cutting me off from the future, they are changing me into a thing'.

Our transcendence towards freedom should be perceived as an activity nurtured by those who fashion the world in which we act; a 'constructive movement' whose success is conditioned on my participation. With this in mind, companies should not even

[2] This term is used by Beauvoir to refer to those present who present themselves to the world as defenders of certain values.

consider drafting codes of ethical conduct for their employees, let alone undertake initiatives for global 'AI Ethics'. 'AI Ethics' is not a standalone enterprise decoupled from the rest of the business ecosystem. For this reason, companies that want to build 'ethical AI' will demonstrate a loyal and constant commitment not to build, but to live ethically; ethically towards their employees and their clients; and ethically towards their goals for a better and sustainable future. There is no 'ethical AI', unless there is an 'ethical' supply chain, clean from conflict materials; an 'ethical' UX/UI design, free from manipulative dark patterns; an 'ethical' workforce, comprised of ethical micro-workers; an 'ethical' decision-making process, conditioned on the parameter of climate justice; and eventually, an 'ethical' reconsideration of the 'data-driven' necessity.

The shareholder-centric approach that dominated the economic culture of our world for over a century, had a severe impact on the profit formulas of digital businesses. Being 'data-driven' has now become the new 'common sense' for our economy; a Gramscian 'hegemonic culture' [6] which ended up being regarded as natural. But this relentless strife for accelerating the 'data-drive' always comes with trade-offs and ethical compromises. For this reason, an authentic commitment to ethics will require business model reconsiderations.

Such a process will begin by institutionalizing the appropriate mechanisms for building awareness on what the company is doing at any given moment. Instead of declining to comment, companies will be open to public scrutiny and as subjects by themselves, they will be critiqued by their employees and the public based on their choices, and not by their motives or declarations. Ethical companies will abandon dishonest and bad-faith attempts to mask the range of available options by putting them below the carpet of some fictional inevitability; they will assume responsibility for every choice they make. A Sartrean existentialist, for example, would not require a tech-giant to give away the data it uses for developing its systems. Instead, she would demand from the company not to hide this choice behind bad faith arguments such as proprietary rights. As long as you freely and consciously choose to lock your treasure away from public investigation and research, Sartre is happy[3]. Pretending you have lost the keys is, however, problematic.

Similarly, companies that want to suffer the anguish and risk of becoming ethical will be called to cultivate an open ecosystem for their employees; a business environment that is more democratic and open to dialogue and inquiry. In such an environment, employees will be encouraged to question the moral choices of the company and debate ethical issues that they encounter at the workplace. They will not be kept in the dark about the initiatives of their employer and there will be no secret teams working in mystifying branches. A new model that wishes to inspire morality will be a business ecosystem where individuals will enjoy a zone of autonomy and will be incentivized to understand and embrace full responsibility for their actions as a prerequisite for enjoying a democratic environment.

But what are, someone may reasonably ask, the nature of this responsibility and the burden of this freedom? Indeed, if left vague, the argument would risk to contradict itself.

This brings us to the second pillar of the suggested framework; How individuals who act and design in the AI world could fit in such a new business environment? How they, in turn, undertake the onerous task of pursuing an ethical career?

4.2 Towards an authentic self-awareness

4.2.1 Acceptance of freedom The pathway towards the conceptualization of our true freedom is a reflective process to comprehend and disclose: 1) all the 'objects' that constitute our 'given world and cannot be altered; 2) the potential elements that could lure us in evading our responsibility thereby inhibiting our transcendence to the awareness of our freedom and 3) the ontological possibility of every choice that we have at any given moment.

For example, picture a data scientist that has undertaken the task of building a job-recruitment tool. Given the strict deadline, she has limited time to explore the datasets with which she will feed her system. As a result, she cannot exclude the possibility of discriminatory results. In understanding her situation, the data scientist will first have to identify the boundaries of her 'given world'. She will need to be aware of all the elements she cannot change i.e. the deadline, her working hours, her family responsibilities or the luring bonus. Secondly, she has to take ownership of her actions. The deadline, her work or family are nothing more than pieces of a puzzle that altogether shape her situational facticity. She can neither change them nor can she utilize them to escape the transcendence process towards undertaking full responsibility for her final choice and actions, because this will be bad faith. Thirdly, and perhaps more importantly, she must acknowledge the ontological possibility of every choice. For example, she could do whatever possible given the resources at hand or ask to work extra hours. She could as well mitigate some of the system's errors after the delivery of the system or disclose its weaknesses in due time and risk authorship of and bonus from the task because 'someone will do it'. She could finally blow the whistle and join the emerging tech-resistance. Whatever she eventually decides to do, it will be her choice. Accepting it is her first step to transcendence.

Heidegger believed that to exist is to be historical [8]; to understand my existence, I must set it side-by-side to *something* that defines my historical heritage. For Heidegger, every time I act, I take part in a narrative unity that recollects the past to give meaning to the present; and 'when I choose, I exemplify a standard for the others as well'[18]. On the same realm, Marcuse, considers this historical dimension along with the operational rationality to consist the two opposite dimensions of our universe: 'The suppression of [the historical dimension] in the societal unity of operational rationality is a *suppression* of history, and this is not an academic but a political affair' he writes[10].

[3] Beauvoir is not and we will soon see why.

Today, there are those developers who are ‘just engineers’ and whose job is not ‘to take political stances’ because they are afraid of ‘letting the perfect be the enemy of the good’[7]. This category of developers exists as a narrative continuation of all those coders that have historically perceived their job as the apotheosis of functionality: the strife to reach an objective ends even if you have to objectively reformulate our problem along the way[12]. For them, the prioritization of functionality defines their fate, not deterministically, but by providing *something* that they inherited from their historical situation to claim them and understand their existence. They are *developers* because thousands of developers over the last decades have shaped a relevant standard of what it means to exist as *a developer.*

Today, this very historical heritage is challenged by a distinct historical block of developers who have moved from the safe ‘we just build’ to the creative and (r)evolutionary ‘we won’t build’; these developers reject outright the dominance of operational rationality; for these developers, before wondering *what* or *how* to build, you need to ask *why*[16]? What this historical alternative fundamentally declares is that ‘your freedom as a developer expands well beyond your professional status-quo’. Developers do not just build products anymore. They fashion experiences and through them, they create futures. A developer of this block exists not only to build but also –and perhaps fundamentally- for not to build. She/he ultimately becomes a new ‘*something’* for other developers to reflect on and claim their existence, thereby challenging the traditional 'common sense' of what it means to exist as a *developer.*

Each one of these historical blocks comes with different conceptualizations of freedom but irrespective of how we might feel about our role in the world, there are responsibilities we cannot evade.

4.2.2 Acceptance of a universal responsibility In his seminal public lecture entitled ‘Existentialism as Humanism’[15], Sartre shaped the notion of such a responsibility. Admittedly, if the latter is to be perceived holistically, it might seem frightening. It is of no surprise that he considered a life in which we experience our responsibility, as a life full of anguish; an anguish rooted in two equally daunting conceptions.

Firstly, it derives from the realization that we are responsible for what we do and there is nowhere we can lay this burden other than our shoulders. No abstract principles, no inescapable present situations, and no inevitable future outcomes; There are not strange powers and there is no God. As Beauvoir further remarks: 'A God can pardon, efface, and compensate. But if God does not exist, [man's] faults are inexpiable'. Everything we think or say contrary to this is nothing but a bad faith attempt to hide, mask our responsibility and escape our transcendence-again in vain.

Yet the second source of the same anguish might perhaps seem more intimidating to those aspiring to lead an ethical life. Because, even when I accept my full responsibility for everything I choose, I have to acknowledge that what I do in my life is an implicit statement of what everyone else ought to do. The Sartrean idea of *universalism* is the first compass for the ethical well-being of the individuals. It is a powerful message saying: ‘act as If the entire world is watching you and is waiting to mirror your choices’. It is an ‘Onward’ call that invokes our subjectivity and urges it to canvass the image of the world it dreams of living in. A message that renders every choice we make an example of the life we want to pursue.

To illustrate this, think of a scientist that presents her machine learning model on how to embed fairness into algorithmic decision-making systems at the technical track of a global conference. Firstly, she understands that the guarantees provided by statistical techniques for her proposed model do not absolve her of her responsibility to accept the outcomes of her model's failures and unintended consequences. She understands that her responsibility extends well beyond theoretical guarantees. Second, and more importantly, by proposing her model, she understands that she is not just providing a mathematical model but, in fact, proclaiming to others that her model leads to a better outcome for all the people involved including, but not limited to, data subjects. Her responsibility begins, not ends, with the model. Through her model, she is now responsible to society at large[4].

4.2.3 Act and design for freedom 'OK, I am now a free man and I know my universal responsibility. What's next?' someone would reasonably wonder. Indeed, Sartre gave us the compass of universalism to illuminate the normative context of our responsibility but has left the respective context of freedom on a subjective and formal basis. 'One of the chief objections leveled against existentialism is that the precept "to will freedom" is only a hollow formula and offers no concrete content for action', Beauvoir admits. 'You are a free man', Sartre awkwardly answered to a student who asked for his advice on the existential dilemma he was facing whether to join the Resistance or stay home to help his mother. 'But what am I supposed to do with my freedom, sir?' I wish he asked. He did not and thus we missed a chance to listen to the Sartrean teleological perspective on freedom. Beauvoir filled that gap. And, to my perspective, her ideas on the *telos,* the scope of freedom are valuable not only for the AI community but for the entirety of our world.

Similar to the child’s example in the ‘*Ethics of Ambiguity*', when we connect to the Web, we cast ourselves to a universe we have not helped to establish; a universe that has been fabricated without us and to which the only thing we can do is submit. In such a world, data are assets, cookies are necessary, advertisements are inevitable, the 4G network needs subscription fees, social media are connecting people and private companies profit from our inputs. These are all given facts that seem to us as inevitable as our breath. They are always there -and they will always be- waiting patiently for our connection and submissions. Like children in a world of adults, we do not

[4] I am grateful to the first anonymous reviewer for her/his contribution specifically on -but not limited to- this point. The example of the scientist in this paragraph is copied by her/his comments on another example I had used to illustrate the universal responsibility of the scientist. The comments provided during the review process made me realize that my initial perspective was narrow.

want to change anything and we do not even think of the possibility of doing so. We are happily naïve about staying always connected to this chaotic universe of collective irresponsibility.

The inconvenient truth is that most of us will keep feeling that way until the day we die. Just like the child that casts itself in a world of adults where it does not even think of the possibility of intervening to change the order of things, we, the children of this new world, will spend our digital life exercising our freedom only within the given boundaries that we found upon connection and according to those rules that we abide by when entered; or other rules that will be fashioned for us –again- by others. We have yet to raise ourselves to the consciousness of this enthrallment; Yet, there exactly lies the context for action for all of those who dream for an open future.

The most powerful message that Beauvoir conveyed through her work is her view of what it means to will freedom. According to her, no freedom can be enjoyed in solitude. There is no individual freedom severed from the rest of the world. To will freedom, to want other people to explore their options for liberation and, subsequently to will all of them to be free is one and the same thing. To will freedom is to will the freedom of others.

In our emerging digital world, no one will have the luxury of feeling free in the solitude of her computer. Her cookies will be processed just like any other visitor's; her data will be accessed and shared among various entities with or without her consent; and her Alexa will hear her argument with her significant other just like anyone's else; If you are lucky enough to have learnt some techniques and developed some way of securing your data and privacy then again you will not be substantially free. You will have only achieved a larger room to play in. In such a world, the responsibility lies in the hands of those who know the architecture of the place and the rules of the game. It's them who are commissioned to call for freedom, design for freedom and act for freedom. The responsibility for an open future lies primarily on their shoulders. By willing to transcend to their subjectivity and by renouncing everything 'given', these people are our only escape from being hardened and lost into the absurdity of our 'given world'. Simultaneously, we, the happily naïve children of this world, are the only possible way for their transcendence towards freedom. 'For a freedom wills itself genuinely only by willing itself as an indefinite movement through the freedom of others', Beauvoir keeps reminding the reader.

The moral compass that points at the ethical life is thus located in our thrust for our subjective and collective freedom. The designer, the CEO, the lobbyist or the activist have one thing through which to filtrate their available options when confronting ethical decisions: *which option will eventually magnify the freedom of others affected by my choice*? Inasmuch as these people who hold the keys for our open future and everyone else who accepts the possibility of his/her subjective freedom, are determinant in undertaking the anguish and risk of being free, they shall withstand dishonest attempts of bad faith that wane their will for freedom. Instead, they always need to bear in mind that: 'To want existence, to want to disclose the world, and to want [men] to be free are one and the same will'[2].

Hence, coupled with the *imperative* for the transformation of the business culture and environment, the suggested alternative framework for furnishing ethics within organizations requires that: 1) individuals will understand, embrace and strive for their universal responsibility; 2) they will accept their absolute freedom that expands well beyond the professional boundaries; and simultaneously -as well as necessarily- 3) they will crusade for the freedom of others. It is, primarily, the responsibility of those who know the rules of the game to act following the freedom of others and act to raise their consciousness towards the plethora of the ontologically possible choices at hand and their responsibilities towards an open future.

5 Conclusion

Beauvoir's and Sartre's philosophical contribution lies in the unconditional rejection of the deontological connotation of ethics. For them, ethics is not an end; there can be no end in ethics. No one will ever reach a point of being 'ethical' for there can be neither an absolute aim nor a perpetual subjectivity enclosed and certified in the sphere of the given world. Ethics is, and will always be, a means towards our transcendence to freedom summoned on our choices regarding the life we want to pursue and the world we want to live in.

In such a context, those who evangelize their ethical character and objectivity are nothing more than people confined in their 'given world'; lured by the comfort of inevitability, they deny, either unconsciously or purposefully, the very possibility of a choice for their freedom and the freedom of others. But there would be no meaning in life if history were to be perceived as an inexorable and everlasting 'mechanical unrolling', and our existence as an object fulfilling itself just by fueling the engines of this perennial Sisyphean task. The only possible way to discuss and/or implement ethics with people that reject their subjectivity and their role in the world is through abstraction, and there is no point in doing so. Conversely, it can potentially lead to opposite results by providing people with a natural yet bad faith escape button to mantle their choices and stifle their transcendence.

Regarded as a means, however, ethics becomes a way of pursuing your life. And like subjects, corporations can themselves decide to lead an ethical life. Beauvoir describes ethics as a 'game': 'the characteristic feature of all ethics is to consider human life as a game that can be won or lost and to teach man the means of winning'. Ethics is thus transformed in a way for the tech giants to present themselves to the world not as defenders of certain values but as subjects that strive to do the right thing in their daily life; by doing so they concurrently entreat everyone else to follow their lead. Ethics also lies in the way corporations regard the role of their employees and in the way they respect and raise the consciousness of their clients; But most of all, ethics is mirrored in the way certain entities respond to the challenge of achieving a better, sustainable and open to all, future. Corporations that want to walk through the strenuous pathway of ethics cannot but:

1. review the ethics of their profit formulas and avoid the procrastination that limits the discussion within the fictional AI boundaries;
2. be open, responsible and avoid the dishonest attempt to mask their true options either behind abstraction or under the mythical fallacy of inevitability;
3. nurture a corporate environment that will prepare and embrace their employee's autonomous transcendence and assumption of freedom and responsibility;

Today, we are at a crossroads, awkwardly looking at various directions. The younger generation has no idea what it means to be offline and at the same time, the older generation finds it difficult to adapt old patterns to new virtual spaces. The technological evolution surpassed our ability to conceptually cope with its speed; reforming or adapting our thesis in the world has been an existentially puzzling task. We have gradually, unwillingly and unconsciously become large anonymized computational assets within an uncontrollably expanding and unsustainable world whose ecosystem struggles to maintain even the minimum balance. People discuss and write best-sellers on how technology reinvents what it means to be human, disregarding the fact that accepting such a mechanistic attitude as inevitable will eventually engineer a world where only the fittest, the reinvented, will survive. But if there is something that defined our existence in this world, it is that we built societies where the least fit could not only survive, but thrive.

It is high time we realized our freedom. For this to happen, those who hold the keys to our future must take every possible step to re-establish themselves and the rest of us within our rights and help us restore our personal and collective sense of freedom. In a world where everyone is connected to others, my freedom springs from the freedom of my connections. There is no individual freedom here for there can be no individual digital sphere to connect to, no individual society to interact with and no individual Earth to live and prosper. It is time for the 'one-dimensional' data subject either to become human again or to stay captive of its 'given world' and annihilate its existence to the vanity of a fictional inevitability. As long as an individual consciously and freely chooses one or the other, freedom shall flourish. For the process of transcendence, anguish is inescapable. It always is when we decide to lift ourselves to the consciousness of our freedom. Beyond that, there is no easy way out. There never was.

ACKNOWLEDGMENTS

I sincerely thank all the anonymous reviewers whose comments helped me improve this paper. I also thank my PhD Supervisors, Dr Ben Sanders and Dr Emma Nottingham for their ongoing support as well as Dr Martina Hutton and the members of the Doctorate Roundtable 2019-2020 for their useful comments on an earlier version of the paper.

REFERENCES

[1] Philippe D'Anjou. 2011. An Ethics of Freedom for Architectural Design Practice. *Journal of Architectural Education 64, 2*, (March 2011), 141-147. DOI: https://doi.org/10.1111/j.1531-314X.2010.01137.x

[2] Simone de Beauvoir. 1947 (New ed. 2002). The Ethics of Ambiguity. Kensington Publishing.

[3] Simone de Beauvoir. 1965 (New ed. 1987). The Force of Circumstance. Penguin Classics.

[4] Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, Kate Crawford. AI Now Report. 2017. Retrieved December 4, 2018 from: https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html

[5] Thomas Flynn. 2006. Existentialism: A very short introduction. Oxford University Press.

[6] Antonio Gramsci. 1935 (New ed. 2005). Selections from the Prison Notebooks. Lawrence & Wishart Ltd

[7] Ben Green. 2018. Data Science as Political Action: Grounding Data Science in a Politics of Justice. arXiv: arXiv:1811.03435. Retrieved from https://arxiv.org/abs/1811.03435v1

[8] Martin Heidegger. 1927 (rev.ed. 2010). Being and Time: A Revised Edition of the Stambaugh Translation. State University of New York Press.

[9] Kevin T. Jackson. 2005. Towards Authenticity: A Sartrean Perspective on Business Ethics. *Journal of Business Ethics 58, 1* (June 2005). 307-325

[10] Herbert Marcuse. 1964. One-dimensional Man. Routledge & Kegan Paul Limited.

[11] Thomas Metzinger. 2019. Ethics washing made in Europe (April 2019). Retrieved December 4, 2018 from: https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html

[12] Samir Passi and Solon Barocas. 2019. Problem Formulation and Fairness. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19). ACM, New York, NY, USA, 39-48. DOI: https://doi.org/10.1145/3287560.3287567

[13] Microsoft Research Blog. Guidelines for Human-AI interaction design, (February 2019), Retrieved December 4, 2018 from: https://www.microsoft.com/en-us/research/blog/guidelines-for-human-ai-interaction-design/

[14] Jean-Paul Sartre. 1943 (2nd ed. 2003) Being and Nothingness: An Essay on Phenomenological Ontology. Routledge.

[15] Jean-Paul Sartre. 1946. Existentialism Is a Humanism. *Public lecture given in 1946.* Retrieved December 4, 2019 from: http://www.mrsmoser.com/uploads/8/5/0/1/8501319/english_11_ib_-_no_exit_-_existentialism_is_a_humanism_-_sartre.pdf

[16] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19). ACM, New York, NY, USA, 59-68. DOI: https://doi.org/10.1145/3287560.3287598

[17] Julia Powles and Helen Nissenbaum. 2018. The Seductive Diversion of 'Solving' Bias in Artificial Intelligence (December 2018). Retrieved December 4, 2018 from: https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53

[18] Edward Zalta (Ed.). 2017. The Stanford Encyclopedia of Philosophy (Winter 2017). Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/win2017/entries/existentialism/