AI and Ethics
https://doi.org/10.1007/s43681-021-00047-2

ORIGINAL RESEARCH

Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics

Johann Jakob Häußermann[1,2] · **Christoph Lütge**[2,3,4]

Received: 17 December 2020 / Accepted: 2 March 2021

Abstract
Today, due to growing computing power and the increasing availability of high-quality datasets, artificial intelligence (AI) technologies are entering many areas of our everyday life. Thereby, however, significant ethical concerns arise, including issues of fairness, privacy and human autonomy. By aggregating current concerns and criticisms, we identify five crucial shortcomings of the current debate on the ethics of AI. On the threshold of a third wave of AI ethics, we find that the field eventually fails to take sufficient account of the business context and deep societal value conflicts the use of AI systems may evoke. For even a perfectly fair AI system, regardless of its feasibility, may be ethically problematic, a too narrow focus on the ethical implications of technical systems alone seems insufficient. Therefore, we introduce a business ethics perspective based on the normative theory of contractualism and conceptualise ethical implications as conflicts between values of diverse stakeholders. We argue that such value conflicts can be resolved by an account of deliberative order ethics holding that stakeholders of an economic community deliberate the costs and benefits and agree on rules for acceptable trade-offs when AI systems are employed. This allows AI ethics to consider business practices, to recognise the role of firms, and ethical AI not being at risk to provide a competitive disadvantage or in conflict with the current functioning of economic markets. By introducing deliberative order ethics, we thus seek to do justice to the fundamental normative and political dimensions at the core of AI ethics.

Keywords AI ethics · Business ethics · Contractualism · Deliberation · Stakeholder engagement

1 Introduction

Today, due to growing computing power and the increasing availability of comprehensive, high-quality datasets, so-called artificial intelligence (AI) technologies are increasingly being used in almost all sectors and are thus entering many areas of our everyday life [1, 2]. Yet, the use of AI-based algorithmic systems raises ethical questions, calls societal beliefs into question and challenges many fundamental values [3, 4]. This concerns, for example, questions of discrimination and fairness, privacy and human autonomy in semi-automated decision making, risks of individual and social surveillance or threats to democracy through dynamic misinformation in social media and to human life through autonomous weapon systems or drones [5, 6]. Addressing the complex social, ecological and ethical consequences the development and use of AI systems might have, the emerging field of AI ethics seeks to establish normative approaches both on a theoretical as well as practical level which mitigate adverse effects and enhance the advantages of AI for the benefit of society.

Bringing together several different concerns about its evolution, we identify five crucial shortcomings of the first two waves of AI ethics. Based on this analysis, we introduce a business ethics perspective of deliberative order ethics claiming that at the core the use of AI systems may lead to fundamental value conflicts which to resolve AI ethics needs to be adequately equipped. In short, we argue that by too

✉ Johann Jakob Häußermann
johann-jakob.haeussermann@iao.fraunhofer.de

[1] Center for Responsible Research and Innovation, Fraunhofer-Institute for Industrial Engineering (IAO), Stuttgart, Germany

[2] TUM School of Governance, Technical University of Munich, Munich, Germany

[3] Chair of Business Ethics, Technical University of Munich, Munich, Germany

[4] Institute for Ethics in Artificial Intelligence, Technical University of Munich, Munich, Germany

Published online: 23 March 2021

narrow a focus on technical systems, current AI ethics tends to ignore the context of using AI, namely their integration into business practices and economic markets. The question then becomes how AI ethics could include a broader normative perspective which acknowledges the wider societal embeddedness of AI innovation. In response to this question, we advocate complementing AI ethics with a normative theory of business ethics that makes it both theoretically more solid and practically better applicable given the conditions under which AI innovation is (mostly) carried out today. Specifically, we present a contractualist approach of deliberative order ethics which stipulates that value conflicts triggered by the use of AI systems should be resolved by the stakeholders of an economic community deliberating and agreeing on mutually beneficial rules for balancing benefits and costs and acceptable trade-offs between diverse values. In this way, the ambition is to make the use of AI a matter of pluralistic value creation. Thus, acknowledging the political dimension of AI ethics, our approach of deliberative order ethics helps to address the fundamental normative questions raised by the use of AI in society [7–13].

This article proceeds as follows: in the following Sect. 2, we first outline the evolution of the first two waves of AI ethics before, we then aggregate five crucial shortcomings at the threshold of an emerging third wave. The next Sect. 3 proceeds by introducing the normative theory of order ethics and refining it in contrast to integrative social contracts theory (ISCT) as most proliferated theory of contractualist business ethics. Building on this, we then develop the concept of deliberative order ethics and discuss our approach in light of similar existing reasonings of the AI ethics debate (Sect. 4). We then examine whether or not and to which extent our proposal may successfully address the five shortcomings identified in Sect. 2 (Sect. 5). We conclude by summarising our reasoning and highlighting both its purpose and relevance as well as its limitations (Sect. 6).

2 Toward a third wave of AI ethics

2.1 From principles to practice

Although AI today is a highly interdisciplinary field, it can be described as a subfield of computer science which includes a range of technologies to create algorithmic systems that aim to reproduce human capabilities of intelligence [14]. Already established as a field of academic research since the 1950s, recent increases in computing power and the growing availability of large datasets allowed disillusions of the 1970s and late 1980s known as AI winters to be overcome. Today, it is particularly methods of machine learning and so-called neural networks that enable self-learning systems to be developed which, trained with the corresponding data, can ultimately perform even relatively complex tasks [15]. Based on different techniques of learning such as supervised, unsupervised, reinforcement or deep learning, AI thus allows the creation of algorithmic systems that assist humans by their ability to perform tasks in a highly adaptive and (semi-)autonomous manner. AI systems are already widely used in almost all sectors of society, from manufacturing, agriculture, trade, finance and medicine to government and public administration. Applications range from digital assistants such as chatbots, language translation tools, recommender systems of varying complexity in the consumer sector or professional contexts, to applications for autonomous driving or complex robotic systems and face recognition technologies. However, the enormous potential and the broad range of possible applications do not only promise economic and business value. Often there are far-reaching social consequences for individuals and society as well as the environment. Ethical issues in the development and use of AI systems are raised, for example, with regard to the protection of individual rights, autonomy and privacy, risks of biases and discrimination based on characteristics such as skin colour, race or gender, the lack of accountability of AI-supported decisions, or risks of undesirable individual or social surveillance. Mittelstadt et al. [5] and recently Tsamados et al. [6] describe six types of ethical concerns. In addition to traceability, these include epistemic concerns about inconclusive, inscrutable or misguided evidence on the one hand and normative aspects such as unfair outcomes and transformative effects on the other. One well-known example is the case of a recruiting tool developed by Amazon which was designed to identify the most suitable candidates among the applicants based on data on previous career paths within the company. However, as the system revealed to discriminate heavily against women and systematically favoured male applicants, Amazon had to withdraw it completely. Another high-profile case is provided by COMPAS, a system designed to help courts assess the risk of recidivism of defendants. Despite a high overall accuracy, however, it turned out that the probability of being wrongly assigned a high risk of recidivism was twice as high for a black offender than for a white offender ('false positive'), while white offenders were twice as likely to be wrongly assigned a low risk ('false negative') [16].

In view of the increasing use of AI and its vast influence on individuals and society, the debate about its ethical implications has attracted growing attention from the public, businesses, the academic community, and politics. To harness the benefits of AI while at the same time taking appropriate account of the ethical risks involved, a number of different actors from science [17, 18], politics [19–23], industry [24–26], as well as professional associations [27] and civil society [28] have developed principles and guidelines to enable the ethical and responsible use of AI. Although

their focus varies in detail, cross-cutting issues and trends can be identified. Jobin et al. [29] summarise a total of 11, Hagendorff [30] 6 and Floridi et al. [3] 5 overarching principles. Using different review methods, Jobin et al. [29] and Hagendorff [30] highlight the principles of transparency, fairness or accountability. The principle of transparency, for example, aims primarily at disclosing the functioning of AI systems to make results explainable and interpretable. In this way, damage can be averted, (legal) justifiability verified and trust strengthened [29]. The principle of fairness seeks to prevent undesirable bias and resulting forms of discrimination to ensure diversity and equality. Accountability aims to ensure that decisions are justified in a comprehensible manner and that the distribution of responsibility is clarified in advance. From a more integrative perspective, the different principles and guidelines have been summarised with regard to established principles of bioethics of beneficence, non-maleficence, autonomy, justice, and explicability [3, 31]. While beneficence is to ensure that the use of AI promotes overall wellbeing and is consistent with sustainability and the common good [32], the principle of non-maleficence aims to prevent potential damage caused by the use of AI [3]. In view of (semi-)autonomous systems, the principle of autonomy stipulates that people should always retain the last decision-making power or "the power to decide which decisions to take". Justice encompasses the effects that AI systems have on societies in terms of unfair discrimination, but also on social cohesion and solidarity, and aims to ensure that the costs and benefits of the use of AI systems are fairly distributed within society [3, 33]. Finally, the principle of explicability, which is the only one specifically for the context of AI, shall ensure that users and those affected by an AI system are able to understand and comprehend its results and that the distribution of responsibility is clear. We summarise the quest for principles and guidelines as the first wave of AI ethics. In view of the increasing use and impact of AI systems on individuals and society, it reflects the need to develop and use AI systems in line with a set of ethical values.

Even though the transition is certainly fluid, we take approaches tackling a concrete implementation of ethical AI as a second wave of AI ethics. One influential case is the ACM Conference on Fairness, Accountability and Transparency (ACM FAccT), formerly ACM FAT*, which evolved into an active community concerned with the ethical design of AI in close connection to relevant technical issues. In particular questions of explainable AI [34–39] or issues of fairness [40–47] have emerged as productive fields of research. But also more governance-oriented approaches to the practical implementation of ethical AI play an important part, for example with regard to a professional code of conduct for developers [48, 49], a more direct involvement of ethicists in the development of AI systems [50, 51], or in terms of checklists [52, 53], adapted internal structures [54], suitable impact assessment frameworks [55] and auditing processes [56] or a value-based AI label [57]. Finally, perspectives from the law concern the ethical design of AI at the interface with regulatory issues [58–62]. In summary, Morley et al. [63] provide a comprehensive overview of a variety of approaches and tools for the integration of ethical aspects in the development of AI systems. They develop a typology which relates the different approaches to implementing the five overarching principles according to Floridi et al. [3], and assigns them to seven phases of an algorithmic development process. Overall, we conclude that a first wave of AI ethics, in view of the impact on individuals and societies, has put forward appropriate ethical principles to guide the development and use of AI systems. The second wave builds on this and looks into how principles can be implemented and how guidelines can be put into practice. Although the difficulty of operationalisation and practical implementation is often emphasised [29], the variety of approaches presented indicate that there are nevertheless a number of promising efforts in progress.

Based on the first and second waves in AI ethics described above, we argue that there are indications of a third wave, the upshot of which is not yet clear. Based on critical analyses of its evolution, we identify five key shortcomings of current AI ethics which we discuss in the next section.

2.2 Five shortcomings of current AI ethics

Following the quest for appropriate ethical principles and initial considerations on their practical implementation, a number of concerns have been voiced about the ensuing trends in AI ethics. In the following, we will consolidate different concerns to delineate the current status of AI ethics. We argue that current critique can be summarised under five key shortcomings. First, AI ethics neglects the importance of business practices, without which, however, the ethical assessment of the use of AI systems is based on an incomplete picture [7, 8, 10, 30, 64]. Second, AI ethics is characterised by a form of technical solutionism which not only narrows the view of problems but also of options for action [7, 8, 10, 30, 63, 65]. Closely related to this we find, third, a focus on individuals, both in terms of the effects of AI systems and the responsible actors [7, 8, 10, 30, 63]. Fourth, the principle-focused approach of AI ethics faces problems in its practical implementation, on the one hand with regard to the necessary operationalisation of general principles, and on the other hand, in terms of accountability and guaranteeing the intended effects [7, 10, 29]. And finally, the unclear relationship between AI ethics and the legal regulation of AI is criticised, which, among other things, leads to AI ethics being misused by powerful corporations to prevent or at least delay legal action [7, 29, 66, 67]. In the following, we

discuss these five shortcomings and their relevance to the field of AI ethics.

2.2.1 AI ethics neglects the business context of developing and employing AI systems in society

Although seemingly trivial, it is worth noting that it is mostly firms that commercialise AI systems and introduce them to markets at the end of an innovation cycle. Against this background, it seems reasonable to assume that an ethical assessment should take into account the business context of AI systems. In fact, even if an AI system is completely ethically designed on a technical level—if this is possible at all and whatever that may mean with regard to say fairness, privacy or safety in particular [13, 44, 68]—major ethical questions may arise. Think, for example, of the risks of dual use or the cases in which employees from Google or Microsoft have voiced public protest against the potential use of some of their companies' products for immigration and law enforcement agencies, military purposes or foreign governments [64, 69, 70]. Or take the already widespread and various use of AI systems in recruiting, which raise questions about whether decisions about the future of people based on (psychological) profiling are legitimate and desirable. When and how are (semi-)automated decision-making processes about people's career prospects and opportunities for personal development societally desirable? Or, as Tasioulas [33], p. 65 puts it: "Consider, for example, the plight of long-term unemployed people whose job applications are routinely rejected by the automated systems that now dominate workforce recruitment. After months or even years of applying unsuccessfully for jobs, those individuals may never once have their application read and evaluated by a fellow human. Even if we assume that the relevant algorithm meets a good standard of functionality, i.e. it is just as effective, efficient and compliant with norms of appropriateness as the average human recruiter, the fact that it is a non-human mode of decision-making is worrisome. It is hard to pin down the worry very precisely, but the thought is roughly that the job seeker is subjected to a cold, alienating, and ultimately potentially disrespectful process because his application never comes to the attention of a fellow human being. So much is suggested in this extract from a recent *Guardian* article: "It's a bit dehumanising, never being able to get through to an employer," says Robert, a plumber in his forties who uses job boards and recruiters to find temporary work. Harry, 24, has been searching for a job for 4 months. In retail, where he is looking, "just about every job" has some sort of test or game, anything from personality to maths, to screen out applicants. He completes four or five tests a week as jobs are posted. The rejections are often instant, although some service providers offer time-delay rejection emails, presumably to maintain the illusion that a person had spent time judging an application that had already failed an automated screen' [71]". Hence, beyond issues of fairness or privacy questions arise as to whether its use may lead to ethically questionable business models, such as e.g. attention hacking [8], whether the use of certain infrastructures such as cloud services directly or indirectly promotes competition-distorting monopoly structures [cf. 72] or whether power balances relating to existing infrastructures are shifted through the use of AI [64].

The problem of too narrow a view can also be substantiated from a more technical perspective. Using AI methods such as machine learning, so-called optimisation methods are often applied, which can calculate different models and optimisation functions on the basis of training data and defined optimisation goals, since an analytical solution to the problem is not possible [23, 73, 74]. The use of optimisation technologies as a central element of AI systems shows two things. On the one hand, it illustrates that focussing on the individual protection of, say, privacy on a technical level does not allow the dynamic effects to be controlled in terms of profiling or manipulation of groups or societies [64]. Even if this means that companies are less interested in qualitative insights into individual data but only need the data for the statistical, probably even decrypted optimisation of services [64], this shows all the more that AI ethics' focus on the technical improvement of the system itself does not grasp the full picture. Instead, it is crucial to include questions about the acceptability of consequences, potential side effects and the legitimacy of a product, service or business model into the ethical evaluation. Second, the increasing use of optimisation technologies highlights the fact that their ethical evaluation is a complex and often inherently political undertaking which can only be answered through societal discourse and public deliberation. Take, for example, the at first sight rather innocuous optimisation of routes of public school buses in Boston [75]. It demonstrates that in addition to more efficient bus routes to reduce costs, traffic volume and CO_2 emissions, major health issues and different individual needs of children, e.g. with special needs, must be taken into account. The multiplicity of different variables to be included in an optimisation function poses an immense challenge to achieve a fair result with acceptable trade-offs with which those affected are satisfied [76–78].

A focus on the ethical design of AI systems on a technical level thus risks ignoring essential and fundamental aspects. Even if a system is technically mature and meets the highest standards of accuracy, fairness and privacy, its use may be ethically problematic because it overlooks trade-offs or may reinforce structural social injustices as in the case of predictive policing [10]. While ethical aspects at the micro-level of the technical system constitute a key element of AI ethics today, crucial business decisions and practices that implement these systems in products, services and business

models have been largely neglected. As a result, however, questions relating to, e.g. the concentration of power, practices of attention hacking, or concerning structural injustices such as institutional racism or problematic profit motives are ultimately not being addressed [8]. Moreover, too narrow a focus not only assumes that ethical challenges arise from flawed or inadequate design of the AI system [7, 8], but also limits the scope of possible options for action in the sense of a technical fix [7, 8, 30, 63]. The ethical relevance of business practices and the wider societal context shows that a focus on "better building" [8] is insufficient as ethical implications go beyond an ethical design of AI as a technical system and AI ethics cannot be "solved" but should rather accompany the use of AI continuously [7, 65]. In this light, approaches to "ethics by design" [79–81] may reveal similar limitations insofar as they are based on the assumption that ethical questions can be dealt with exclusively or predominantly at the level of the design of a system. The implicit assumption of moral causation in the sense that poor ethics on the part of the responsible developers are the source of bad designs which in turn produce harmful outcomes [8] reflects at least a limited understanding, in the worst case, it indicates more fundamental normative shortcomings. Although the relevance of conflicts between short-term profit interests and truly ethical AI have been recognised [7, 30, 63, 64], such aspects often remain outside the current focus on ethical design. However, this should not be seen as imposing an apparently incompatible opposition between business on the one hand and ethics or society on the other [cf. 30]. Nor should it mean that the commercial exploitation of AI is in itself ethically problematic. The point is that due to its narrow focus AI ethics does not include an integral part of AI systems as developed and employed in society without considering business models, business practices, their potential wider impacts and the general societal context which they are part of. The ultimate danger here is for AI ethics to become ineffective and powerless [66, 67, 82]. In response to this shortcoming, the challenge is, therefore, to expand AI ethics in such a way that the use and integration of AI systems in business practices and the necessary negotiation of legitimate (optimisation) goals and trade-offs can also adequately be taken into account.

2.2.2 AI ethics is biased toward a technological solutionism

Another reported deficiency of current AI ethics lies in the tendency to ignore the question of whether and when the use of AI systems may be less appropriate than another solution [7, 8, 10, 30, 65]. At least three different elements can be distinguished with respect to this type of technological solutionism. First, following a technically driven perspective, AI ethics seems to take technical progress and the development and use of AI as given and somewhat unchangeable [8, 10]. Yet this loses sight of the fact that technical progress always takes place within the scope of economic, political and social conditions. To the extent that technical advances are thus always the result of the societal conditions under which they are achieved, they are normatively shaped and not invariable. The development and use of new technologies like AI is, therefore, always informed by societal values, no matter how hidden they may be. This implicit adoption of technological determinism raises a second element. According to this, technological solutionism leads AI ethics to neglect the question of whether an AI system is in fact the most suitable and effective solution for the problem at hand [10]. The question as to when a (semi-)automated decision-making system is actually the best choice, whether human decisions may be useful in a specific case [30] or whether the cause of the problem is not rather to be found on a structural and systemic level [64, 83], is of utmost ethical importance. Or as Greene et al. put it [8], p. 2127: the "ethical debate is largely limited to appropriate design and implementation—not whether these systems should be built in the first place." Finally, third, this kind of technological solutionism implies restricting AI ethics to technical solutions to address ethical challenges. However, this not only limits the range of possible courses of action and levels at which changes are necessary for ethical AI. It also narrows the view of where and which ethical questions arise at all: when holding a hammer, everything looks like a nail. This tendency of a technical fix in AI ethics thus risks overlooking important ethical questions, curtailing complex, ethical questions and thus avoiding a wider societal debate [7, 8]. But fundamentally, as the examples above reveal, "AI ethics is effectively a microcosm of the political and ethical challenges faced in society" [7], p. 505. Recognising this means, among other things, that more emphasis must be placed on the question of the ethical appropriateness of (the use of) an AI system, e.g. in relation to the causes of the problem to be solved and possible (non-) technical alternatives.

2.2.3 AI ethics succumbs to an individualist focus

The first two points of criticism are closely linked to the aspect of an individualistic focus. As pointed out in the example of optimisation technologies above, AI ethics mainly examines the ethical implications in relation to individuals, i.e. whether the privacy of persons is sufficiently protected, whether persons are unfairly discriminated against or whether the results of AI systems are sufficiently comprehensible for its users. However, this overlooks ethically relevant effects that the use of AI systems may have on groups or society as a whole [63, 64]. While Morley et al. [63] highlight the role of trust, questions of societal monitoring, control and governance and their political impact, particularly on democratic societies, are often discussed in public debate

[84–86]. An overly individualistic focus, therefore, risks not addressing important ethical consequences at the societal level. In addition, Hagendorff [30] points to a noticeable omission of more often than not hidden social and ecological costs, such as the outsourcing of necessary labelling of data sets to so-called "clickworkers" or the extensive energy consumption caused by necessary hardware services. While this may be understood as a weakness in relation to the first wave of AI ethics, the problem also persists when it comes to the question of implementation. To the extent that there is a tendency to implement ethics in the sense of "better building" by means of technical solutions, mainly developers and data scientists are assumed responsible for ethical action. In addition to the application of appropriate technical measures, this is reflected, for example, in the development of professional ethics [48, 49], the teaching of ethics to AI practitioners [87] or tools such as checklists [52, 53, 88] directed at developers and data scientists. But also critical contributions, which rather belong to an emerging third wave of AI ethics, sometimes tend to argue with a focus on individuals as relevant actors for ethical AI [10]. The point is not to say that this perspective on appropriate action would be unjustified or ineffective—because it certainly is not. Instead, we wish to highlight that this form of an individual focus tends to lose sight of the role of the organisational level, i.e. of businesses, their strategies and business models, but also of questions of internal governance and corporate culture [cf. 65] as important levers for ethical AI. In the words of Mittelstadt [7], p. 505: "This approach conveniently steers debate towards the transgressions of unethical individuals, and away from the collective failure of unethical organisations and business models." Consequently, the lack of an individualistic orientation shows two things. On the one hand, a wider deliberative approach is needed to discuss and assess the complex social impacts appropriately. Second, the role of organisations and companies as actors should be given more weight as responsibility for ethical action should not be assigned at the individual level alone.

2.2.4 AI ethics is problematic in its implementation and lacks accountability and clear impact

The fourth weakness can be summarised as the problem of implementing ethical principles. One reason for this lies in the often very abstract and vague formulation of ethical principles, which leave room for different interpretations [29]. This results not only in the risk of divergent interpretations, unclear claims and negative effects on trust [63], but also in a rather vague basis for attempts of operationalisation and implementation in legal, organisational or technical contexts. Beyond that, the challenge of translating abstract ethical principles into specific requirements may be one reason for the focus on technical solutions, given that technical parameters provide precise specifications for the implementation of ethics. On a social, political, legal, governance or corporate culture level, the field of possible measures and methods of implementation appears to be much more diverse—and hence much more complicated. The abstract formulation of ethical principles thus leads to the considerable difficulty of developing and implementing approaches for their practical implementation [9, 29, 30, 62, 63, 82], not least on a legal level [61]. Besides suitable tools and measures, this also applies to the definition of responsible actors and accountability structures that ensure that the principles are complied with at all [7, 29]. While appropriate approaches to the implementation of AI ethics are urgently needed, their mere existence is not sufficient. Effective structures and robust processes need to be established, evaluated and documented to enable a sustainable impact of AI ethics [7]. What this shortcoming of AI ethics shows is not only the difficulty of putting ethics into practice, especially in a business context. It also points to the fundamental discrepancy between normative goals and practical approaches, often due to the lack of an explicit and theoretically sound normative framework [10] to justify both particular normative goals and the means of their effective implementation. As a consequence, a third wave of AI ethics should focus on substantiating normative goals based on a solid theoretical foundation to derive practical approaches and counteract a gap between formulated principles and their practical implementation.

2.2.5 AI ethics lacks a clear relationship to legal regulation

Finally, a fifth weakness can be summarised as the often unclear relationship to the legal regulation of AI systems. The dynamism with which the first and second waves of AI ethics were triggered and large technology corporations dominated the resulting public discourse led to concerns that industry could determine the ethical standards to be applied to AI [67]. Although the concern is closely linked to the economic power of many large corporations, the often conceptually ambiguous relationship of AI ethics to pivotal legal issues, such as the impact of AI on existing legislation or the need for further legal regulation, contributes to this concern. Rességuier and Rodrigues [66] argue that this is due to an underlying law conception of ethics which misunderstands the role of ethical principles and thus risks the practical effectiveness of AI ethics. Beyond mere virtual signalling, the ambiguous use of "ethics" on a communicative level may be tactically exploited to influence the public debate and prospective legislation. An undefined relation of ethics and law thus risks AI ethics being misused to soften, delay or prevent hard legal regulations [7, 10, 29, 62, 64, 89]. Moreover, even at the political level, the relationship between ethics and law sometimes seems to be unclear in the context of AI [89], which may become problematic in view

of the time delay in legislative processes reacting to rapid technological developments such as AI [61]. Although the danger of "ethics washing" thus seems reasonable [6, 10, 61, 63–67], it is important to note that from a conceptual perspective the relationship of ethics to questions of legal regulation can be considered complementary in principle [90]. In contrast to legal legislation, ethics is particularly helpful when legislation is unavailable, requires ethical interpretation or counterbalance, or when something that is (still) legal should be avoided for ethical reasons, or something not yet legally required should be done for ethical reasons [82]. Even though AI ethics lobbying, that is "the malpractice of exploiting digital ethics to delay, revise, replace, or avoid good and necessary legislation (or its enforcement) about the design, development, and deployment of digital processes, products, services, or other solutions" [82], p. 188, poses a significant risk of undermining serious ethical efforts, a conceptually clear demarcation from questions concerning the legal containment of AI systems is possible. In short, we conclude that the fifth identified shortcoming requires a third wave of AI ethics to clearly determine its legitimate role and promote appropriate communication activities. Building on a solid normative foundation, AI ethics should thus describe both its tasks and limitations.

As this review of the evolution of AI ethics demonstrates, at the beginning of a third wave some key steps need to be taken to ensure that AI ethics can make an effective long-term contribution to technology, the economy and society. Based on the principles-led approaches of the first and manifold efforts for the practical implementation of the second wave, ethical implications of AI-based business models and business practices on a societal level need to be brought more into focus. In addition to "better building" [8] the goal of "better managing", in the sense of considering the wider social, economic and ecological consequences, needs to become a key element of AI ethics. The call for a transition to "microethics" [30, cf. 91] should, therefore, be complemented by a perspective of "macroethics", which deals with the ethics of products and services at the level of markets and the organisational relationship between businesses and society as a whole. Second, this includes to extend current approaches that take the development and deployment of AI systems as a given and irreversible fact and concentrate on technological answers, so that the deployment of AI systems as such can be reflected and wider options for action are enabled. Third, a third wave of AI ethics needs not only to take greater account of the wider impacts on societies but also focus on businesses at an organisational level as responsible actors for ethical behaviour. Future approaches to AI ethics should finally adopt conceptually clear and transparent demarcation of the legal regulation of AI and openly address challenges in implementing ethics. Besides practical approaches for everyday business, the implementation of AI ethics should also comprise issues of effectiveness, accountability, and the justification of both normative goals and proposed measures. In conclusion, the five concerns as described above point to the weakness of current AI ethics in recognising fundamental normative challenges and acknowledging the inherent political dimension of AI ethics [7, 8, 10, 13]. This manifests itself in the neglect of the business context, a strong focus on ethical design and a primary attribution of responsibility to individuals such as developers. Against this background, we aim in the following chapter to offer a first step towards complementing AI ethics by drawing on established normative theories from business ethics.

2.3 Order ethics as business ethics approach to AI

In this chapter, we present a contractualist theory of business ethics arguing that it provides a suitable normative approach to AI ethics. Since a comprehensive introduction to the philosophical foundations of contractualism or business ethics is beyond the scope of this article, we focus on the aspects essential to our reasoning. We first introduce the concept of order ethics and then contrast it with integrative social contracts theory (ISCT) as the most prominent example of contractualist business ethics.

Business ethics deals with the question of the possibility of ethical behaviour in a market economy which is driven by the principle of competition [92]. Despite early contributions on AI from a business ethics perspective [93] and the fact that the impact of AI on business ethics has been recognised [94, 95] and conceptualised by several authors [69, 88, 96, 97], business ethics approaches are hardly found in the current AI ethics debate [69, 98], but in no case from a contractualist perspective.

As a concept of contractualist business ethics, order ethics refers to constitutional versions of contractualist theory [99, 100] which provide for the fundamental attribution of basic rights, e.g. based on human rights, via a constitutional contract, and thus go beyond more reductionist approaches of contractualism building on J. Locke or R. Nozick. Although to some extent similar to J. Rawls' contractualist *Theory of Justice* [101], two key distinctions can be made [102]: first, the negotiation of contractual conditions does not take place in an idealised setting behind a veil of ignorance but is shifted to the real-world situation of business ethics. Second, a constitutional version of contractualism does not seek to derive normative principles that determine a just social order but reflects solely on the normative foundations of economic action based on the assumption of self-interested persons. In this sense, it is an economic approach to business ethics that promises to be particularly compatible with business practice.

The starting point of order ethics are value conflicts which are addressed and aimed to be resolved from a

contractualist perspective. As will be shown, it is this fundamental approach that makes order ethics a promising complement to AI ethics. Confronted with conflicts between different values—ranging from individual interests, to social norms or ethical values—order ethics assumes that no recourse to a certain substantial normative principles is possible, however, they may be defined and justified in advance in any form, but that a solution can only be reached by agreeing to a rule for the benefit and in the interest of all parties involved. Ethical conflicts in this sense are to be negotiated and resolved only through a solution that settles the conflict in the sense of a voluntary agreement on the basis of individual consent, but not through reference to higher normative principles. In this sense, the contractualism of order ethics is both more and less ambitious than Rawls' understanding of it [102]: less ambitious, as no attempt is made to justify overarching normative principles, and more ambitious, as this means that contractual renegotiation in the face of ethical conflicts takes place under real-world conditions with all the associated entanglements and complications. For order ethics, the level on which agreements are made is essential. Based on the distinction between action and rules [100], order ethics holds that ethical conflicts can only be resolved in a justifiable manner at the level of the conditions for action. In this way, order ethics responds to the risk of ethical behaviour being crowded out, since more often than not it is not rewarded at the level of individual actions in a competitive environment [103]. Typically conceptualised in the form of the prisoner's dilemma, order ethics thus reflects the problem of cooperative behaviour (in competitive markets): only if ethical standards are set at the level of rules can individual ethical behaviour be reasonably required since otherwise they will be subject to some form of sanction. With this in view, order ethics advocates the following concerning the notion of rules [citing 100, 102], p. 692:

1. Only changes in rules can change the situation for all participants involved at the same time.
2. Only rules can be enforced by sanctions—which alone can change the incentives in a lasting way.
3. Only by incorporating ethical ideas in (incentive-compatible) rules can competition be made productive, making individuals' moves morally autonomous in principle. With the aid of rules, of adequate conditions of actions, competition can realise advantages for all people involved.

First of all, rules need not be understood in a narrow economic or political sense as they can also be drawn from ideas from other areas of society such as culture, philosophy or arts [102]. What is further important is that corresponding rule changes or new rules designed to resolve ethical conflicts do not conflict with individual actions, so that no counteracting incentives on the level of rules arise [104]. The shift from ethics to the level of rules means that ethical conflicts should be clarified by deriving more general rules that apply not only to the specific individual case at hand but to at least one specific group of conflicts and actors. In this sense, it is about finding rules of distribution of goods and not about determining one particular distribution of goods [102]. Not least, an agreement on the level of rules facilitates the consent of all parties involved. Although order ethics thus underlines the importance of an appropriate general framework for ethical behaviour through the concept of rules, rules should not be put into one with laws. Instead, order ethics seeks to provide a conceptual supplement to laws and the general legal framework based on the theory of incomplete contracts [105] which may also raise ethical conflicts. Incomplete contracts occur, for example, when obligations are not sufficiently clarified, when it is difficult to assess whether a contract has been complied with or when its enforcement is difficult [104]. To the extent that it is impossible to adequately equip all contracts for all possible future scenarios and to amend incomplete contracts, their occurrence is necessary and cannot be avoided. The resulting scope for interpretation of legal contracts, which deal with complex issues or claim validity over a long period of time, should thus not be seen as a shortcoming but rather as an advantage in dynamic environments by allowing flexibility and adaptability. Order ethics understands the role of ethics in managing the openness of incomplete contracts, including the resulting uncertainty and possibly emerging conflicts [102, 104, 106]. This allows order ethics to define the place of ethics and to specify its relationship to legal regulations. "Order", therefore, does not refer to the legal framework but to all other formal and informal rules and agreements which seek to enforce ethical behaviour, for example at the level of individual sectors or groups of firms [98].

By shifting the focus of ethics to the level of rules, order ethics finally emphasises the contractualist criterion of mutual benefits [102, 106, 107]. Accordingly, given the absence of overarching normative principles, only such an agreement can be normatively justified which offers benefits for each individual or party involved on the basis of his or her individual values and interests. In this context, possible advantages are to be understood broadly and include not only monetary or financial benefits, but everything that people take to be advantages [106]. In practical terms, firms should resolve ethical conflicts arising, for example, from previously missing, impossible or unintentional legal regulations by means of adapted or new rules, which are in the interest of every stakeholder involved and thus generate mutual benefits. For only when real win–win situations are created [102], a normatively justified solution can be claimed. This does not imply that firms should abandon a

business management perspective but rather that they must improve their economic calculations by incorporating the values of various stakeholders and, for example, taking into account long-term effects on reputation [106].

In a nutshell, the core elements of order ethics can be summarised as follows:

1. Building on contractualism as normative theory, order ethics argues that ethical conflicts cannot be resolved by reference to overarching normative principles (reasonable pluralism).
2. Instead, ethical conflicts ought to be solved by adapted or new rules to which each stakeholder involved consents based on their individual values.
3. The normative criterion is the mutual advantage that is to be achieved by a respective agreement.

We conclude the brief introduction of the concept of order ethics by highlighting some of its main advantages. First, the concept of order ethics is rooted in a fundamentally pluralistic view of society. According to this, a multitude of different values can be legitimately held, which ultimately may come into conflict with each other. In resolving these conflicts, no shared basis of common values of any kind should be assumed but rather each individual value is accepted as normatively justifiable. This offers a key advantage over other ethical theories of business ethics such as utilitarian approaches. Roughly speaking, the latter assume that in the face of an ethical conflict, the option that yields the greatest possible (measurable) benefit should be chosen. However, this not only bears the risk of delivering highly counterintuitive results but more importantly it requires the maximisation of utility, however, defined and justified as universal ethical norm. Second, a contractualist approach seems to be better equipped than stakeholder theories of business ethics to reconcile claims of different stakeholders, balance incommensurable conflicts of values or solve problems of collective action, given that contract theories were originally formulated to address these very issues [108]. Third, by aiming at rule changes, crucial constraints of operating in a competitive environment can be taken into account. For as Morley et al. [63], p. 2161, note, it is highly plausible that not least in the context of AI, an ethical approach would constitute a competitive disadvantage for any single "first mover". Fourth, order ethics allows us to specify the relationship between ethics and legislation, which is of particular relevance to AI ethics. Before we discuss the implications in more detail at the end of this chapter, we first contrast the introduced concept of order ethics with the probably most prolific theory of contractualist business ethics, namely integrative social contracts theory (ISCT).

2.4 A cursory comparison of two contractualist theories of business ethics: order ethics and ISCT

Just as order ethics, integrative social contracts theory (ISCT), originally developed by Donaldson and Dunfee [109–111], stems from a contractualist basis. For ISCT too, the central question is how conflicts between different or differently prioritised values and norms can be overcome. In contrast to order ethics, however, ISCT assumes a macrosocial contract which sets the conditions for microsocial contracts. Although Donaldson and Dunfee do not assume a strong hypothetical setting in the sense of Rawls' veil of ignorance as the contractors know at least their basic preferences and values, they nevertheless assume that "information about their personal economic endowments and roles in society" [112] is unknown. Under the four terms of the macrosocial contract, the following conditions are set out [109, 110, 112]:

1. Local communities may specify ethical norms for their members through microsocial contracts (called "moral free space").
2. Norm-generating microsocial contracts must be grounded in informed consent buttressed by a right of community members to exit and to exercise voice within their communities.
3. To be obligatory (legitimate), a microsocial contract must be compatible with hypernorms.
4. In case of conflicts among norms satisfying principles 1–3, priority must be established through the application of rules consistent with the spirit and letter of the macrosocial contract.

According to ISCT, the actual discussion of ethical conflicts is moved to the level of microsocial contracts, which all members of a local community must agree to for the agreement to be considered an authentic norm. The members of a community have the right to leave the agreement and to give voice to their position. Individuals may be members of several economic communities, defined as "self-circumscribed group of people who interact in the context of shared tasks, values, or goals and who are capable of establishing norms of ethical behaviour for themselves" [110], p. 262. Decisive for the legitimacy of microsocial contracts is their compliance with so-called hypernorms, certain universal ethical principles such as those expressed in human rights [109, 110]. These can be either procedural hypernorms such as the right to exit and voice, substantive hypernorms such as respect for human dignity or structural hypernorms such as the right to property or necessary social efficiency [112]. Lastly, ISCT stipulates that conflicts between microsocial norms will be resolved by so-called priority rules, provided

that they are in line with hypernorms. A total of six such rules decide how to deal with conflicts in case of doubt. A large part of the practical implementation of ISCT, besides the identification of stakeholders of a community, rests thus in the empirical determination of microsocial authentic norms [113] and the identification of relevant hypernorms. Ever since its original introduction in the mid-1990s, ISCT was criticised [e.g. 108, 114–120] and defended [112, 121] and has become an important cornerstone in the debate on contractualist business ethics.

Starting from this rough summary of ISCT, some similarities and differences to the theory of order ethics can be noted, given they both provide approaches of contractualist business ethics. Fundamentally, both concepts bear similarities in their contractualist foundation, according to which they start from two different levels, constitutional and post-constitutional rules based on Buchanan [99] in the case of order ethics and macro- and microsocial contracts in the case of ISCT. Furthermore, against the backdrop of their similar theoretical framework, both approaches emphasise the role of individual consent for the legitimate validity of rules or authentic norms, with ISCT particularly emphasising the role of consent in the sense of engaging in a practice [110]. Beyond these underlying similarities, however, significant differences in the way the concepts are further elaborated can be identified. Most noticeable appears to be the handling of ethical conflicts. While order ethics seeks to resolve conflicts through mutually advantageous rule changes, ISCT establishes hypernorms, i.e. universally justified principles [102]. Referring to third normative principles, however, creates serious problems of justifiability, legitimacy and empirical identification given the assumption of reasonable pluralism. In addition, in practice, the identification of hypernorms seems to result in a much more complicated process as it involves high justification standards. In this light, the concept of order ethics seems to be more suitable for the context of AI as it works on the grounds of weaker normative requirements. The second difference we notice concerns the perspective from which ethical conflicts are approached. ISCT considers these conflicts to be deficiencies of a market economy and that they should be corrected accordingly [102]. Not least, this limits the scope of possible outcomes of ISCT to standards like code of conducts and results in a mechanical approach to business ethics [122], which, as we have shown above, is not adequate for the context of AI. Order ethics, on the other hand, aims to achieve mutual benefits for all stakeholders of an ethical conflict through rule changes and in this sense, it strives to ethically improve the market economic system. Again, order ethics seems to offer a better approach for the context of AI. Because especially in a dynamic and rapidly developing field of technology, it is important to actively shape innovations through ethics. This type of productive perspective is facilitated by an opportunity-oriented approach rather than an approach geared to remedying deficiencies. Moreover, solving emerging ethical conflicts through a set of six priority rules seems to present a somewhat rigid [116, 118] and probably conservative [123] framework, which appears to be ill-suited for the dynamic context of AI.

Nevertheless, we would like to point out one aspect which we think is worth being added to order ethics from the concept of ISCT in the context of AI. This concerns the characterisation of economic communities as respective subjects of ethical decisions. We find that this conceptualisation fits particularly well into the concept of order ethics as it provides a suitable starting point for its procedural expansion in the context of AI. In the next section, we will argue for a procedural amendment of order ethics providing a practical method to deal with ethical conflicts between values and interests in the context of AI.

3 Community-in-the loop: the concept of deliberative order ethics

3.1 Bringing business ethics to AI: a procedural extension of order ethics

In the following section, we will introduce order ethics as a theory of normative business ethics to the field of AI. To this end, we advocate that order ethics provides a suitable framework of normative business ethics to complement AI ethics as presented in the first part of the paper. However, we also argue for a procedural addition through deliberative stakeholder engagement that provides a suitable methodological extension to debate value conflicts and agree on trade-offs via adequate rules.

The starting point of order ethics is the question of how to deal with ethical conflicts that may arise for firms given the competitive environment of international market economies. We believe that this approach to ethics provides a valuable addition to the predominant perspective of current AI ethics considering the shortcomings as identified above. As Wempe [124] explains, ethical conflicts between different norms and values may arise due to globalisation, increasing complexity, increasing specialisation. This applies especially to the context of AI. Importantly, the perspective of ethical conflicts allows issues beyond ethical design to be brought into the focus. Since besides conflicts between accuracy, accountability or fairness [125], in particular conflicts between very diverse and complex issues have to be taken into account when assessing an AI system. Ultimately, the assessment of conflicts between different values, norms or interests is about determining the necessary trade-offs and negotiating which solution and distribution of costs and benefits is acceptable for all parties involved. Some of

these conflicts are already inherent in the concepts currently employed by AI ethics, such as fairness or privacy, the application of which, therefore, requires a thorough normative analysis [7, 13, 35, 68, 126, 127]. Other potential trade-offs include, for example, those between the intended purpose of an AI system and resulting costs for employees in terms of layoffs or training, costs for suppliers or other partners in terms of systemic risks or resulting dependencies, complex social or ethical costs in terms of gains in flexibility, risks to surveillance and privacy, direct or indirect costs to society through monopolisation effects or beneficial alternatives that are being pushed aside, or costs for the environment from energy consumption or the mining of raw materials. While some of these conflicts and trade-offs may be explicitly considered and perhaps even included in the cost calculation of a system and business model, such as the risks of safety and security, others, especially unintended and longer-term consequences, are often difficult to identify at all [11]. Take the example of the above-mentioned recruiting systems. What effects does the increased use of AI-based recruiting systems have on applicants, on the human resource management in firms and on the labour market in general? Under what conditions does their use seem acceptable to all stakeholders in the long term? The example of optimisation technologies illustrates the complexity of the conflicts: how should benefits and drawbacks for children, parents, teachers, schools, public administration and bus companies in terms of health effects, cost and time savings and environmental effects be best organised for all stakeholders? Along similar lines, Whittlestone et al. [9] describe such conflicts as tensions with which AI ethics is confronted. By summarising four such key tensions in general terms, they highlight the challenge of assessing costs and benefits. It becomes clear that the identification and judgement of such value conflicts is a political task by its very nature, which involves the social negotiation of different values, conflicts and trade-offs [9].

Insofar as the contractualist theory of order ethics starts out from precisely such ethical conflicts, the approach seems particularly apt to complement AI ethics at this point. As order ethics is based on reasonable pluralism respecting the multitude of values that prevail in society, no substantial basis in the sense of a certain set of shared values is assumed. For the context of AI, this means that all values and interests must be given equal consideration in emerging conflicts, without any of them being in any way given a lower valuation than others. Nor would it be possible to reduce conflicting values to some kind of common basic value. No matter how great a challenge this presents for order ethics, it is essential to recognise the pluralism of values. Order ethics now provides for agreements on the level of rules to which the stakeholders involved agree on the basis of the normative criterion of mutual advantage [106]. At this point, we propose to add an important procedural element to order ethics to develop and agree by means of participation and deliberation on a suitable measure at rule level, in which all stakeholders can realise their values, i.e. achieve benefits of some kind. Our proposal thus amounts to the following: to deal with conflicts between different values and to arrive at an assessment of trade-offs and a fair distribution of costs and benefits of an AI-based product or service, stakeholders responsible for or affected by a business practice should formulate a rule through a deliberative participation process to which all can agree on the basis of their own interests. We argue that participation is the appropriate method for deriving an eligible rule (or set of rules) since the legitimate interest of stakeholders is already manifested in the criterion of mutual advantages and the collaborative, co-creative development of an eligible rule is, therefore, the most effective way to meet it. To negotiate complex value conflicts in the context of AI, the participatory involvement of stakeholders as well as the cooperative consultation is necessary since only in this way relevant values and interests as well as diverse consequences and benefits and costs can be identified, and ultimately legitimate trade-offs balanced. First, only in this way can the diverse values and potential costs be determined because for most of them there are neither any validated data or parameters nor standards for their evaluation. Second, only a deliberative process allows decisions to be made on whether trade-offs are acceptable and whether the balance between advantages and disadvantages is societally desirable. The question of which social groups (e.g. children, children with special needs, teachers, and bus drivers) should benefit or bear which disadvantages, how health effects should be weighed against cost savings and environmental improvements are complex societal negotiation processes. Similarly, in the case of AI systems in recruiting, questions may arise such as how to reconcile efficiency gains for firms with potential benefits and harms for certain groups of applicants, potentially increasing dissatisfaction and emotional distress for applicants, or with increased insecurity in labour markets. Deliberating about the different costs and benefits for respective stakeholders is, therefore, a suitable approach to prioritise values, decide on trade-offs and thus do justice to the political dimension of the problem. Only through participatory and deliberative exchange can a societal consensus and, building on this, an agreement be found which provides acceptable benefits for all. In our view, participation and deliberation are the appropriate methodological strategies to make AI ethics, within the framework of order ethics, a societal and political debate on the consequences of AI-based business practices at the level of organisations and actors. With the procedural supplement to order ethics presented here, we hope to adapt the crucial step of rule changes to the context of AI ethics.

One important element is the question of legitimate stakeholders. For order ethics, it is central to develop ideas on

the level of rules, so that ethical behaviour does not cause a competitive disadvantage for individual actors. Depending on the individually defined scope, stakeholders may be, for example, those who belong to a specific industry or a specific area of application of AI systems, such as AI in recruiting or human resource management, or AI in the public sector or for public infrastructures. Stakeholders include those involved in the development and employment of AI systems as well as those potentially affected, in particular specific groups from civil society. Ultimately, the identification of relevant stakeholders depends on the precise definition of the specific scope that the rule to be developed should cover. It is likely, however, that this can only be finally determined in the participatory deliberation process itself, as it is often anything but trivial to decide at which level a rule is effective and compatible with competition. For this purpose, we suggest borrowing the term community from ISCT [109, 110] to describe as an economic community a group of stakeholders who are interested in the ethical governance of AI systems on the basis of a shared interest in a specific field of application.

Furthermore, the issue of rules is essential. In general, rules can be drawn from a wide variety of conceptual ideas, and therefore, do not need to be legitimised by a specific legal, political or economic background [102]. Rather, the aim is to give voice to the pluralism and capabilities of deliberative participation processes through creative rules. The only requirement is that beyond resolving one individual case and assessing the costs, benefits and trade-offs of a concrete AI system, the rules must apply to at least a certain group of corresponding products, services or AI-based business practices. With regard to the examples consulted, this might include rules for AI-based business practices in recruiting or human resource management.

The participatory and deliberative extension of order ethics can be further explored in the light of some critiques of ISCT, which argue that its rather static approach is not sufficiently equipped for dynamic contexts of changing norms and conflicts [116, 118, 122, 128]. Burg [122], for instance, analyses ISCT's concept of authentic norms and criticises Donaldson and Dunfee's recurrent recommendation of corporate codes as an appropriate measure. According to him, this form of "mechanical business ethics" seems problematic: "At their best, codes are merely levers for internal and external stakeholders to hold organisations and organisational actors accountable by stating what is obvious to nearly everyone. At their worst, codes present an ethical façade that is only marginally related to manifest organisational norms, to be treated as the punch line of a joke about how one should behave within an organisation ('Check the code of conduct!')" [122], p. 675. Not least, this point is reminiscent of the problems of a too principled approach to AI ethics described above. Alternatively, Burg advocates an approach of deliberative business ethics which establishes and prioritises norms by an open process of stakeholder dialogue and ultimately reaches agreements based on consent. Similarly, Phillips and Johnson-Cramer [116] have criticised the lack of dynamism in ISCT arguing that the described mechanisms of exit and voice do not adequately reflect the dynamic processes of norm evolution. For a dynamic addition to ISCT they propose four principles, including the principle of community discourse "to create systems for the exercise of voice" [116], p. 298. Calton [118] also formulates a more dynamic and process-oriented supplement to Donaldson and Dunfee's ISCT. According to him, ISCT's reference to hypernorms and the defined priority rules are too inflexible and thus unsuitable to deal with the manifold and dynamic value conflicts in a pluralistic context. He introduces a dialogic twist, allowing stakeholders to find a fair agreement in an interactive learning process. Such a dynamic dialogue process is able "to unleash the full reflective potential of a social contracting theory of business ethics" [118], p. 344. Overall, it can be noted that ISCT, as the most advanced theory of contractual business ethics, has already been enriched by various participatory and deliberative approaches [108, 119–121]. We argue that the advantages of such an extension can also be applied to order ethics in the context of AI. However, since the focus of this article is not on a conceptual extension of order ethics, some essential issues have to remain outstanding. What remains to be clarified, for example, are the specific criteria for the identification of stakeholders [cf. 116], whether particular types of rules may be differentiated and what requirements for consent may be derived. Here, it can at least be stated that consent in the sense of an ongoing collaborative process [cf. 13, 122] would not only enable a constant monitoring and adjustment of rules for rapidly changing business practices, but could also play an important role in terms of accountability. Furthermore, the challenges of process design and the different starting conditions must also eventually be addressed. In particular, standards must be set that adequately take into account the heterogeneity of the stakeholders involved. How can different levels of knowledge, power imbalances and different cultures and languages be managed in such a way that a fair deliberation process is possible? While it is clearly worth building on prior work from related fields, future research would need to further specify the participatory deliberation process and relevant criteria.

Since a more comprehensive explanation of the proposal goes beyond the scope of this article, we summarise our reasoning as follows:

1. When introducing AI-based business models, **conflicts** between different values, norms and interests may arise over the distribution of benefits and costs of deploying AI.

2. To decide on a societally desirable distribution of costs and benefits and agree on acceptable trade-offs, the **deliberative participation** of the relevant economic community is necessary.
3. Through engaging in a participatory process of deliberation, the economic community, i.e. stakeholders of using AI in a particular field of application, ultimately establishes overarching **rules** that enable ethical behaviour without creating competitive disadvantages.
4. Assuming the same legitimacy of the different interests and values, the decisive normative criterion is that all stakeholders of the community agree to the rules on the basis of **mutual advantages**.
5. As a result, rules are to be created through inclusive deliberation of the economic communities, enabling ethical AI business practices in the sense of **pluralistic value creation**.

In other words, our proposal is to complement the third wave of AI ethics with a stakeholder engagement approach, according to which, whenever conflicts arise between different values, firms engage in a participatory and deliberative process with the relevant economic community to develop rules that enable ethical behaviour in the field of given business practices. Stakeholders of an economic community refers to all such parties who are involved in any way related to the use of AI systems in a specific field of application, whether as developers, users, or affected person or group of civil society. What is crucial here is the theory-based normative criterion of mutual benefits on the basis of which stakeholders consent to an agreement. Only if all stakeholders recognise satisfactory gains in the ratio of costs and benefits, and in this sense win–win situations are created, may a rule legitimately claim validity and be considered as enabling ethical business practices. While in the case of public school bus services, it seems rather intuitive that all parties involved should benefit from the introduction of an AI-based system, this becomes even more complex in the case of AI-based recruiting systems. What follows is that not only, say, developers and firms as users but also potential applicants must benefit from the use of the respective systems. What is thus characteristic is the aim of creating shared value instead of unilateral business value in terms of financial profits for firms involved. This is in line with Schormair and Gilbert [129] who present a framework for creating shared value in situations of value conflicts among stakeholders. Comparing approaches of agonistic and deliberative stakeholder engagement, they argue for an integrative approach based on a process of discursive justification. Recognising stakeholder value pluralism, they develop a five-step procedural framework which helps to resolve value conflicts by steps of discursive sharing and potentially leads to pluralistic stakeholder value creation. It is worth pointing out that both approaches do not attempt to resolve the problem of value conflicts by referring to monistic normative theories or consensus in the sense of an agreement on single values, but rather seek to realise different values and thus mutual benefits within a procedural framework. For AI ethics, this means that it should not (only) be about "solving" ethical problems but also the creation of more comprehensive benefits or even the promotion of the common good. The perspective of pluralistic value creation helps to establish "AI for good" [32] not as a subfield but as the core prospect of AI ethics.

4 Discussion: contractualism, deliberation and AI ethics

In the recent debate on AI ethics, a few different contractualist or deliberative ideas have been put forward [9, 11–13, 130].[1] In the following section, we discuss some of these proposals to refine the approach of deliberative order ethics as outlined above.

Most prominently, Rahwan [11] has introduced social contract theory to the AI ethics debate by arguing for a conceptual framework of society-in-the-loop (SITL). Based on the paradigm of human-in-the-loop (HITL), he applies the idea that at some point of the algorithmic system a human is involved to provide monitoring and supervisory functions, and adapts it to a more general societal level. As Rahwan [11], p. 7, puts it: "While HITL AI is about embedding the judgement of individual humans or groups in the optimisation of AI systems with narrow impact, SITL is about embedding the values of society, as a whole, in the algorithmic governance of societal outcomes that have broad implications." Recognising the ethical and societal implications AI systems may have, he makes use of social contract theories in a broader tradition referring to Hobbes, Locke and Rousseau as well as to Rawls and Gauthier in modern times as an adequate framework to deal with fundamental value conflicts and the question to find fair distributions of costs and benefits and acceptable trade-offs. As societies today become increasingly governed by AI-based algorithms, SITL seeks to expand the general social contract to the realm of algorithmic and AI-assisted decision making. To be able to agree on acceptable trade-offs, both quantifying externalities as well as ways to articulate values and societal expectations are needed to evaluate AI systems. Rahwan then discusses different methods and techniques that

[1] Contractualist arguments in the context of AI have also been put forward on a technical level or looking at specific ethical challenges, such as justifiability [135] or autonomous vehicles [136]. But also from a perspective of discourse ethics criteria of rational communication have been developed to manage algorithmic accountability [137].

could be used to "bridge the society-in-the-loop gap", from value-sensitive design to crowdsourcing and data-driven tools such as computational social choice, as well as deliberation between stakeholders and public engagement [11], pp. 10–11. While he is not proposing one specific methodological avenue to resolve value conflicts and agree on trade-offs, he nevertheless seems to recognise the significance of public engagement as not only experts alone can decide on societal values, but it is precisely through interaction and deliberation that values and norms emerge and can eventually be agreed upon.

In general, Rahwan's and our approaches presented above have much in common as they both seek to provide an ethical framework for AI based on the normative theory of contractualism. Against this background, our account of deliberative order ethics is similar to Rahwan's SITL in the conceptualisation of AI ethics as conflicts between different values and interests with regard to the trade-offs AI systems may imply. While both approaches use contractualist theory to address necessary trade-offs, we focus on stakeholder participation and deliberation as appropriate method to find an agreement. Rahwan on the other hand seems to remain rather open in this regard yet emphasises the need for quantifiable tools to measure human values [11], p. 9. Although we agree that this would help streamline negotiation processes, we remain somewhat sceptical about quantifiable parameters as prerequisite to be considered. As the pluralistic approach is explicitly acknowledging the equal authority for any value and interest that stakeholders may hold, thus including economic as well as broad social or ecologic values, we think that the quantifiability condition risks excluding and disadvantaging some values and thus unjustifiably reduces the deliberative arena. Perhaps the most important difference, however, lies in the different levels at which the approaches are ultimately intended to have practical effect. Whereas Rahwan defines the SITL framework in contrast to the HITL paradigm that operates on a micro-level of individual technical systems; our approach focuses on the organisational level of firms based on contractualist business ethics. The SITL thus "looks more like public feedback on regulations and legislations than feedback on frequent micro-level decisions" [11], p. 12. Deliberative order ethics by contrast seeks to enable ethical behaviour of firms at an intermediate level, below the level of legislation and above individual measures of corporate governance. As pointed out above, the rules which deliberative order ethics aim to establish for ethical behaviour not to constitute a competitive disadvantage can be understood as soft law. In Rahwan's framework, negotiation and public engagement would instead take place at the regulative level of hard law. While both our approaches thus acknowledge the crucial political dimension of AI ethics, deliberative order ethics supplements the picture by adding a political and discursive level between legislation and the micro level of technology. It is in this sense that we call our approach *community-in-the-loop* (CITL) adapting Rahwan's reasoning and based on our definition of economic community above. We argue that in general society expresses values and preferences through the political system in place, assuming they are democratic societies. We agree with Rahwan in that more adequate hard legislation in the context of AI is needed and that more participation and deliberation on the level of democratic political systems would be of great value, yet for ethical AI based on participation to be most effective, we believe an approach of *community-in-the-loop* would be appropriate.

In the context of algorithmic accountability, Binns [12] proposed a concept based on the democratic ideal of public reason. Binns explores the question of algorithmic accountability, i.e. the right of individuals to know what principles and considerations lie behind an algorithm-based decision to be able to understand and, if necessary, contest it [12], p. 547. The challenge of algorithmic accountability, according to Binns, is, therefore, to make the implicit values of technical systems understandable and justifiable in such a way that they might persist in a pluralistic environment. Thus, the task is not only to identify the epistemic and normative assumptions inscribed in the development and design of technical systems, but to provide explanations and justifications that are acceptable to all (potentially) affected individuals. However, in a society in which individuals legitimately hold differing values, a divergence not only in epistemic but particularly in normative standards seems likely. Thus, for algorithmic accountability to promote the legitimacy of algorithmic-based decisions, Binns suggests that an account of public reason be taken as a basis. Put simply, this states that despite existing differences, there must be universal rules and principles "provided they are suitably public and shared by all reasonable people in the society" [12], p. 549. The dilemma of justifying epistemic and particularly normative assumptions in the context of pluralistic societies may thus be overcome by referring to universal principles that establish the shared standard as a frame of reference.

Although Binns' focus is on a more specific problem, it bears similarities to deliberative order ethics not only in that he draws from ethics and political philosophy, but also in that he uses a similar problem description as starting point. Hence, both approaches start from the question of how conflicts between different values and interests triggered by the use of AI-based systems are to be solved, given the assumption of reasonable value pluralism. The solutions, however, as provided by Binns' approach to algorithmic accountability in terms of public reason on the one hand and our proposal for an approach of deliberative order ethics on the other hand, show two main differences. First, our proposal starts at an intermediate level of business ethics, while Binns like Rahwan [11] starts at the wider political level

of democratic societies. Second, Binns' account of public reason provides for value conflicts to be resolved by reference to universal principles, while order ethics attempts to refrain from assuming any universal normative principles. At this point, Binns emphasises that in advocating a public reason-flavoured form of algorithmic accountability, no particular form of public reason should be presupposed. Instead "the precise content of these common principles is expected to emerge from a process of reflective equilibrium between equal citizens" [12], p. 550. To some extent, the process of reflective equilibrium on a societal level to identify universal ethical principles is similar to the deliberative process of order ethics to describe community specific ethical rules. Thus, while Binns seeks to resolve pluralistic conflicts of values in the context of AI by establishing principles and rules at a societal level, we believe that from a business ethics perspective, such rules are most effective and practicable at the community level in terms of specific fields of application. Would it not be plausible to assume that for principles of algorithmic accountability too, different rules might be useful depending on the area of application, but still general enough as it is shared by a group of stakeholders, an economic community? Would it not be possible, for example, to have different shared epistemic and normative standards for algorithmic accountability, depending on whether an online retailer refuses to provide me with a specific offer or whether my application for a vocational training position has been rejected? And still others when I receive my tax return? Despite the differences in the proposed responses, we find in Binns' proposal a corroboration of our concept of deliberative order ethics as it similarly highlights the issue of value conflicts underpinning AI ethics and seeks a solution based on rules and standards developed by public-deliberative dialogue.

Moreover, Wong [13] introduced a deliberative approach to the question of algorithmic fairness based on the accountability for reasonableness framework (AFR). Similar to the criticism of a too limited technical focus raised in Sect. 2, he argues that algorithmic fairness is mainly conceived as a technical challenge [13], p. 227. However, as it turns out, the concept of fairness is in itself controversial and different definitions exist, each with their own implications. Above all, however, this shows that first, it is mathematically impossible for an algorithm to fulfil different fairness measures at the same time, i.e. fairness claims other than those implemented are necessarily violated. Second, trade-offs between fairness and other factors in the design of algorithms arise, e.g. between fairness and performance or accuracy, or between fairness and safety [13], p. 229. As a result, algorithmic fairness is ultimately a question of conflicting values and interests of involved stakeholders and, in this sense, an inherent political task. In response to this challenge, Wong proposes to rely on the accountability for reasonableness framework (AFR) based on Daniels and Sabin [131]. Recognising the pluralistic nature of liberal democracies, the AFR presents a procedural framework to establish a "'process or procedure that most can accept as fair to those who are affected by such decisions. That fair process then determines for us what counts as fair outcome'" [131, quoted from [13]], p. 233. To this end, AFR formulates four conditions for decision-making processes to be considered fair and legitimate: a publicity condition, a relevance condition, a revision and appeals condition, and a regulative condition. What Wong thus proposes is to use a procedural account of public deliberation to engage in a "genuine exchange of reasons" and facilitate social learning to find common ground on appropriate fairness measures which are acceptable to all. In this way, Wong's approach is similar to Binns' [12] proposal in that both emphasise the political dimension of conflicting values and develop a framework for societal responses. Yet while Binns deduces his concept especially against the backdrop of reasonable pluralism, Wong argues on the basis of the internal features of algorithmic fairness [13], p. 239. Both Binns and Wong thus explore the question of how a shared normative basis for an ethical approach to AI can be found in the face of differing and conflicting values and interests, once in relation to algorithmic accountability and once in the context of algorithmic fairness. Wong's proposition of an AFR differs from our approach to deliberative order ethics mainly by its explicitly practical claim [13], p. 241, which offers concrete criteria for process design to determine which values or which conceptualisation of individual values should be adopted in the design and use of AI systems. In this way, Wong's proposal can be seen as a supplement on a practical level of implementation. Regarding the implementation of deliberative or ethics, in particular the revision and appeals condition and criteria for the development of rules require further elaboration. What remains open, however, is how AFR's criteria should be applied in practice. Insofar as the proposal addresses specific AI systems on the one hand, but on the other hand is based on the overarching level of the political system as a whole (deliberative democracy) [13], p. 238, the connection between the two levels does not seem trivial. It is here that the added value of our account of deliberative or ethics comes into play as it attempts to provide a more precise description of who (firms) wants to achieve what (establishment of rules) and how (deliberative stakeholder engagement) and at which level (economic communities). This kind of methodological middle step of normative political theory on the one hand, and precise criteria for process design on the other, not only facilitates the practical application but also allows normative theory to be differentiated and expanded over different levels.

Finally, Whittlestone et al. [9] and Rosenbaum and Fichman [9] explored the wider societal dimensions of ethical

issues in the use of AI. Rosenbaum and Fichman, like Binns [12], focus on the question of algorithmic accountability and point out the complexities of technical and sociotechnical approaches. Summarising different ways forward they point to the societal and political dimensions of digital justice "by moving away from a focus on the algorithm itself" [130], p. 243. As mentioned above, Whittlestone et al. [9] argue that the normative dimension of AI ethics should be understood as tensions between different values and principles and thus agree with our proposal to take value conflicts as a starting point as well as with the ideas of Rahwan [11], Binns [12], Wong [13]. They echo the concern that fundamental normative conflicts may not be resolved with a technical or principled view [7, 8, 10]. Rather, they underline the difficulty to achieve acceptable results when dealing with value conflicts or tensions: "Making these trade-off judgements will be a complex political process. Weighing the costs and benefits of different solutions can be an important part of the process but alone is not enough, since it fails to recognise that values are vague and unquantifiable, and that numbers often hide complex value judgements. In addition, resolving trade-offs will require extensive public engagement, to give voice to a wide range of stakeholders and articulate their interests with rigour and respect" [9], p. 199. Moreover, trade-offs may not be unavoidable, e.g. if further research and development promises solutions that reduce or even avoid trade-offs [9]. Should we then wait for the technology to be applied in the future or should we use existing applications? Or does this perhaps point to precisely those instances where a technical, AI-based solution may not be the best alternative [83]? Whittlestone et al. [9], therefore, stress the need for stakeholder engagement and deliberation to evaluate costs and benefits and find agreement on acceptable trade-offs. Our account of deliberative order ethics picks this up and introduces a normative theory of contractualist business ethics to provide a both normatively firm and practicable approach to AI ethics.

Overall, the discussion shows that a broadening of AI ethics towards the fundamental normative conflicts between different values and interests on a societal level is needed [9, 11–13, 130]. We understand the proposals by Rahwan [11], Binns [12], Wong [13] and Whittlestone [9] as an affirmation that the starting point of value conflicts as adopted by deliberative order ethics is suitable for addressing the ethics of AI on a comprehensive societal level. We also agree with Rahwan [11], Wong [13], Whittlestone [9] and, to some extent, Binns [12] that in view of the conflicts and trade-offs to be negotiated, a participatory and deliberative approach is appropriate for negotiating a fair distribution of costs and benefits and acceptable outcomes in a mutual exchange between affected stakeholders. Based on contractualist business ethics, our approach further specifies that firms should involve not only stakeholders of a specific business model, but the whole economic community in the sense of all stakeholders involved and potentially affected by the use of AI systems in a specific application area, say recruiting, not least from civil society. Engaging in participatory deliberation, firms should ultimately establish rules that enable mutual advantages to be created for the entire economic community. Thus, while the basic theoretical approach is similar to other current contributions, our approach is distinct in that it introduces a business ethics concept based on a solid theoretical framework. We argue that a business ethics perspective is useful for linking normative-theoretical considerations to the relevant implementation context of AI innovation, to both conceptually enhance AI ethics and improve its practical application. In this respect, our proposal of deliberative order ethics provides a valuable addition to the existing contractualist and deliberative proposals in AI ethics.

Recently, Himmelreich [132] argued that AI ethics should turn to political philosophy to take greater account of the collective decisions that are evoked by the use of AI systems. Political philosophy would then be able to add three basic concerns to the conceptual toolkit of AI ethics: reasonable pluralism, individual agency, legitimate authority. As explained above, contractualist business ethics is fundamentally based on the question of preserving reasonable pluralism and the possibilities of legitimate authority. Along these lines, our theory-based proposal contributes to complement AI ethics with political philosophy. Insofar as we are developing a concept of business ethics on such a theoretical basis, we even go one step further and argue that AI ethics needs not only political philosophy but also normatively sound contributions of business ethics [102, 133, 134]. It is precisely from such an integrated approach that AI ethics can be pushed forward both in terms of normative concepts and in practice.

4.1 Problems solved? Addressing current shortcomings with deliberative order ethics

Recognising the beginnings of a third wave, we started our reasoning with a critical discussion of the current state of AI ethics. To this end, we summarised five shortcomings of a first and second waves of AI ethics. In this section, we will, therefore, briefly discuss whether or not and to what extent the concept of deliberative order ethics may successfully address current weaknesses.

4.1.1 AI ethics neglects the business context of developing and employing AI systems

The first shortcoming concerned the focus on the technical level of AI systems, which tends to neglect the integration in an entrepreneurial and wider societal context. In response to

this, deliberative order ethics offers a valuable contribution as it starts at the level of firms and their business practices which may create value conflicts. The deliberative approach involving the stakeholders of an economic community does not focus on the ethical design of the technical system alone, but rather on the question of how the use of AI in a specific societal context is acceptable for all involved, i.e. how it can be implemented to the benefit of all. The object of ethical scrutiny is thus not only the AI system on a micro-level but the AI system in the context of its commercialisation, and its impact on markets, the environment, individuals and society. In this way, primarily responsible actors are not individual developers but firms on an organisational level and their strategic behaviour in economic communities. By introducing a business ethics perspective, the business context of the use of AI systems is systematically taken into account. At the same time, the political dimension of AI ethics is reflected by focusing on value conflicts and deliberatively negotiating acceptable trade-offs with all stakeholders, not least from civil society. Drawing from political philosophy, deliberative order ethics thereby is able to "address fundamental normative and political tensions" [7], p. 501, prompted by AI systems.

4.1.2 AI ethics is biased toward a technological solutionism

Closely connected to the first issue of too narrow a focus on technical systems alone, the second problem refers to the problem of technological solutionism. By focusing on the ethical design of AI systems, the question of whether or not a particular AI system should be built and employed in the first place or whether there is perhaps another, possibly non-technical alternative that can better solve the problem at hand moves out of sight. Here, too, deliberative order ethics seems to offer an approximation to the problem by broadening the scope of AI ethics. Insofar as rules for a fair distribution of costs and benefits and acceptable trade-offs are negotiated in a deliberative procedure, what is at stake is a more comprehensive evaluation of AI-based business practices. This includes the question of whether the use of AI systems appears appropriate and reasonable in view of the identified costs and benefits. Although its goal is not to devise possible alternatives, the deliberative establishment of rules defines the framework (level playing field) within which the use of AI systems is socially acceptable. By formulating the conditions accordingly, minimum thresholds can be set and certain AI applications may thus be ruled out.

4.1.3 AI ethics succumbs to an individualist focus

Third, we pointed out that the focus on AI systems and their ethical design leads to ethical action primarily to be located at the level of individuals such as data scientists and developers. However, this fails to recognise the role of the organisational level, both internally in terms of governance mechanisms and corporate culture and externally in terms of its integration in markets and the wider societal contexts. Here the perspective from business ethics helps to bring the role of firms as organisations into focus. In doing so, deliberative order ethics concentrates less on internal aspects than on external interactions and the behaviour of businesses in society. Accordingly, responsibility for ethical behaviour is no longer attributed (only) to developers but to firms as social actors who need to establish general rules for the ethical use of AI systems on the basis of their own interests. In this sense, firms have the responsibility to initiate or actively participate in respective deliberative processes if they intend to establish a new business segment or expand an existing one using AI.

4.1.4 AI ethics is problematic in its implementation and lacks accountability and clear impact

The fourth weakness relates to the difficulty of making abstract principles manageable in practice, identifying effective approaches and ensuring their normative soundness (gap between a variety of tools and normative justification). Here, the approach of deliberative order ethics can only partly provide an adequate answer as it fell outside the scope of this article to spell out its practical application. Nevertheless, in response to the observed implementation problems, our approach advocates a systematic combination of normative theory with an applied perspective of business ethics. However, the concept needs to be fleshed out for its implementation and its effectiveness needs to be critically evaluated.

4.1.5 AI ethics does not clarify its link to legal regulation

Finally, we pointed out the ambiguous relationship between AI ethics and legal regulation, which has been criticised by a number of authors. Here, our proposal provides a thorough clarification. Based on a contractualist theory of normative business ethics [103] and the concept of incomplete contracts [105], the complementary character of ethics can be specified [cf. 82]. Deliberative order ethics highlights the relevance and need of legislation and hard law regulating the development and use of AI set clear and binding rules enabling fair competition. It is only in addition to these regulations that order ethics seeks to address inevitable gaps for ethical rules. Thereby, ethics may also function as participatory creation and testing of ethical rules for as long as the legislative process is still under way [cf. 61], and rules might even become legislation at some point. According to the complementary nature of the relationship, deliberate order ethics should thus under no circumstances provide a basis for avoiding or delaying legislation.

To conclude, complementing AI ethics with deliberative order ethics thus offers several benefits. By integrating ethical considerations on the business practices surrounding the use of AI systems in society, it forces firms to identify and analyse the diverse impacts the employment of AI may have, to discuss costs and benefits from a comprehensive perspective with all stakeholders from the community and explicitly formulate rules for acceptable trade-offs that allow added value for all. Thereby value conflicts caused by critical business practices become the subject of an open dialogue. As a result, both new and already existing problematic business practices can be revealed and put up for discussion [cf. 12]. Furthermore, the deliberative approach not only helps uncover relevant values and interests in society, it also initiates the weighing up and explicit balance of different societal values, from economic to social and ecological values. Establishing societal standards for a desirable use of AI, deliberative order ethics offers an avenue designed to make the use of AI beneficial for everyone in society. Beyond avoiding adversarial effects, this makes pluralistic value creation the ultimate ambition of AI ethics.

5 Conclusion

The increasing use of AI systems not only presents great opportunities for many important areas of society such as medicine or climate protection, it also raises profound ethical questions and challenges fundamental societal values. Recognising these impacts, the field of AI ethics emerged developing both theoretical guidelines as well as practical tools addressing issues such as unfair discrimination or algorithmic accountability.

In this article, we introduced a procedural account of a deliberative order ethics to complement AI ethics. To this end, we first presented the current state of AI ethics which results in our review in two first waves of AI ethics. At the threshold of the beginning of a third wave we consolidate different concerns by arguing that in its current form AI ethics is facing at least five crucial shortcomings: AI ethics tends to neglect the business context of developing and employing AI systems, it is biased toward a technological solutionism, succumbs to an individualist focus, is problematic in its implementation and lacks accountability and clear impact, and does not clarify its link to legal regulation. Building on this critique, we first introduced the contractualist concept of normative business ethics called order ethics. Contrasted with ISCT as the most proliferated theory of contractualist business ethics, we argue that deliberative order ethics provides an adequate approach to deal with the complex value conflicts firms may trigger through AI in pluralistic societies. Order ethics holds that these conflicts should be resolved by adequate rules so that ethical behaviour does not lead to a competitive disadvantage to which stakeholders agree based on mutual benefits. Second, we proposed a procedural expansion arguing that it is through participation and deliberation that stakeholders of an economic community may adequately discuss costs and benefits and agree on rules for acceptable trade-offs when using AI systems in their respective field. Thereby, deliberative order ethics ultimately seeks to make the use of AI systems a matter of pluralistic value creation. The role of ethics in AI thus becomes, among other things, to ensure that AI creates diverse societal and ecologic value in combination with financial business value.

By complementing AI ethics with an approach of business ethics, we aim to integrate the level of business practices into the considerations of AI ethics and highlight the organisational role of firms for achieving ethical AI. Since AI systems and most other emerging technologies are at least commercialised and brought to society by firms through new business models or enhanced products and services, we believe that one cannot achieve truly ethical AI without addressing key issues of business ethics. From a favourable point of view, building "ethical" AI systems that are then part of questionable and dodgy business practices and markets does not seem to cover the whole picture. At worst, it is part of a deceitful strategy and irresponsible. Some perspectives in AI ethics seem to implicitly assume an opposing relation of business and ethics which ultimately leads to the conclusion that truly ethical AI can only be possible beyond businesses and markets in their current logic and structure. This, however, means admitting that, at least in the short and medium term, ethical AI business practices cannot gain wider application. Although we are somewhat sympathetic to such a sincerely critical and more idealistic view, our ambition is to strive for realistic change to achieve ethical AI at all, given the current circumstances. At the critical point at which AI innovation currently stands, it is thus particularly important to reflect the normative foundations of both economics and businesses and to make AI ethics an endeavour of business ethics too. To this end, we have made a first contribution with this article.

Combining two previously separated fields of research, this article shows some important limitations. First, as we focused on demonstrating how a business ethics perspective may provide a valuable complementation to AI ethics addressing some of its current shortcomings, a detailed description of the concept of deliberative order ethics and what a concrete application might involve exceeded the scope of this article. Among other aspects, this would entail more conceptual detail with regard to criteria for deliberation processes, rules and consent as well as an exemplary description of its implementation in practice. Second, our analysis of the first two waves of AI ethics only covers major trends and was thus unable to do full justice to the diverse and dynamic field that is emerging today. An adequate

review for the purpose of a systematic critique of the field would provide the subject of an entire article. Alone the focus of this article was a different one.

These limitations, however, also provide the basis for further research. Future research agendas should seek to both systematise and consolidate the very diverse and dynamic field of AI ethics as well as foster the development of solid practical approaches. More specifically, further research needs to develop practical approaches and tools based on well-founded normative claims and evaluate their effectiveness empirically.

Finally, this article contributed to a diversification of causation narratives about (un)ethical AI [8]: without taking into account the role of business practices and markets, AI ethics risks never reaching its goals. Therefore, political philosophy and business ethics are urgently needed complements to ensure that AI ethics remains a theoretically sound and practically effective effort.

Acknowledgements We would like to thank two anonymous reviewers for their positive and constructive feedback.

Author contributions Not applicable.

Funding Open Access funding enabled and organized by Projekt DEAL. Not applicable.

Availability of data and materials Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

References

1. Perrault R, Shoham Y, Brynjolfsson E, Clark J, Etchemendy J, Grosz Harvard B, Lyons T, Manyika J, Carlos Niebles J, Mishra S (2019) The AI index 2019 annual report. Stanford, CA
2. Benaich N, Benaich N (2019) State of AI report. London, United Kingdom
3. Floridi, L., Cowls, J., Beltrametti, M., et al.: AI4People—an ethical framework for a good AI Society: opportunities, risks, principles, and recommendations. Minds Mach **28**, 689–707 (2018)
4. Dignum, V.: Ethics in artificial intelligence: introduction to the special issue. Ethics Inf Technol **20**, 1–3 (2018)
5. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: mapping the debate. Big Data Soc **3**, 205395171667967 (2016)
6. Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., Floridi, L.: The ethics of algorithms: key problems and solutions. SSRN Electron J (2020). https://doi.org/10.2139/ssrn.3662302
7. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat Mach Intell **1**, 501–507 (2019)
8. Greene D, Hoffmann AL, Stark L (2019) Better, nicer, clearer, fairer: a critical assessment of the movement for ethical artificial intelligence and machine learning. In: Proc. 52nd Hawaii int. conf. syst. sci., pp 2122–2131
9. Whittlestone J, Alexandrova A, Nyrup R, Cave S (2019) The role and limits of principles in AI ethics: towards a focus on tensions. AIES 2019—proc 2019 AAAI/ACM conf AI, ethics, soc, pp 195–200
10. Green, B.: Data science as political action: grounding data science in a politics of justice. SSRN Electron J (2020). https://doi.org/10.2139/ssrn.3658431
11. Rahwan, I.: Society-in-the-loop: programming the algorithmic social contract. Ethics Inf Technol **20**, 5–14 (2018)
12. Binns, R.: Algorithmic accountability and public reason. Philos Technol **31**, 543–556 (2018)
13. Wong, P.: Democratizing algorithmic fairness. Philos Technol **33**, 225–244 (2020)
14. Mccarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E.: A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. AI Mag **27**, 12 (2006)
15. Russell, S.J., Norvig, P.: Artificial intelligence: a modern approach, 3rd edn. Pearson Education, Harlow (2016)
16. Larson J, Mattu S, Kirchner L, Angwin J (2016) How we analyzed the COMPAS recidivism algorithm. In: ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
17. The Montreal declaration (2017)
18. Future of Life Institute (2018) Asilomar AI principles
19. High-Level Expert Group on Artificial Intelligence (HLEG) (2019) Ethics guidelines for trustworthy AI. Brussels
20. AI4People (2018) The AI4People's ethical framework for a good AI Society: opportunities, risks, principles, and recommendations. Brussels
21. OECD (2019) Recommendation of the council on artificial intelligence
22. UK House of Lords SC on AI (2017) AI in the UK: ready, willing and able? London
23. Datenethikkomission (2019) Gutachten der Datenethikkommission. Berlin
24. Deutsche Telekom (2018) Digital ethics: guidelines on AI. https://www.telekom.com/resource/blob/544508/ca70d6697d35ba60fbcb29aeef4529e8/dl-181008-digitale-ethik-data.pdf
25. Microsoft (2018) Responsible AI. https://www.microsoft.com/en-us/ai/responsible-ai
26. Google (2018) Artificial intelligence at Google: our principles
27. IEEE (2017) Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems, version 2
28. Partnership on AI (2016) Tenets. https://www.partnershiponai.org/tenets/
29. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat Mach Intell **1**, 389–399 (2019)

30. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Minds Mach **30**, 99–120 (2020)
31. Floridi, L., Cowls, J.: A Unified framework of five principles for AI in Society. Harv Data Sci Rev **1**, 1–13 (2019)
32. Floridi, L., Cowls, J., King, T.C., Taddeo, M.: How to design AI for social good: seven essential factors. Sci Eng Ethics **26**, 1771–1796 (2020)
33. Tasioulas, J.: First steps towards an ethics of robots and artificial intelligence. J Pract Ethics **7**, 49–83 (2019)
34. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L (2019) Explaining explanations: an overview of interpretability of machine learning. Proc—2018 IEEE 5th int conf data sci adv anal DSAA 2018, pp 80–89
35. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artif Intell **267**, 1–38 (2019)
36. Madumal P, Miller T, Vetere F, Sonenberg L (2018) Towards a grounded dialog model for explainable artificial intelligence. arXiv:1806.08055
37. Arrieta AB, Díaz-Rodríguez N, Del Ser J et al (2019) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. arXiv:1910.10045
38. Páez, A.: The pragmatic turn in explainable artificial intelligence (XAI). Minds Mach (2019). https://doi.org/10.1007/s11023-019-09502-w
39. Gunning D (2019) DARPA's explainable artificial intelligence (XAI) program. In: Proc. 24th int. conf. intell. user interfaces—IUI '19. ACM Press, New York, New York, USA, pp ii–ii
40. Lee MSA, Singh J (2021) The landscape and gaps in open source fairness toolkits. In: CHI conference on human factors in computing systems (CHI '21), 8–13 May 2021, Yokohama, Japan. ACM, Yokohama. https://doi.org/10.1145/3411764.3445261
41. Hellman, D.: Measuring algorithmic fairness. Va Law Rev **106**, 811–866 (2020)
42. Holstein K, Wortman Vaughan J, Daumé H, Dudik M, Wallach H (2019) Improving fairness in machine learning systems. In: Proc. 2019 CHI conf. hum. factors comput. syst.—CHI '19. ACM Press, New York, New York, USA, pp 1–16
43. Barocas S, Hardt M, Narayanan A (2019) Fairness and machine learning. http://fairmlbook.org
44. Kleinberg J (2018) Inherent trade-offs in algorithmic fairness. In: Abstr. 2018 ACM int. conf. meas. model. comput. syst. ACM, New York, NY, USA, pp 40–40
45. Chouldechova A, Roth A (2018) A snapshot of the frontiers of fairness in machine learning. Commun ACM 63(5):82–89. https://doi.org/10.1145/3376898
46. Pessach D, Shmueli E (2020) Algorithmic fairness. arXiv:2001.09784
47. Corbett-Davies S, Goel S (2018) The measure and mismeasure of fairness: a critical review of fair machine learning. arXiv:1808.00023
48. Boddington P (2017) Towards a code of ethics for artificial intelligence. Springer, Cham. https://doi.org/10.1007/978-3-319-60648-4
49. Bertelsmann Stiftung (2018) Ethik für Algorithmiker: Was wir von erfolgreichen Professionsethiken lernen können. https://doi.org/10.11586/2018033
50. Bonnemains, V., Saurel, C., Tessier, C.: Embedded ethics: some technical and ethical challenges. Ethics Inf Technol **20**, 41–58 (2018)
51. McLennan, S., Fiske, A., Celi, L.A., Müller, R., Harder, J., Ritt, K., Haddadin, S., Buyx, A.: An embedded ethics approach for AI development. Nat Mach Intell **2**, 488–490 (2020)
52. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Jan DB (2020) Datasheets for datasets. arXiv:1803.09010
53. Madaio MA, Stark L, Wortman Vaughan J, Wallach H (2020) Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In: Proc. 2020 CHI conf. hum. factors comput. syst. ACM, New York, NY, USA, pp 1–14
54. Rakova B, Yang J, Cramer H, Chowdhury R (2020) Where responsible AI meets reality: practitioner perspectives on enablers for shifting organizational practices. arXiv:2006.12358
55. Schiff D, Rakova B, Ayesh A, Fanti A, Lennon M (2020) Principles to practices for responsible AI: closing the gap. arXiv:2006.04707
56. Raji ID, Smart A, White RN, Mitchell M, Gebru T, Hutchinson B, Smith-Loud J, Theron D, Barnes P (2020) Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. https://doi.org/10.1145/3351095.3372873
57. AI Ethics Impact Group (2020) From principles to practice: an interdisciplinary framework to operationalise AI ethics. Bertelsmann Stift. https://doi.org/10.11586/2020013
58. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv J Law Technol **31**, 841 (2018)
59. Calo, R.: Artificial intelligence policy: a primer and roadmap. UCD L Rev **51**, 399 (2018)
60. Wachter S, Mittelstadt B (2019) A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI. Columbia Bus Law Rev 2019(2):494–620. https://doi.org/10.7916/cblr.v2019i2.3424
61. Larsson, S.: On the governance of artificial intelligence through ethics guidelines. Asian J Law Soc **00**, 1–15 (2020)
62. Coeckelbergh M (2019) Artificial intelligence: some ethical issues and regulatory challenges. Technol Regul. https://doi.org/10.26116/techreg.2019.003
63. Morley, J., Floridi, L., Kinsey, L., Elhalal, A.: From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci Eng Ethics **26**, 2141–2168 (2020)
64. Veale, M.: A critical take on the policy recommendations of the EU high-level expert group on artificial intelligence. Eur J Risk Regul (2020). https://doi.org/10.1017/err.2019.65
65. Metcalf, J., Moss, E., Boyd, D.: Owning ethics: corporate logics, silicon valley, and the institutionalization of ethics. Soc Res An Int Quart **86**, 449–476 (2020)
66. Rességuier, A., Rodrigues, R.: AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data Soc **7**, 1–5 (2020)
67. Benkler, Y.: Don't let industry write the rules for AI. Nature **569**, 161–161 (2019)
68. Binns R (2020) On the apparent conflict between individual and group fairness. In: Proc. 2020 conf. fairness, accountability, transpar. ACM, New York, NY, USA, pp 514–524
69. Martin, K., Shilton, K., Smith, J.: Business and the ethical implications of technology: introduction to the symposium. J Bus Ethics **160**, 307–317 (2019)
70. Whittaker M, Crawford K, Dobbe R et al (2018) AI now report 2018. AI Now Institute, New York
71. Buranyi S (2018) "Dehumanising, impenetrable, frustrating": the grim reality of job hunting in the age of AI. In: Guardian. https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai. Accessed 30 Oct 2020
72. Haucap, J.: Markt, Macht und Wettbewerb: Was steuert die Datenökonomie. Nicolai Publishing, Berlin (2018)
73. Kulynych B, Overdorf R, Troncoso C, Gürses S (2018) POTs: protective Optimization Technologies. FAT* 2020—Proc 2020 conf fairness, accountability, transpar, pp 177–188
74. Gürses, S., Overdorf, R., Balsa, E.: Stirring the pots: protective optimization technologies. In: Bayamlioglu, E., Baraliuc, I.,

Janssens, L., Hildebrandt, M. (eds.) Being profiled, pp. 24–29. Amsterdam University Press, Amsterdam (2019)
75. Bertsimas, D., Delarue, A., Martin, S.: Optimizing schools' start time and bus routes. Proc Natl Acad Sci USA **116**, 5943–5948 (2019)
76. Scharfenberg D (2018) Computers can solve your problem. You may not like the answer. What happened when Boston Public Schools tried for equity with an algorithm. In: Boston Globe. https://apps.bostonglobe.com/ideas/graphics/2018/09/equity-machine/. Accessed 30 Oct 2020
77. Ito J (2018) What the Boston School Bus schedule can teach us about AI an MIT team built an algorithm to optimize bell times and bus routes. The furor around the plan offers lessons in how we talk to people when we talk to them about artificial intelligence. In: Wired. https://www.wired.com/story/joi-ito-ai-and-bus-routes/. Accessed 30 Oct 2020
78. Crockford K, Ito J (2017) Don't blame the algorithm for doing what Boston school officials asked. In: Boston Globe. https://www3.bostonglobe.com/opinion/2017/12/22/don-blame-algorithm-for-doing-what-boston-school-officials-asked/lAsWv1Rfwqmq6Jfm5ypLmJ/story.html?arc404=true. Accessed 30 Oct 2020
79. Dignum V, Baldoni M, Baroglio C, et al (2018) Ethics by design. In: Proc. 2018 AAAI/ACM conf. AI, ethics, soc. ACM, New York, NY, USA, pp 60–66
80. d'Aquin M, Troullinou P, O'Connor NE, Cullen A, Faller G, Holden L (2018) Towards an "Ethics by Design" methodology for AI research projects. In: Proc. 2018 AAAI/ACM conf. AI, ethics, soc.—AIES '18. ACM Press, New York, New York, USA, pp 54–59
81. Aizenberg, E., van den Hoven, J.: Designing for human rights in AI. Big Data Soc **7**, 1–14 (2020)
82. Floridi, L.: Translating principles into practices of digital ethics: five risks of being unethical. Philos Technol **32**, 185–193 (2019)
83. Green, B.: The Smart Enough City: putting technology in its place to reclaim our urban future. MIT Press, Cambridge (2019)
84. Zeng, D., Chen, H., Lusch, R., Li, S.-H.: Social media analytics and intelligence. IEEE Intell Syst **25**, 13–16 (2010)
85. Yang, K.C., Varol, O., Davis, C.A., Ferrara, E., Flammini, A., Menczer, F.: Arming the public with artificial intelligence to counter social bots. Hum Behav Emerg Technol **1**, 48–61 (2019)
86. Yeung, K.: 'Hypernudge': big data as a mode of regulation by design. Inf Commun Soc **20**, 118–136 (2017)
87. Goldsmith J, Burton E (2017) Why teaching ethics to AI practitioners is important. AAAI-17 Work AI, Ethics, Soc, pp 110–114
88. Ryan, M., Stahl, B.C.: Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. J Inf Commun Ethics Soc (2020). https://doi.org/10.1108/JICES-12-2019-0138
89. Wagner, B.: Ethics as an escape from regulation: from "ethics-washing" to ethics-shopping? In: Hildebrandt, M. (ed.) Being profiled. Cogitas Ergo Sum, pp. 84–90. Amsterdam University Press, Amsterdam (2018)
90. Floridi, L.: Soft ethics and the governance of the digital. Philos Technol **31**, 1–13 (2018)
91. Floridi, L.: Information ethics: on the philosophical foundation of computer ethics. Ethics Inf Technol **1**, 37–56 (1999)
92. Moriarty J (2017) Business ethics. Stanford Encycl. Philos.
93. Khalil, O.E.M.: Artificial decision-making and artificial ethics: a management concern. J Bus Ethics **12**, 313–321 (1993)
94. Kaplan, A., Haenlein, M.: Rulers of the world, unite! The challenges and opportunities of artificial intelligence. Bus Horiz **63**, 37–50 (2020)
95. Kaplan, A., Haenlein, M.: Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz **62**, 15–25 (2019)
96. Martin, K.: Ethical implications and accountability of algorithms. J Bus Ethics **160**, 835–850 (2019)
97. Bartneck C, Lütge C, Wagner A, Welsh S (2021) An introduction to ethics in robotics and AI. Springer, Cham. https://doi.org/10.1007/978-3-030-51110-4
98. Lütge, C.: There is not enough business ethics in the ethics of digitization. In: Ciulla, J.B., Scharding, T.K. (eds.) Ethical bus. Leadersh. Troubl. Times, pp. 280–295. Edward Elgar Publishing, Cheltenham (2019)
99. Buchanan, J.M.: The limits of liberty. Between anarchy and leviathan. Chicago University Press, Chicago (1975)
100. Brennan, G., Buchanan, J.M.: The reason of rules: constitutional political economy. Cambridge University Press, Cambridge (1985)
101. Rawls, J.: A theory of justice. Harvard University Press, Cambridge (1971)
102. Luetge, C., Armbrüster, T., Müller, J.: Order ethics: bridging the gap between contractarianism and business ethics. J Bus Ethics **136**, 687–697 (2016)
103. Luetge, C.: The idea of a contractarian business ethics. In: Handb. philos. found. bus. ethics, pp. 647–658. Springer Netherlands, Dordrecht (2013)
104. Luetge, C.: Contractarian foundations of order ethics. In: Order ethics an ethical framew. soc. mark. econ., pp. 3–17. Springer International Publishing, Cham (2016)
105. Hart, O.: Incomplete contracts and control. Am Econ Rev **107**, 1731–1752 (2017)
106. Luetge, C.: Economic ethics, business ethics and the idea of mutual advantages. Bus Ethics A Eur Rev **14**, 108–118 (2005)
107. Heugens, P.P.M.A.R., van Oosterhout, J., Kaptein, M.: Foundations and applications for contractualist business ethics. J Bus Ethics **68**, 211–228 (2006)
108. Wempe, B.: On the use of the social contract model in business ethics. Bus Ethics A Eur Rev **13**, 332–341 (2004)
109. Donaldson, T.J., Dunfee, T.W.: Ties that bind: a social contracts approach to business ethics. Harvard University Press, Boston (1999)
110. Donaldson, T., Dunfee, T.W.: Toward a unified conception of business ethics: integrative social contracts theory. Acad Manag Rev **19**, 252–284 (1994)
111. Donaldson, T., Dunfee, T.W.: Integrative social contracts theory: a communitarian conception of economic ethic. Econ Philos **11**, 85–112 (1995)
112. Dunfee, T.W., Donaldson, T.J.: Integrative social contracts theory. In: Wiley encycl. Manag., pp. 1–5. Wiley, Chichester (2015)
113. Dunfee, T.W.: Business ethics and extant social contracts. Bus Ethics Q **1**, 23–51 (1991)
114. Wempe, B.: Four design criteria for any future contractarian theory of business ethics. J Bus Ethics **81**, 697–714 (2008)
115. Soule, E.: Managerial moral strategies—in search of a few good principles. Acad Manag Rev **27**, 114–124 (2002)
116. Phillips, R.A., Johnson-Cramer, M.E.: Ties that unwind: dynamism in integrative social contracts theory. J Bus Ethics **68**, 283–302 (2006)
117. Boatright, J.R.: Contract theory and business ethics: a review of ties that bind. Bus Soc Rev **105**, 452–466 (2000)
118. Calton, J.M.: Social contracting in a pluralist process of moral sense making: a dialogic twist on the ISCT. J Bus Ethics **68**, 329–346 (2006)
119. Reisel, W.D., Sama, L.M.: The distribution of life-saving pharmaceuticals: viewing the conflict between social efficiency and economic efficiency through a social contract lens. Bus Soc Rev **108**, 365–387 (2003)

120. Van Buren, H.J.: If fairness is the problem, is consent the solution? Integrating ISCT and stakeholder theory. Bus Ethics Q **11**, 481–499 (2001)
121. Dunfee, T.W.: A critical perspective of integrative social contracts theory: recurring criticisms and next generation research topics. J Bus Ethics **68**, 303–328 (2006)
122. Burg, R.: Deliberative business ethics. J Bus Ethics **88**, 665–683 (2009)
123. Husted, B.W.: A critique of the empirical methods of integrative social contracts theory. J Bus Ethics **20**, 227–235 (1999)
124. Wempe, B.: Extant social contracts and the question of business ethics. J Bus Ethics **88**, 741–750 (2009)
125. Corbett-Davies S, Pierson E, Feller A, Goel S, Huq A (2017) Algorithmic decision making and the cost of fairness. In: Proc. 23rd ACM SIGKDD int. conf. knowl. discov. data min. ACM, New York, NY, USA, pp 797–806
126. Kleinberg, J., Mullainathan, S., Raghavan, M.: Inherent trade-offs in the fair determination of risk scores. Proc Innov Theor Comput Sci **67**, 1–23 (2017)
127. Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C.R.: Discrimination in the age of algorithms. J Leg Anal **10**, 1–62 (2018)
128. Ast, F.: The deliberative test, a new procedural method for ethical decision making in integrative social contracts theory. J Bus Ethics **155**, 207–221 (2019)
129. Schormair, M.J.L., Gilbert, D.U.: Creating value by sharing values: managing stakeholder value conflict in the face of pluralism through discursive justification. Bus Ethics Q **31**, 1–36 (2020)
130. Rosenbaum, H., Fichman, P.: Algorithmic accountability and digital justice: a critical assessment of technical and sociotechnical approaches. Proc Assoc Inf Sci Technol **56**, 237–244 (2019)
131. Daniels N, Sabin JE (2002) Setting limits fairly: can we learn to share medical resources? https://doi.org/10.1093/acprof:oso/9780195149364.001.0001
132. Himmelreich, J.: Ethics of technology needs more political philosophy. Commun ACM **63**, 33–35 (2019)
133. Heath, J., Moriarty, J., Norman, W.: Business ethics and (or as) political philosophy. Bus Ethics Q **20**, 427–452 (2010)
134. Moriarty, J.: On the relevance of political philosophy to business ethics. Bus Ethics Q **15**, 455–473 (2005)
135. Loi, M., Ferrario, A., Viganò, E.: Transparency as design publicity: explaining and justifying inscrutable algorithms. Ethics Inf Technol (2020). https://doi.org/10.1007/s10676-020-09564-w
136. Leben, D.: A Rawlsian algorithm for autonomous vehicles. Ethics Inf Technol **19**, 107–115 (2017)
137. Buhmann, A., Paßmann, J., Fieseler, C.: Managing algorithmic accountability: balancing reputational concerns, engagement strategies, and the potential of rational discourse. J Bus Ethics **163**, 265–280 (2020)