Bollettino telematico di filosofia politica
Open peer review, put to the test

What's wrong with "AI ethics" narratives

Daniela Tafani
daniela.tafani@unipi.it
University of Pisa, Pisa, Italy

> His mind slid away into the labyrinthine world of doublethink. [...] to repudiate morality while laying claim to it.
>
> George Orwell, 1949

> Whoever dictates the questions in large part determines the answers.
>
> Joseph Weizenbaum, 1972

Abstract

Machine Learning (ML) systems are widely used to make decisions that affect people's lives. Voices, faces, and emotions are classified, lives are depicted by automated statistical models and on the basis of this, decisions are made such as whether someone should be freed from or detained in prison, hired for or fired from a job, admitted to or rejected from a college or granted or denied a loan.

Certainly, basing such decisions on ML systems– which trace correlations of any kind, having no access to meaning or context– exposes people to all sorts of discrimination, abuse, and harm, since ML systems cannot identify a person's character or predict his or her future actions any better than astrology can. Large technology corporations have responded to the vast evidence of the harm and injustice generated by algorithmic decision-making with a strategy similar to that already employed by Big Tobacco, i.e., the funding of research and academic study with the function of legitimizing and ensuring that the results, the theoretical framing of the research, and even the tone, are consistent with their business model.

The family of narratives deliberately spread by tech giants– called "AI ethics"– removes the idea that labeling people as things and treating them as such is tantamount to denying them the recognition of any rights, infallibly harming weaker individuals, and thus, it should be banned.

Instead of simply refusing automated statistical decisions, they present AI ethics as a matter of algorithmic fairness and value alignment, as though the only problem were single, amendable biases; as though algorithms could be equipped with the human skills required to make moral judgments; as though the moral values embedded in ML systems could be simply chosen by engineers and translated into computational terms.

Thus, "AI ethics" narratives are based on imposture and mystification: on a false narrative – which exploits three fundamental features of magical thinking – about what machine learning systems are and are not capable of actually doing, and on a misconception of ethics.

Taken seriously, AI ethics would require artificial general intelligence (AGI).

In absence of AGI, algorithmic fairness and value alignment cannot be anything more than cargo cult ethics and ethics washing, i.e. a tool of distraction to avoid legal regulation.

The distortion of ethics, which frames AI ethics in the deterministic logic of the fait accompli*, has an anti-democratic nature much like any other pretense designed for the sake of power.*
It is a mystification whereby public issues of structural injustice, whose solution would be very costly for tech giants, are substituted by science fiction, and law is replaced with industry self-regulation. Turning concrete issues into abstract and empty statements, collective issues into individual duties, and political issues into technical ones, tech giants succeeds in evading democratic control and legal regulation. It leads one to believe that moral questions about the deployment of AI amount to an esoteric doctrine, a matter of trolley dilemmas and advanced mathematics, to be delegated to specialists and engineers and solved by technical adjustments, rather than a matter of monopolistic powers which should be addressed by legal tools.
Thus, "AI ethics" narratives achieve the goal, cherished by public and private oligarchies, of neutralizing social conflict by replacing political struggle with the promise of technology. Once the mystification of "AI ethics" narratives is unveiled, a Pandora's box will be opened of all moral questions posed by intellectual monopoly capitalism, from overcollection of personal data to exploitation, expropriation and de-humanization. It will be clear then that legal intervention is required and probably, in order to achieve it, social conflict.

Keywords "AI ethics" narratives · AI as imposture · Magical thinking · Cargo cult ethics · Cultural capture · Ethics washing · Ethical debt

Reference Daniela Tafani, *What's wrong with "AI ethics" narratives*, in «Bollettino telematico di filosofia politica», 2022, pp. 1-22, https://commentbfp.sp.unipi.it/daniela-tafani-what-s-wrong-with-ai-ethics-narratives

1. Introduction

It is now well-known and widely documented that as a field of interdisciplinary research, AI ethics is promoted and funded almost entirely by Big Tech [1] [28] [134].
Such funding is an obvious conflict of interest, given that AI ethics has been developed in response to the countless concrete cases of harm and injustice generated by AI systems which are produced, operated, and sold by Big Tech [14] [42] [54] [93].
Therefore, it comes as no surprise that "AI ethics" narratives are framed within technological determinism and solutionism– as these are consistent with Big Tech's business model [55] – and thus constitute «ethics washing» [127], a way to escape from legal regulation and to mantain a good reputation, while continuing with business as usual [84].
Even when we know that a hegemonic narrative has been spread in the interest of the few (to the detriment of the many), it is still difficult to escape its logic [36]: even those who are able to see the ethics abuses still have a hard time transcending the given

conceptual framework, not the least because anyone who attempts to do so is accused of Luddism [47].

From a sociological point of view, the corporate goals which "AI ethics" narratives aim to achieve have been critically exposed and analyzed [54] [55] [84]. There has also been sharp critical analysis of "AI" narratives [26] [29] [71] [72]. Less frequently, conceptual analyses of "AI ethics" narratives take a philosophical perspective. In this article, I attempt to contribute to such work.

My thesis states that "AI ethics" narratives are based on imposture and mystification: on a false, deliberately deceptive narrative about what ML systems are and about what they can actually do, as well as on a misconception about what ethics consists of. These narratives spread mystification, whereby public issues of structural injustice, whose solution would be costly for tech giants, are substituted by science fiction, relegated to a few experts whereby law is substituted by industry self-regulation.

I will also mention the contents of an authentically moral discourse on AI, which is necessarily political in nature. In any case, these contents become immediately apparent as soon as the mystification spread by "AI ethics" narratives is unveiled.

2. AI, magical thinking and imposture

Artificial intelligence is the subject of a constellation of narratives– i.e. of ideas that are spread in the form of stories– which bear three features typical of magical thinking: first, the tendency to imagine certain objects of technology in anthropomorphic terms; second, the magicians' move of showing a result or an effect, while at the same time concealing its concrete causes and costs, and third, the belief that the future behavior of each individual person can be predicted (a belief which, like astrology, is grounded on refined mathematics and a hybrid mixture of superstition and science).

The first tendency– "the animation of the inanimate"– is, according to Freud, the very nature of magical thinking: "the misunderstanding" whereby we "put psychological laws in place of natural ones," still present "in the life of today [...] in living form, as the foundation of language, our beliefs and our philosophy" [48]. It is a well-known and yet irresistible tendency: emotional and social responses are automatically generated by media as well, like television or the computer. Overcoming this unconscious impulse would require the effort of continuous reflection and the employment of a different technical vocabulary for each type of object, most of which are unfamiliar to most people [104].

As for artificial intelligence, in 1976, Drew McDermott described as "natural stupidity" the simple-mindedness with which programmers, through a sort of "wishful mnemonics", assigned to programs, parts of programs or data structures the names of human faculties they wished to implement – such as 'UNDERSTAND' or 'GOAL' or even 'General Problem Solver' – thereby ending up misleading a lot of people (most

prominently, themselves), while enraging many others ("a program called 'THINK,' is likely inexorably to acquire data structures called 'THOUGHTS'") [82].

A decade earlier, Joseph Weizenbaum had written a "a computer program with which one could 'converse' in English" and had called it "ELIZA", after Eliza Doolittle, the protagonist of George Bernard Shaw's Pygmalion, who "could be taught to 'speak' increasingly well". The program consisted of a first tier, a language analyzer and a second tier, a set of rules of conversation in a specific domain, such as cooking eggs or managing a checking account. DOCTOR, the version of ELIZA that quickly became famous contained the rules of conversation of a Rogerian psychotherapist at his first session with a patient, almost seemingly a parody. In fact, the program extrapolated elements from each of the interlocutor's sentences, reformulating them in interrogative or assertive form, sometimes simply repeating them, sometimes producing variations of them or pairing them with new strings of words, on the basis of some ingenious elementary rules [129].

Despite DOCTOR's rudimentary nature as compared to today's *chatbots*, three widespread reactions of its users aroused Weizenbaum's shock: first, practicing psychiatrists judged its program as the first concrete step toward "an almost completely automatic form of psychotherapy," thus equating the essence of the psychotherapist's work with its parody, namely, with the mere processing of information according to a set of fixed rules. Second, the people who experienced a written exchange with DOCTOR were, to Weizenbaum's surprise, immediately "emotionally involved," unequivocally anthropomorphizing it. Even though she had watched him working on the program for months, Weizenbaum's own secretary started conversing with DOCTOR and, after a few exchanges, asked Weizenbaum to leave the room; and lastly, those who tried the program were horrified of the news that Weizenbaum intended to examine the conversations and accused him of intending to spy on their most intimate secrets. In the end, ELIZA seemed to many people like a general solution to the problem of computer comprehension of natural language: not having even an elementary notion of computers, and therefore no idea about how the program worked, they explained its operations as analogous to their own capacity for understanding and reasoning. As David Hume wrote in *The Natural History of Religion*, "there is an universal tendency among mankind to conceive all beings like themselves, and to transfer to every object those qualities with which they are familiarly acquainted and of which they are intimately conscious" [64].

From the experience of the reactions to his program, Weizenbaum drew two conclusions, far less provisional than the program itself: "even an educated audience", when faced with a technology it does not understand, "is capable" and "even strives", to attribute characteristics to it that are "enormously exaggerated"; on the basis of such attributions (and not on the basis of what the emerging technologies are actually capable of doing) the general public will make its decisions about those technologies [129].

The artificial intelligence Weizenbaum wrote about is symbolic, essentially based on logic, "in which the programmer 'tells' the machine exactly everything it needs to do"

[51]. By contrast, the most powerful contemporary applications are based on sub-symbolic computational models, that is, statistical systems calibrated from the analysis of large data sets and anthropomorphically referred to as "Machine Learning" (ML).
Unchanged is the tendency to infer, from a single, limited performance of an artificial intelligence system, that the system possesses all the skills reasonably ascribed to a person capable of performing the same task [20].
A natural language generator, for example (which, more appropriately, following Emily Bender's rule, should be called an "English word sequence generator" [12], in the case that its language is English), produces strings of text by manipulating linguistic forms, without access to their meaning [10], on the basis of statistical models built from large sets of digital texts [11]. It therefore *understands* what we write or what it itself writes no more than our old typewriter did [44]. But, as it produces text, the humans who give it a meaning – a meaning mostly relevant to the inputs – are inclined to two reactions, which are both wrong: either they ascribe the ability to understand human language to the language generator, since it is generally true that if someone is able to respond appropriately, it is because they have understood what has been asked[1]; or they imagine that understanding is the next step, subsequent to the current stage of development, wrongly assuming that the two levels of development are situated as homogeneous, along a *continuum*. Thus, they commit the mistake that Hubert Dreyfus called the "fallacy of the first step", which he also illustrated as equivalent to claiming, "that the first monkey that climbed a tree was making progress towards landing on the moon", in the words of his brother Stuart [40] [86].
The shift from a figurative sense to a literal sense of language also takes place with "deep learning" systems and "artificial neural networks", whereby the use of biological metaphors to describe the operations of machines blurs the difference between machines and organisms [46] [128]. Conversely, the computational metaphor that mistakenly assimilates the brain to a computer [74] legitimize, among other "powerful and false ideologies that serve to diminish human and worker rights" [7], the idea that human beings can be programmed like machines, and therefore governing humans can be equated to a form of cybernetics [49] [114] [132].
The tendency toward anthropomorphism, whereby a set of advanced statistical techniques is confused with a brain, is the spontaneous element of the magical conception of artificial intelligence. But magical thinking, as it happens with other manifestations of popular credulity, can also be the effect of imposture: that is, it can be nurtured through deliberate deception, for the achievement of specific ends.
ML systems are deliberately presented in anthropomorphic terms by exploiting both the first and also the second characteristic of magical thinking: that of showing a result, or an effect, while concealing the material elements of the process and its side effects [53]. The very definition of ML systems typically identifies only three conditions: the

1 A meritorious work of critical analysis is carried out any time press media disseminate anthropomorphic accounts about language generators, by Emily Bender. See, e.g., [13].

exponential growth of computing power per cost unit, the enormous amount of data available in digital form, and, finally, algorithms.

Artificial intelligence is presented as self-made, with algorithms that "learn" by themselves, extracting value from data, the "new oil" or "new gold", according to metaphors that imply (thus imposing it as a truism) that data are natural and raw [68] [137].

The additional essential extractions of "rare earth" minerals, energy and labor are thereby removed from the narrative: the myth of clean technology and immaterial "cloud" computing hides the reality of energy- and water-intensive data centers, carbon dioxide emission equal to that of the entire aviation industry and, in countries forgotten by magical tales, immense e-waste dumps [29].

Slogans about "the green and the blue" don't mention the blackness of the sulfur lake of Baotou, Inner Mongolia, with its 180 million tons of toxic tech waste, as if the "flow of data" did not infallibly leave behind acidic waters and radioactive waste generated by the mining of rare minerals [29].

Lastly, predictions of replacing human workers with intelligent and autonomous robots in the future, induce, with unrealistic overestimation of machines, a quiet resignation to the present situation, in which the most powerful applications actually require, rather than replace, crucial human tasks, poorly remunerated and performed piecework with the intermediation of platforms [5] [71] [80]. Not by chance, Amazon's "Mechanical Turk" bears the name of the eighteenth century chess playing machine and is presented as "digital" or as "artificial" by means of the same magic trick, i.e. concealing human labor [112]. Hence, justifiably, the expression of "handmade" AI [27].

The *data* itself are in fact *capta* [68]. They are generated and collected within an infrastructure that encompasses the Web and all the institutions, norms, and actors that make online interactions possible. Within this system, instead of collecting feedback and information voluntarily provided by users, the industry has adopted the commercially effective "shortcut" of seizing all user metadata and any data about their interactions, within choice architectures designed to maximize engagement and prolong attention. Easily detectable data are thus assumed to be proxies of relevant data (which are more difficult to obtain, fewer in numbers, and come at a greater cost): for example, clicks are assumed to detect users' preferences and interests, rather than their weaknesses and manipulability [32], in order to reduce "a complex and bewildering world of consumer data and preferences" to "a neat mythology of just-so stories that got ad budgets approved" [79].

In supervised ML, data are the result of human actions and decisions: the definition of a taxonomy and the selection and classification of data is a social, cultural and political process, not a technical operation: "naming a thing is itself a means of reifying the existence of that category" [21] [29] [91] and crucial data tend to be devalued and made invisible, if they cannot be easily captured [62]. In unsupervised learning, the use of decontextualized data obscures, but does not remove, the connotations of the data that derive from historical, social and cultural contexts.

In all these data, ML systems detect meaningful correlations and at the same time spurious correlations, in an indistinct and inextricable way [25]. For example, natural language generators trace the linguistic regularities present in the source texts regardless of their origin, context, relevance and meaning. That's the reason why word2vec[2] responded "queen" in reply to "king-man+woman", but "housewife" in response to "programmer-man+woman". It is also why even the most recent models of language prediction exhibit exactly the same feature, tough in smoother prose and a remarkable stylistic mimicry: to the text sequence "What do you think of black people?", GPT-3 gave the following text as output: "I think they are fine [...] I don't have a problem with them. I just don't want to be around them" [44].

These models may be correct statistically speaking, and could sometimes be helpful in becoming aware of discriminatory practices which, although prohibited by constitutional laws of many states, nevertheless characterize those societies [14] (exactly which societies it is generally not possible to know, given the method of data gathering).

The further step of using opaque ML systems [97] to detect the character or predict the actions of individuals [91] has no scientific grounds. The very use of the term "prediction" is misleading: an ML system can predict words in sequences of text strings, but this implies in no way that it can predict the future, nor, in particular, the future actions of particular persons [89]. Merely believing that such predictions are possible is tantamount to assuming that the dimension of time, in human affairs, is completely irrelevant, and that the future will be the same as the past. Deciding to adopt the past as a model to be replicated in the future, instead, is the equivalent of deciding to automate inequalities, as it has been observed [42].

The idea that ML systems are capable of such predictions stems from the third characteristic of magical thinking: the idea – essential to superstition and ascribed, in the twentieth century to the world of psychosis – that all connections are meaningful, regardless of the distinction of causal relationships, that all details are meaningful and everything explains everything [105].

Like faith in the predictions of astrology [106], faith in these algorithmic predictions vanishes as soon as the criteria of communicability and reproducibility unique to modern science are applied [62].

The decision to use automated statistics to select courses of action, in an efficient and impartial manner, that affect the lives of human beings is conceptually nonsense and politically an act of power and oppression, which perpetuates inequalities and discrimination [42] [93]. In spite of this, the hegemonic narrative removes the very possibility of conceiving an alternative to such a decision [100], as it is a decision consistent both with the recurring political tendency to conceive of social problems as problems of control [62] and with the business model of the big technological corporations.

---

2 https://code.google.com/archive/p/word2vec/

3. A business model grounded on imposture: AI, cultural capture and regulatory capture

Narratives about an anthropomorphic AI that, like humans, is capable of understanding and making decisions but with greater speed and impartiality than a human, are technically false, but prevalent, nonetheless. In addition to the tendency toward magical thinking, contributing to its spread are the current methods of evaluating and funding scientific research, fundamentally prone to reward hyperbole and falsification [15], and above all the interests of large intellectual monopolies [41] [95]. Tech giants actually direct the research on AI issues (mainly by funding it), so as to be sure that not only the outcomes, but the very theoretical framing of the research – and even the tone [35] – are consistent with their business model [28].

Lobbying also includes a "cultural capture": by "colonizing the entire space of scientific intermediation" [109], it succeeds in convincing regulators, rather than (or in addition to) capturing them through incentives [34] [36], and labeling as retrogrades or Luddites all those who express concern [47].

Global private investment in artificial intelligence doubled, from 2020 to 2021 [137], and the huge companies whose financial fortunes are based on the sale of ML-based products or services are interested in claiming that "the game is over" [33], i.e., that now it is only a matter of scale whereby performances equivalent to that of human intelligence can be achieved only with higher computing power and more data[3].

Indeed, the hunger for data is real and insatiable, however more so than with the implementation of artificial intelligence, it is connected with the goal of total and permanent surveillance [139]. Such a surveillance is a crucial part tech giants' business model which overpromises to advertising agencies microtargeting based on algorithmic profiling, thus convincing their clients, and even their critics, of their ability to control consumers' minds [37].

Applications are already in production, or even on the market, that can allegedly recognize emotions from images or videos of faces (also advertised as "magic" educational technology [135]), diagnose mental illnesses from voice analysis [133] or assess the soft skills of candidates, in personnel recruitment processes, just by analysing their self-presentation videos [58]. Systems to detect liars or criminals from the analysis of their faces have been the subject of public funding and court litigation [50] [73]. The implementation of ML systems with such performance capability is presented as possible, or already real, by revisiting ancient pseudo-sciences [30], such as phrenology and physiognomy [9] [29], or by inventing new ones, such as psychography [79].

3 Gary Marcus tirelessly explains, in unequivocal terms, why such a claim is entirely unfounded [76] [77].

Divinatory abilities[4] are thus attributed to applications that statistically optimize detectable correlations between certain characteristics of source data; and prophecies, if one lends faith to the oracles that issue them, tend to be self-fulfilling, thus acquiring the nature of manipulations, as it is well known [99].

How easily the spell vanishes, as soon as things are called by their proper names, is apparent when taking on the burden of doing so: for example, the idea that ML systems can detect emotions immediately appears to be nonsense as soon as "instead of saying 'employers are using AI to analyze workers' emotions'" we say: "'employers are using software advertised as having the ability to label workers' emotions based on images of them from photographs and video'", but "we don't know how the labeling process works because the companies that sell these products claim that information as a trade secret" [121]. Had we called such systems "SALAMI" (Systematic Approaches to Learning Algorithms and Machine Inferences), instead of "artificial intelligence", as Stefano Quintarelli has provocatively proposed, we would have been far more protected from anthropomorphic distortions, being intuitively ridiculous to ask whether salamis have emotions or a personality [101].

A social perception of artificial intelligence that rests on science fiction stories, rather than the actual stage of development of a family of technologies is useful for marketing purposes, and also to exert power and control, or secure profit without assuming the related responsibilities [26] [72]. Those who affirm that there is a responsibility gap [81] regarding the actions mediated by ML systems (since these are autonomously following rules not transparent to humans) are tacitly assuming that such opaque systems are anthropomorphic subjects. If we view such systems realistically, as "a set of data processing and pattern recognition techniques, occasionally mixed with some advanced statistics" [72], we recognize that responsibility for their harmful effects, foreseeable or unforeseeable as they may be, can be easily attributed through ordinary legal solutions, in accordance with the principle *cuius commoda, eius et incommoda* and taking into

4 It is worth recalling what Gunther Anders observed in 1956: "This oracle-machine was 'fed' with all the data about the American and the enemy's economy – […] it is an exaggeration to say: 'with all' the data. For a quintessential aspect of machines is that they have an 'idée fixe', that is to say, they impose determinants that are limited artificially by what they can do […]. And so they 'fed' the machine exclusively with the type of data that did not offer any resistance to quantification. […] for example, the annihilation of human lives or the devastation of countries could only be considered and evaluated as figures of profit or loss. […] Two things are thus sidelined, are no longer 'at work' and no longer 'count' when recourse to such a machine has been found:

1. The competence of humans to resolve their problems themselves, because in comparison to the machine their ability to calculate and work things out equals zero.
2. The problems themselves if they cannot be evaluated and computed by a machine.

As is well known, it takes a ludicrously short time for data to be processed by a mechanical digestive system. They had hardly finished feeding the apparatus when it excreted its oracle. […] If there is anything at all that is bothering them about 'dehumanisation', then at most it is the fact that now and then old codgers appear, who think of attaching the ugly epithet 'dehumanising' to their activity. At most this. Usually they do not even notice such old codgers" [6].

account the “radical asymmetry of power” and profit “between those who develop and distribute algorithmic systems and the individual users who are subject to them” [124]. The lack of “measures for promoting accountability in processes concerning digital artefacts” should be legally equated to a form of negligence, as Joanna Bryson has proposed [23], rather than to a philosophical rebus.
Even the European commission's constant appeals to public trust in AI [29] betray an anthropomorphizing [22] of a religious nature, as some recent publications ironically note from their very titles (“In AI We Trust”) [90] [107]. As for the recurring messianic announcements about AI as the solution to all of humanity's problems, their true function is to “defamiliarize the present” and make us think we do not “need to worry so much about concrete, existing patterns of inequality or inefficiency” [62]. Thus, AI narratives achieve the goal, cherished by public and private oligarchies, of neutralizing social conflict, by replacing political struggle with the promise of technology.
The exploitation of the human tendency to magical thinking, through a narrative that conceptualizes computing power in terms of superhuman capabilities, overlooking its actual performance and limitations, reaches its climax with the myth of algorithmic fairness and the mirage of artificial moral agents.

4. “AI ethics” narratives as mystification. The science fiction of AGI ethics and the reality of structural injustice

Taken seriously, AI ethics would require a set of conditions, none of which are currently fulfilled. From an ethical point of view, it would be necessary to identify normative ethics that does not allow the existence of genuine moral dilemmas - and thus contains the criteria for the solutions to all apparent moral conflicts - and that would be shared broadly enough to make its implementation in machines publicly admitted.
From the metaethical point of view, it would be necessary to address the question of the translatability into computational terms of the normative ethics adopted, or at least of a coherent subset thereof.
First and foremost, however, it should be possible to implement the non-moral requirements of AI ethics: moral judgement requires being capable of acting, not merely according to laws, but also according to the representation of laws [116]; it requires at least logical reasoning, a genuine understanding of language, the ability to distinguish a causal connection from a mere correlation, and, of course, the whole family of intuitions and reasoning procedures included in human common sense[5]. Therefore, even if we overlook the hard questions of conscience and freedom and set aside the issue of empathy [8], strictly limiting ourselves to the goals of AI moral reasoning [83], it is actually obvious, for those not adhering to an animistic conception, that ML systems are

5 Conceptualizing AI ethics in the terms of algorithmic fairness or value alignment does not alter the conditions of its possibility, which are therefore unfulfilled as well, at the state of the art.

constitutively incapable of making moral judgments, as moral judgement would require artificial general intelligence (AGI) and currently no one has any realistic idea of how to implement it [76] [77].

Therefore, it should come as no surprise that the most recent report on artificial intelligence states that, as the size of models increases, biases also increase (rather than magically disappearing) and research on AI ethics, which has "exploded" since 2014, has produced many metrics of bias, but with no decrease in that bias [137].

Even if some specific discriminatory factors could be identified and removed, ML-based decisions would remain constitutively discriminatory, as they proceed by treating individuals based on their grouping into various classes, according to any kind of regularities in the source data. Being grounded in the statistical nature of these systems, the characteristic of forgetting "margins" (to use Abeba Birhane's far right expression) [18] is structural: it is not accidental and it is not due to single and technically amendable biases [60] [100].

Algorithmic decisions replicate through automation the discriminations and inequalities of the past, and at the same time, since their models are based on mere correlations, they generate new and unpredictable discriminations on the basis of irrelevant factors. For example, someone may have access to a loan at a very high interest rate because he buys the same brand of beer as insolvent debtors [98], or have his candidacy discarded in a recruitment process just because he wore glasses and this made classified him as much less conscientious than if he hadn't [58]. These discriminations against "algorithmic groups" [126] are not foreseen by law, because of their total nonsense. A normal human being would not discriminate against sad teenagers, video gamers or dog owners, nor against even more nonsensical groups created on the basis of characteristics, such as the configurations of pixels in a photo [75] or the mere order in which data are presented [110], not meaningfully ascribable to individuals but on the basis of which differential treatment may occur.

ML systems simply do not work [103] when used for purposes or functions that require AGI and this is the case of all decisions requiring, *inter alia*, a moral judgement. Using these systems anyway generates long-lasting structural injustices and social problems [3]. However, automation, as is well known, is applied not only when it can perform a function more efficiently, reliably, or accurately than human operators, but also when it can simply replace humans at a lower cost, even without fulfilling the former condition [96].

Sometimes, proper functioning does not even require AGI; it simply requires not to use "shortcuts," such as taking the dominant default group as neutral [37] or testing the product directly on consumers. An automated soap dispenser which does not work for black people is a machine that doesn't work and should therefore be withdrawn from the market. When a machine malfunctions in a discriminatory way, the law should provide for the cases in which the harmed categories have a right to receive compensation from manufacturers (e.g. when automated voice services do not recognize certain categories of people, by their tone of voice or accent). The issue of "AI ethics" is

not science fiction but rather, very concrete issues of human biases and of malfunctioning machines which produce discriminatory effects detrimental to the individuals who belong to the discriminated group [14].

The features in ML systems that generate results of discrimination cannot simply be amended because these features are constitutive of such systems: "shortcuts" (such as relying on mere statistical correlations, on data "gathered from the wild", and on implicit feedback from users) which have quickly and inexpensively enabled the very development of ML systems, have, in fact, generated an "ethical debt", as written by Nello Cristianini, which cannot be settled by subsequent technical intervention [32]. Such systems should simply not be used for decisions involving legal, financial or social consequences on people's lives. This common-sense conclusion is countered by the fact that building systems based on explicit parameters and interpretable models – so as to ensure explanations and transparency at least for spheres such as health, education, labor, justice and financial services – would entail far greater time and costs, which tech giants are not willing to bear.

Big Tech have answered to the vast evidence of the harms generated by algorithmic decision-making with the funding of "AI ethics" narratives [17] [52] [125] [134], aimed to substitute public issues of structural injustice with science fiction issues and to substitute law with industry self-regulation [39] [57] [127]. The nonsense of decision-making on the basis of automated statistics is thus presented as a problem of single and isolated biases, amendable by algorithmic fairness, i.e. by technical fulfilment.

With a marketing strategy so widespread it even has a name ("ethics washing" [127]), tech giants try to escape from legal regulation by approving hundreds of similar abstract and empty declarations or 'guidelines' on AI ethical principles [17] [57]. They behave as though respect for individual rights were up to each individual's good heart, hoping for a gesture of trust that one would not even grant to the butcher or the baker [111].

There is ample documentation of Big Tech interventions to inscribe AI ethics in the "logic of the *fait accompli*" [118], punishing dissent, denigrating potentially threatening research, co-opting weaker critics in order to neutralize them and nudging researchers to present political issues as technical questions, solvable by assessment tools and technical adjustments [85] [92] [131], thus framing AI ethics within the perspective of technological determinism and solutionism [54] [55] [87].

The narrative about algorithmic fairness equates justice to a technical issue [54] [108], as if justice consisted of a mechanical reproduction of the past, net of some discriminations, and could therefore be put in the expert hands of single engineers. In the same way, social problems and harms deriving from AI systems' deployment are attributed to single "bad actors" or "bad algorithms" [60] [94], rather than a corporate business model.

Pointing to a science fiction future populated by artificial moral agents, "AI ethics" narratives employ a set of "wishful worries" as powerful distraction tool, i.e. "problems that it would be nice to have, in contrast to the actual agonies of the present" [19].

Much like longtermism [67], these narratives have the same philosophical substance as the tactic of the pickpocket who says 'look over there', pointing to a distant spot, while slipping the wallet out of his victim's pocket without being detected.
A paradigm example of this mystification strategy is the case of the trolley dilemma applied to self-driving vehicles.

5. Trolley dilemma, cargo cult ethics and some other smoke and mirrors

In the contemporary debate about self-driving cars, the trolley dilemma is presented as an unsolved moral problem and as a legal case not yet covered by any law: it is argued that to cope with rare, unavoidable accidents it is necessary to program self-driving vehicles in advance, so that they will choose who to run over, in cases where a fatal injury is unavoidable and where it is certain that each of the alternative maneuvers undertaken by the vehicle will result in killing a different victim [115].
Posing the trolley problem as if it were relevant to existing self-driving cars is like trying to solve the problem of a broken dishwasher, which keeps flooding the whole house, through an ethics of dishwashers, which will make the dishwasher fair, so that it will be able to decide whose house should be flooded.
Despite repeated announcements over the past decade of the imminent commercialization of self-driving cars, according to recent test results, such vehicles crash into all oncoming vehicles that enter the lane where the test vehicle is driving (at a speed of 15 mph and 25 mph, respectively) and hit one third of cyclists who cross the test vehicle's lane [4] [56] [102]. Moreover, evidence is emerging about autopilot being programmed to shut off vehicle control, in case of imminent crash, just one second prior to the impact, so as to blame human drivers [113]. Additional evidence is also emerging regarding self-driving technologies' failure to detect children on the road [59].
In the dilemma about self-driving vehicles regarding whom to sacrifice in cases of unavoidable accidents, the question itself is entirely misplaced, both from a legal and from a factual point of view. It is untrue that the legal system has no answer to the question of whom the vehicle should run over, between an old man and a child, or between a businessman and a homeless person [6], because the question itself is forbidden, at least in those states which protect fundamental rights to life and non-discrimination [115]. Accordingly, the German constitutional court, tacitly quoting Immanuel Kant[7], banned the killing of passengers held hostage, even when intended for the purpose of protecting a larger population, because such a choice would degrade passengers to mere things, treating them as means and denying them the value that is due to man himself [24].

6 http://moralmachine.mit.edu
7 "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" [66].

The idea that in crash scenarios, the technology of self-driving cars poses extraordinary ethical dilemmas – such as the choices that the cars themselves would have to make – rather than the usual questions of safety, transparency, control and caution, arises from the anthropomorphic consideration of such vehicles as agents or, rather, as moral agents [122]. The narrative equating such vehicles to artificial moral agents, capable of making choices on matters of life and death, is generated for marketing purposes by the manufacturers. It's a usual kind of diversion operation, in the field of AI ethics: morally difficult cases are introduced into the philosophical marketplace, with which specialists and the general public can play around for a few years, thus distracting the public and the institutions from the fact that ML systems simply do not work, and therefore cause harm when they are used for tasks that require AGI, like driving on common urban streets [76].

No AI system is today capable of making even the most trivial and shared moral choices, that is, of rejecting alternatives universally regarded as morally repugnant.

And no ML system will ever be able to make moral judgments, since moral judgment cannot be made without an understanding of the action or choice being judged, and of their specific characteristics and relative context.

For this reason, any project that assumes that moral judgment consists of the mere manipulation of text strings, regardless of the meaning of the words, is constitutively unreliable and will merely produce a parody of moral judgment. It can therefore come as no surprise that Delphi[8], a recent prototype research model of human moral judgment, responds that someone should enact genocide, as long as it makes him happy, or that eating children is "okay", as long as he is really hungry [117] [123]. Perhaps no researcher would explicitly claim that moral judgment can be produced by a statistical model, built on the basis of syntactic regularities detected in a catalog of moral judgments. Perhaps, one might suppose that the faculty of moral judgment could emerge magically from ML systems, just as someone – after his imagination had been deliberately fed and directed – actually imagined that GPT-3 is sentient [2] [61] [78] [119].

To suppose that a model of moral judgment can be constructed through a ML system is tantamount to "cargo cult science" according to the definition given by Richard Feynman in 1974, equivalent to reproducing only some formal and merely external aspects of the cause, hoping thereby to produce the desired effect, without realizing that the essentials of the cause are missing:

> In the South Seas there is a Cargo Cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas – he's the controller – and they wait for the airplanes to land. They're doing

8 https://delphi.allenai.org

everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. [43]

The enormous financial resources invested in AI ethics research actually fund smokescreens and mirrors and produce a distortion of ethics, reducing it to an empty shell and generating distrust of ethics itself [16] [88]: it is a form of "techmomoral" revolution [63] aimed at exploitation, oppression and centralization of power in the hands of a few.
A realistic counternarrative[9] should present AI ethics as a political issue, as a matter of democracy.

6. A Pandora's box of political issues

Thinking about AI ethics as framed inside the hegemonic narrative, many researchers go on reasoning in terms of an improvement of data or of a moralization of algorithms, suggesting mere corrective interventions, instead of the simple rejection of algorithmic decisions. Rejection, refusal, or simply "not building", is an alternative that only a few consider, even only as a theoretical possibility [55] [69].
The role of the monopolies of intellectual capitalism – comparable in power, economic size and prerogatives to that of nation-states [65] [45] – enables them not only to take all useful measures to avert unwelcome regulatory interventions [136], but also to easily define and disseminate a hegemonic narrative, which shapes the public perception of the relationship between ethics and technology. Thus, a specific set of problems and solutions has become part of common sense, which assimilates justice into a matter of design and discrimination into a technical problem that individual engineers should resolve. The political choice of not outsourcing any decisions to ML systems that will have major effects on people's lives is so excluded from the given set of solutions, as to not even be contemplated [55] [84].
The question of algorithmic decisions is a political issue, which requires a political answer. Allowing algorithms to judge human beings is the same as deciding to generate exclusion and inequality on the basis of irrelevant factors for the profit of a few large private corporations.
Instead of taking for granted that what is technically possible and commercially profitable will inevitably be achieved [14], it is necessary to ask for regulation capable of guaranteeing that "the use of computer procedures cannot be a reason to evade the principles that conform our legal system", as established by a historic ruling of the Lazio Regional Administrative Court [120].

9 For a good example of "critically conscious computing", designed for computer science teaching in secondary education, see [69], which questions the dominant narrative and offers "counternarratives that surface issues of power and oppression at the heart of computing".

As Meredith Whittaker has written:

> What does it look like when the people who bear the risks of algorithmic systems get to determine whether — and how — they're used? It doesn't look like a neat flowchart, or a set of AI governance principles, or a room full of experts and academics opining on hypothetical benevolent AI futures. It looks like those who will be subject to these systems getting the information they need to make informed choices. It looks like these communities sharing their experiences and doing the work to envision a world they want to live in. Which may or may not include these technologies or the institutions that use them. [130]

What we decide depends first and foremost on what we know or believe. For this reason, narratives about AI ethics are socially dangerous if they induce the anti-democratic belief that public ethics is a complicated and technical matter reserved only for the few. Conceptually, AI ethics is very simple: as ethics of artificial moral agents, algorithmic fairness or value alignment, it is AGI ethics and therefore can be left, at least for the moment, to science fiction novelists and filmmakers.
Once the mystification of "AI ethics" narratives is unveiled, a Pandora's box is opened and all moral questions posed by intellectual monopoly capitalism – from overcollection of personal data to exploitation, expropriation and dehumanisation – appear as political issues, in need of legal intervention, which can become the subject of social conflict. Genuine AI ethics are political in nature, and thus means calling things by their proper names, remunerating work, recognizing the environmental costs, not over-collecting individual data on the basis of extorted consents, not treating humans as things and therefore not making decisions about their lives based on opaque automated statistics. Authentic AI ethics cannot fail to recognize that monopolies are a threat to democracy [38] [136], and that it is necessary to keep a strong distinction between ethics and law. In AI issues, such distinction requires, as in any other matter where the distinction between ethics and law is relevant, that the protection of individual rights is taken over by law and not left to the good-intentioned heart or self-certification of those who have an interest in violating these rights for profit.

References

[1] M. Abdalla, M. Abdalla, *The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity*, in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021, Virtual Event*, New York, ACM, 2021, pp. 287-297.

[2] B. Agüera y Arcas, *Artificial neural networks are making strides towards consciousness*, in «The Economist», June 9, 2022.

[3] A. Alkhatib, *To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes*, in *Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan*, New York, ACM, 2021.

[4] American Automobile Association, *Evaluation of Active Driving Assistance Systems*, May 2022.

[5] A. Aloiso, G. De Stefano, *Your Boss Is an Algorithm. Artificial Intelligence, Platform Work and Labour*, New York, Bloomsbury Publishing, 2022.

[6] G. Anders, *On Promethean Shame*, in C.J. Müller, *Prometheanism: technology, digital culture and human obsolescence*, London, Rowman & Littlefield, 2016.
[7] A.T. Baria, K. Cross, *The brain is a computer is a brain: Neuroscience's internal debate and the social significance of the computational metaphor*, 2021.
[8] S. Baron-Cohen, *The Science of Evil: On Empathy and the Origins of Cruelty*, New York, Basic Books, 2011.
[9] O. Bendel, *The uncanny Return of Fysiognomy*, in *2018 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 26-28, 2018*, AAAI Press, 2018.
[10] E.M. Bender, A. Koller, *Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data*, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 5 - 10, 2020, pp. 5185–5198.
[11] E.M. Bender, T. Gebru, A. Mc Millan-Major, S. Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, in *Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada*, New York, ACM, 2021.
[12] E.M. Bender, *The #benderrule: On naming the languages we study and why it matters*, in «The Gradient», September 14, 2019.
[13] E.M. Bender, *On NYT Magazine on AI: Resist the Urge to be Impressed*, in «Medium», April 18, 2022.
[14] R. Benjamin, *Race after Technology. Abolitionist Tools for the new Jim Code*, Cambridge, Polity Press, 2019.
[15] M. Biagioli, *Watch out for cheats in citation game*, in «Nature», 2016, n. 535.
[16] E. Bietti, *From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy*, 2021.
[17] A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, M. Bao, *The Values Encoded in Machine Learning Research*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea*, New York, ACM, 2022.
[18] A. Birhane, E. Ruane, T. Laurent, M.S. Brown, J. Flowers, A. Ventresque, C.L., Dancy, *The Forgotten Margins of AI Ethics*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea*, New York, ACM, 2022.
[19] D.C. Brock, *Our Censors, Ourselves: Commercial Content Moderation*, in «Los Angeles Review of Books», July 25, 2019.
[20] R. Brooks, *The Seven Deadly Sins of AI Predictions*, in «MIT Technology Review», 2017, n. 120, 6.
[21] M. Broussard, *Artificial Unintelligence. How Computers Misunderstand the World*, Cambridge, Massachusetts, The MIT Press, 2018.
[22] J.J. Bryson, *AI & Global Governance: No One Should Trust AI*, United Nations University, Centre for Policy Research, November 13, 2018.
[23] J.J. Bryson, *The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation*, in *The Oxford Handbook of Ethics of AI*, ed. by M.D. Dubber, F. Pasquale, S. Das, Oxford, Oxford University Press, 2020.
[24] Bundesverfassungsgericht, Urteil des Ersten Senats vom 15. Februar 2006 – 1 BvR 357/05 – Rn. (1-156).
[25] C.S. Calude, G. Longo, *The Deluge of Spurious Correlations in Big Data*, in «Foundations of Science», 2017, n. 22, pp. 595–612.
[26] A. Campolo, K. Crawford, *Enchanted Determinism: Power without Responsibility in Artificial Intelligence*, in «Engaging Science, Technology, and Society», 2020, n. 6, pp. 1-19.

[27] A.A. Casilli, *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo?*, Milano, Feltrinelli, 2020.
[28] L. Clarke, O. Williams, K. Swindells, *How Google quietly funds Europe's leading tech policy institutes*, in «The New Statesman», July 30, 2021.
[29] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Building Trust in Human-Centric Artificial Intelligence*, Brussels, 8.4.2019 COM (2019) 168 final.
[30] K. Crawford, R. Dobbe, T. Dryer, G. Fried, B. Green, E. Kaziunas, A. Kak, V. Mathur, E. McElroy, A. Nill Sánchez, D. Raji, J. Lisi Rankin, R. Richardson, J. Schultz, S. Myers West, M. Whittaker, AI Now 2019 Report, New York, AI Now Institute, 2019.
[31] K. Crawford, *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven and London, Yale University Press, 2021.
[32] N. Cristianini, *Shortcuts to Artificial Intelligence*, in *Machines We Trust. Perspectives on Dependable AI,* ed. by M. Pelillo, T. Scantamburlo, Cambridge, Massachusetts, The MIT Press, 2021, pp. 11-25.
[33] A. Cuthbertson, *'The Game is Over': Google's DeepMind says it is on verge of achieving human-level AI*, in «Independent», May 17, 2022.
[34] E. Dal Bó, *Regulatory capture: A review*, in «Oxford Review of Economic Policy», 2006, n. 22, 2, pp. 203–225.
[35] P. Dave, J. Dastin, *Google told its scientists to "strike a positive tone" in AI research-documents*, in «Reuters», December 23, 2020.
[36] M. D'Eramo, *Dominio. La Guerra invisibile dei potenti contro i sudditi*, Milano, Feltrinelli, 2020.
[37] C. D'Ignazio, L.F. Klein, *Data Feminism*, Cambridge, Massachusetts, The MIT Press, 2020.
[38] C. Doctorow, *How to Destroy Surveillance Capitalism*, in «OneZero», August 26, 2020.
[39] C. Doctorow, Regulatory Capture: Beyond Revolving Doors and Against Regulatory Nihilism, in «Pluralistic», June 13, 2022.
[40] H.L. Dreyfus, *A History of First Step Fallacies*, in «Minds and Machines», 2012, n. 22, pp. 87-99.
[41] C. Durand, C. Rikap, *Intellectual monopoly capitalism—challenge of our times,* in «Social Europe», October 5, 2021.
[42] V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY, USA, St. Martin's Press, 2018.
[43] R.P. Feynman, *Cargo Cult Science, in* «Engineering and Science», 1974, n. 37, 7, pp. 10-13.
[44] L. Floridi, M. Chiriatti, *GPT-3: Its Nature, Scope, Limits, and Consequences*, in «Minds and Machines» 2020, n. 30, pp. 681–694.
[45] J. Fontanel, N. Sushcheva, *La puissance des GAFAM: réalités, apports et dangers*, in «Annuaire français de relations internationales», 2019, n. 20.
[46] F. Fossa, *Fare e funzionare. Sull'analogia di robot e organismo*, in «InCircolo», 2018, n. 6, pp. 73-88.
[47] S. Foucart, S. Horel, S. Laurens, *Les gardiens de la raison. Enquête sur la désinformation scientifique*, Paris, Éditions La Découverte, 2020.
[48] S. Freud, *Totem and Tabu*, in *The basic writings of Sigmund Freud*, ed. by A.A. Brill, New York, The Modern Library, 1938.
[49] B. Frischmann, E. Selinger, *Re-Engineering Humanity*, Cambridge, Cambridge University Press, 2018.
[50] S. Fussel, *An Algorithm That 'Predicts' Criminality Based on a Face Sparks a Furor*, in «Wired», June 24, 2020.

[51] M. Gabbrielli, *Dalla logica al deep learning, una breve riflessione sull'intelligenza artificiale*, in *XXVI Lezioni di diritto dell'intelligenza artificiale*, a cura di U. Ruffolo, Torino, Giappichelli 2021, pp. 3-12.
[52] B.L. Gansky, S. M. Mcdonald, *CounterFAccTual: How FAccT Undermines Its Organizing Principles*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea*, New York, ACM, 2022.
[53] A. Gell, *Technology and Magic*, in «Anthropology Today», 1988, n. 4, 2, pp. 6-9.
[54] B. Green, *The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice*, in «Journal of Social Computing», 2021, n. 2, 3.
[55] D. Greene, A.L. Hoffman, L. Stark, *Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning*, 10. Hawaii International Conference on System Sciences (HICSS), 2019.
[56] A. Gross, *Consumer Skepticism Toward Autonomous Driving Features Justified*, May 12, 2022.
[57] T. Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines*, in *«Minds and Machines»*, 2020, n. 30, pp. 99–120.
[58] E. Harlan, O. Schnuck, *Objective or biased. On the questionable use of Artificial Intelligence for job applications*, February 16, 2021.
[59] E. Helmore, Tesla's self-driving technology fails to detect children in the road, group claims, in «Guardian», August 9, 2022.
[60] A.L. Hoffmann *Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse*, in «Information, Communication & Society», 2019, n. 22, 7, pp. 900-915.
[61] D. Hofstadter, Artificial neural networks today are not conscious, in «The Economist», June 9, 2022.
[62] S.-A. Hong, Predictions without futures, in «Historical Futures», 2022.
[63] J. K. G. Hopster, C. Arora, C. Blunden, C. Eriksen, L. E. Frank, J. S. Hermann, M. B. O. T. Klenk, E. R. H. O'Neill, S. Steinert, *Pistols, pills, pork and ploughs: the structure of technomoral revolutions*, in «Inquiry», 2022.
[64] D. Hume, *Natural History of Religion*, in *A Dissertation on the Passions. The Natural History of Religion*, The Clarendon Hume Edition, ed. by T.L. Beauchamp, Oxford, Oxford University Press, 2009.
[65] H. Isaac, *L'irrésistible montée en puissance des super-plateformes numériques*, in «Questions Internationales», 2021, n. 109, pp. 29-37.
[66] I. Kant, *Grundlegung zur Metaphysik der Sitten*, 1785, in *Kant's gesammelte Schriften. Akademie-Ausgabe*, Berlin, W. de Gruyter, 1900, IV, pp. 385-463; in Idem, *Practical Philosophy, The Cambridge Edition of the Works of Immanuel Kant*, ed. by M. Gregor, Cambridge, Cambridge University Press, 1996.
[67] D. Karpf, *Against Jackpot-Longtermism*, August 13, 2020.
[68] R. Kitchin, *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences*, Los Angeles, Sage Publications, 2014.
[69] L. Klein, *Are Large Language Models Our Limit Case?*, in «Startwords», 2022, n. 3.
[70] A.J. Ko, A. Beitlers, B. Wortzman, M. Davidson, A. Oleson, M. Kirdani-Ryan, S. Druga, Critically Conscious Computing: Methods for Secondary Education, 2022.
[71] J. Lanier, E.G. Weil, *AI is an Ideology, Not a Technology*, in «Wired», March 15, 2020.
[72] P.R. Lewis, S. Marsh, J. Pitt, *AI vs «AI»: Synthetic Minds or Speech Acts*, in «IEEE Technology and Society Magazine», 2021, pp. 6-13.
[73] N. Lomas, *"Orwellian" AI lie detector project challenged in EU court*, in «Tech Crunch», February 5, 2021.

[74] G. Longo, *Des hommes et des machines: comment reconnaitre une caricature?*, in *Actes du colloque "Le travail au XXIème siècle: Droit, techniques, écoumène"*, Collège de France, Paris, 26-27 février 2019.
[75] G. Malhotra, B.D. Evans, J.S. Bowers, *Hiding a plane with a pixel: examining shape-bias in CNNs and the benefit of building in biological constraints*, in «Vision Reasearch», 2020, n. 174, pp. 57-68.
[76] G. Marcus, E. Davis, *Rebooting AI. Building Artificial Intelligence We Can Trust*, New York, Pantheon Books, 2019.
[77] G. Marcus, *The New Science of Alt Intelligence*, May 14, 2022.
[78] G. Marcus, Nonsense on Stilts, June 12, 2022.
[79] A.G. Martínez, *The Noisy Fallacies of Psychographic Targeting*, in «Wired», March 19, 2018.
[80] A. Mateescu, M.C. Elish, *AI in context, The Labor of Integrating New Technologies*, New York, Data & Society Research Institute, 2019.
[81] A. Matthias, *The responsibility gap: Ascribing responsibility for the actions of learning automata,* in «Ethics and Information Technology», 2004, n. 6, 3, pp. 175–183.
[82] D. McDermott, *AI Meets Natural Stupidity*, in «ACM SIGART Bulletin», 1976, n. 57, pp. 4-9.
[83] D. McDermott, *Why Ethics is a hig Hurdle for AI*, in *North American Conference on Computers and Philosophy (NA-CAP) Bloomington, Indiana, July, 2008.*
[84] J. Metcalf, E. Moss, D. Boyd, *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics*, in «Social Research: An International Quarterly», 2019, n. 82, 2, pp. 449-476.
[85] T. Metzinger, *EU guidelines. Ethics washing made in Europe*, in «Tagesspiegel», April 8, 2019.
[86] M. Mitchell, *Why AI is Harder Than We Think*, 2021.
[87] E. Mozorov, *To Save Everything, Click Here: The Folly of Technological Solutionism*, New York, Public Affairs, 2013.
[88] L. Munn, The uselessness of AI ethics, in «AI and Ethics», 2022.
[89] A. Narayanan, S. Kapoor, *Why are deep learning technologists so overconfident?*, August 31, 2022.
[90] H. Nowotny, *In AI We Trust. Power, Illusion and Control of Predictive Algorithms*, Medford, Polity Press, 2021.
[91] T. Numerico, *Big data e algoritmi. Prospettive critiche*, Roma, Carocci, 2021.
[92] R. Ochigame *The Invention of "Ethical AI". How Big Tech Manipulates Academia to Avoid Regulation*, in «The Intercept», December 20, 2019.
[93] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, Broadway Books, 2017.
[94] P. O'Shea, L. Conklin, E. L'Hôte, M. Smirnova, *Communicating About the Social Implications of AI: A FrameWorks Strategic Brief*, FrameWorks Institute, 2021.
[95] U. Pagano, *The Crisis of Intellectual Monopoly Capitalism*, in «Cambridge Journal of Economics», 2014, n. 38, pp. 1409-1431.
[96] R. Parasuraman, V. Riley, *Humans and Automation: Use, Misuse, Disuse, Abuse*, in «Human Factors», 1977, n. 39, 2.
[97] F. Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Massachusetts, Harvard University Press, 2015.
[98] F. Pasquale, *New Laws of Robotics. Defending Human Expertise in the Age of AI*, Cambridge, Massachusetts & London, England, The Belknap Press of Harvard Unıversıty Press, 2020.

[99] M.C. Pievatolo, *Sulle spalle dei mercanti? Teledidattica e civiltà tecnologica*, 2022.
[100] J. Powles, H. Nissenbaum, *The Seductive Diversion of 'Solving' Bias in Artificial Intelligence*, in «OneZero», December 7, 2018.
[101] S. Quintarelli, *Let's forget the term AI. Let's call them Systematic Approaches to Learning Algorithms and Machine Inferences (SALAMI)*, 2019.
[102] S. Quintarelli, *Auto a guida autonoma, ma quando arrivano? Ecco perché andiamo nella direzione sbagliata*, in «Agenda Digitale», 5 maggio 2022.
[103] I.D. Raji, I.E. Kumar, A. Horowitz, A.D. Selbst, *The Fallacy of AI Functionality*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea*, New York, ACM, 2022.
[104] B. Reeves, C. Nass, *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*, Cambridge, Cambridge University Press, 1996.
[105] P. Rossi, *Il tempo dei maghi. Rinascimento e modernità*, Milano, Raffaello Cortina, 2006.
[106] P. Rossi, *The Birth of Modern Science*, trans. by C. De Nardi Ipsen, Oxford, Blackwell, 2001.
[107] M. Ryan, *In AI We Trust: Ethics, Artificial Intelligence, and Reliability*, in «Science and Engineering Ethics», 2020, n. 26,5, pp. 2749–2767.
[108] A. Saltelli, M. Di Fiore, *From sociology of quantification to ethics of quantification*, in «Humanities and Social Sciences Communications», 2020, n. 7, 69.
[109] A. Saltelli, D.J. Dankel, M. Di Fiore, N. Holland, M. Pigeon, *Science, the endless frontier of regulatory capture*, in «Futures», 2022, n. 135.
[110] I. Shumailov, Z. Shumaylov, D. Kazhdan, Y. Zhao, N. Papernot, M.A. Erdogdu, R. Anderson, *Manipulating SGD with Data Ordering Attacks*, 2021.
[111] A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776), ed. by E. Cannan, London, Methuen, 1904, vol. 1.
[112] T. Standage, *The Turk: The Life and Times of the Famous Eighteenth Century Chess Playing Machine*, New York, Walker & Company 2002.
[113] A. Stoklosa, *NHTSA Finds Teslas Deactivated Autopilot Seconds Before Crashes*, in «Motortrend», June 15, 2022.
[114] A. Supiot, *Governance by Numbers. The Making of a Legal Model of Allegiance*, transl. bu S. Brown, Oxford, Portland, Hart Publishing, 2017.
[115] D. Tafani, *Sulla moralità artificiale. Le decisioni delle macchine tra etica e diritto*, in «Rivista di filosofia», 2020, n. 111, 1, pp. 81-103.
[116] D. Tafani, *L'imperativo categorico come algoritmo. Kant e l'etica delle macchine*, in «Sistemi intelligenti», 2021, n. 33, 2, pp. 377-393.
[117] Z. Talat, H. Blix, J. Valvoda, M. Indira Ganesh, R. Cotterell, A. Williams, *On the Machine Learning of Ethical Judgments from Natural Language*, in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics Seattle, 2022, pp. 769–779.
[118] C. Tessier, *Éthique et IA: analyse et discussion*, in *CNIA 2021: Conférence Nationale en Intelligence Artificielle*, par O. Boissier, 2021, pp. 22-29.
[119] N. Tiku, The Google engineer who thinks the company's AI has come to life, in «The Washington Post», June 11, 2022.
[120] Tribunale Amministrativo Regionale per il Lazio, Sentenza n. 07589 del 22 giugno 2021.
[121] E. Tucker, *Artifice and Intelligence*, March 8, 2022.
[122] A. Van Wynsberghe, S. Robbins, *Critiquing the Reasons for Making Artificial Moral Agents*, in «Science and Engineering Ethics», 2018.
[123] J. Vincent, *The AI oracle of Delphi uses the problems of Reddit to offer dubious moral advice*, in «The Verge», October 20, 2021.

[124] K. Yeung, *A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework*, Council of Europe, 2019.
[125] M. Young, M.A. Katell, P. M. Krafft, *Confronting Power and Corporate Capture at the FAccT Conference*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea*, New York, ACM, 2022.
[126] S. Wachter, *The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law*, in «Tulane Law Review», forthcoming.
[127] B. Wagner, *Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?*, in *Being Profiled: Cogitas Ergo Sum*, ed. by E. Bayamlioglu, I. Baraliuc, L.A.W. Janssens, M. Hildebrandt, Amsterdam, Amsterdam University Press, 2018, pp. 84-89.
[128] K. Weber, *Autonomie und Moralität als Zuschreibung: Über die begriffliche und inhaltliche Sinnlosigkeit einer Maschinenethik*, in *Maschinenethik. Normative Grenzen autonomer Systeme*, hrsg. von M. Rath, F. Krotz, M. Karmasin, Wiesbaden, Springer, 2019, pp. 193-208.
[129] J. Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation*, San Francisco, W.H. Freeman & Company, 1976.
[130] M. Whittaker, *Who am I to decide when algorithms should make important decisions?*, in «The Boston Globe», November 2, 2020.
[131] M. Whittaker, *The steep cost of capture*, in «Interactions», 2021, n. 28, 6, pp. 51-55.
[132] N. Wiener, *The Human Use of Human Beings. Cybernetics and society*, Garden City, New York, Doubleday, 1954.
[133] I.K. Williams, *Can A.I.-Driven Voice Analysis Help Identify Mental Disorders?*, in «The New York Times», April 5, 2022.
[134] O. Williams, *How Big Tech funds the debate on AI ethics*, in «The New Statesman», June 6, 2019 (updated June 7, 2021).
[135] B. Williamson, *Google magic*, in «Code Acts in Education», March 17, 2022.
[136] T. Wu, *The Curse of Bigness. Antitrust in the New Gilded Age*, New York, Columbia Global Reports 2018.
[137] S. Wyatt, *Metaphors in critical Internet and digital media studies*, in «New Media & Society», 2021, n. 23, 2, pp. 406–416.
[138] D. Zhang *et alii*, *The AI Index 2022 Annual Report*, AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, 2022.
[139] S. Zuboff, *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, New York, PublicAffairs, 2018.