



The future of AI in our hands? To what extent are we as individuals morally responsible for guiding the development of AI in a desirable direction?

Erik Persson¹ · Maria Hedlund²

Received: 12 August 2021 / Accepted: 28 November 2021 / Published online: 20 December 2021
© The Author(s) 2021

Abstract

Artificial intelligence (AI) is becoming increasingly influential in most people's lives. This raises many philosophical questions. One is what responsibility we have as individuals to guide the development of AI in a desirable direction. More specifically, how should this responsibility be distributed among individuals and between individuals and other actors? We investigate this question from the perspectives of five principles of distribution that dominate the discussion about responsibility in connection with climate change: effectiveness, equality, desert, need, and ability. Since much is already written about these distributions in that context, we believe much can be gained if we can make use of this discussion also in connection with AI. Our most important findings are: (1) Different principles give different answers depending on how they are interpreted but, in many cases, different interpretations and different principles agree and even strengthen each other. If for instance 'equality-based distribution' is interpreted in a consequentialist sense, effectiveness, and through it, ability, will play important roles in the actual distributions, but so will an equal distribution as such, since we foresee that an increased responsibility of underrepresented groups will make the risks and benefits of AI more equally distributed. The corresponding reasoning is true for need-based distribution. (2) If we acknowledge that someone has a certain responsibility, we also have to acknowledge a corresponding degree of influence for that someone over the matter in question. (3) Independently of which distribution principle we prefer, ability cannot be dismissed. Ability is not fixed, however and if one of the other distributions is morally required, we are also morally required to increase the ability of those less able to take on the required responsibility.

Keywords Distribution of forward-looking responsibility · Effectiveness-based distribution · Equality-based distribution · Desert-based distribution · Need-based distribution · Ability-based distribution

1 Introduction

Artificial intelligence (AI) is becoming increasingly influential in most people's lives. AI is already at work in more applications than many people realise, and the influence of AI is fast increasing. AI is in many ways making our lives more comfortable and has a large potential to improve our lives even further (e.g., [1–5]). It also carries with it risks, however. We are already seeing the effects of so-called algorithmic bias, privacy issues, etc. (e.g., [1–4, 6–8]).

Some authors see even bigger threats, even existential risks to humanity in the further future (e.g., [3, 9–12], see also [13–18]). Others, in turn, think we should not focus too much on the risks, especially not the very long-term risks [18–22].

We believe that speculation about whether AI will be beneficial or dangerous, and to what extent, is not a particularly fruitful endeavour, in particular since the answers to these questions largely depend on us and the decisions we make today, as well as in the future (see also [3, 10]). Rather than speculating about which direction the development of AI will take, we, therefore, think it is more fruitful to think about what we can do today to lead the development of AI in the direction we want. We also believe that the choices we make today, at the infancy of AI development, will be important for determining what direction the development of AI will take in the long run. It is therefore important that

✉ Erik Persson
erik.persson@fil.lu.se

¹ Department of Philosophy, Lund University, Lund, Sweden

² Department of Political Science, Lund University, Lund, Sweden

we make well-informed decisions from the start. It is also important that different parts of society are represented and get a chance to influence the development of AI (e.g., [23, 24]). This raises, among other things, the question about to what extent it is reasonable to claim that individual citizens, consumers, etc. have a moral responsibility to consider how their choices will affect the long-term development of AI, and to make choices that help guiding the development of AI in a direction that is beneficial, safe and democratic (hereafter: ‘guiding the development of AI in a beneficial direction’).

Do we as consumers have a moral responsibility to avoid using AI applications that collect and sell user data to questionable businesses, and do we as citizens have a moral responsibility to inform ourselves about and protest against the use of biased AI applications by the police (c.f. [25]) The responsibility of the individual as a citizen and consumer regarding our environmental impact has been debated for decades among academics, activists, politicians, and business strategists (e.g., [26–31]). We believe the same questions can and should be discussed in relation to AI. Although for responsibility assignments to be effective, more specific goals might be needed ([32]), we are not here going to present any definitive answers to specific questions about what to do in this or that situation, or even to the general question of what responsibility the individual has when it comes to guiding the development of AI in a desirable direction. Our ambition is merely to bring the general question one or two small steps closer to an answer.

There is an ongoing discussion about these issues within environmental ethics, and nowhere is the discussion as vigorous and directly policy relevant as in climate change ethics, where the centre of the philosophical, as well as of the political discussions is the question of responsibility. That is, who should do what to mitigate and adapt to climate change? If we can apply the existing discussions from this field to the question of responsibility for the development of AI, much would be won, since we will be able to make use of and learn from these discussions. Our basic approach in this paper is therefore to make use of the most influential principles for distribution of responsibility in that context and investigate if these principles can be adapted to the question of responsibility for the development of AI, and what they then have to say about this latter question.

2 Some initial clarifications

One thing we should clarify from the start is that this paper is about moral, not legal responsibility. Also, the paper is not going to deal with the question of who should be blamed, fired, or fined when an autonomous machine messes up. This is an important and often discussed question in AI ethics

and law (e.g., [33–35]), but it is not the question we are discussing here. We will instead discuss the responsibility for guiding the development of AI in a desirable direction. That is, we will discuss forward-looking responsibility [31, 32, 36–41] (also known as prospective responsibility [42, 43], projective responsibility [44], or ex ante responsibility [45, 46]).¹ When we need to refer to the other kind of responsibility (that deals with the question of who should be blamed or praised for things that have already happened), we correspondingly call it ‘backward-looking responsibility’ (also known as ‘retrospective responsibility’ [49–51], ‘outcome responsibility’ [52], ‘ex post responsibility’ [45, 46], or ‘accountability’ [44]). When we simply use the term ‘responsibility’ without clarification we will thus mean ‘moral, forward-looking responsibility’ in the sense described above.

The question of forward-looking responsibility for the individual contains several sub-questions. One such sub-question is how to distribute responsibility between individuals on the one hand and other actors, such as governments, companies, and other organisations, on the other (see e.g., [53, 54]). This question, of course, in turn, points at the question of whether other entities than individuals can be assigned responsibility at all (see e.g., [44, 50, 55–64]) for discussion). In a deeper sense, we think they most probably cannot. However, in a more pragmatic sense, entities such as companies, etc. can in practice act as individuals and be assigned responsibility (even though they are not conscious beings themselves and cannot form their own preferences, etc.) (e.g., [65, 66]). This is also the level we are interested in here to make progress in a practical sense that has the potential to actually influence the development of AI. That it is possible and fruitful in practice to talk about responsibility on this level is confirmed by legal systems that treat both companies, nation-states and certain other organisations as legal entities with legal responsibility (e.g., [67]). It is also confirmed by the way most people talk about and interact with such entities. It is not uncommon, for instance, to hear people say things like “it is the government’s responsibility to make it easy for people to choose more sustainable power sources” or “retailers have a responsibility to make sure every node in their supply chain respects human rights”. When people talk about and interact with non-individuals in this way, it is probably not based on a belief that these entities are conscious in themselves or that the entity as such can understand and deliberate about the consequences of “their”

¹ Responsibility in a forward-looking sense is often referred to as remedial responsibility (e.g., [47]), i.e., doing something about harm that has already arisen or that will arise in the future if no measures are taken now. This differs slightly from the use of ‘forward-looking responsibility’ in this article, as our focus is on how to lead the development in a certain direction. A more general notion of forward-looking responsibility as obligation is suggested by [48].

actions, but more probable on an implicit assumption that doing so is practical.

In addition to the question of how to distribute responsibility between individuals and non-individuals, we need of course also to think about how to distribute responsibility among individuals, as well as whether the distribution of responsibility among individuals on one hand and the distribution of responsibility between individuals and other entities on the other, should be treated as two separate questions, or as one question where each individual and each non-individual entity should be seen as nodes in the same distribution scheme. It would doubtlessly be easier to stick with one of the first two options, that is, to discuss either the distribution among types of entities, such as individuals versus companies versus countries, or the distribution among the members of the same type, such as among individuals or among countries, etc. The third, mixed option, inevitably carries with it some practical difficulties, for instance, when it comes to comparing abilities, need and desert between such diverse entities as states, companies, and individual consumers. On the other hand, the mixed option is more in accordance with how other things are usually distributed in the real world (c.f. [68]). Even though resources are, in some instances, first distributed between types of entities and then distributed among the entities of each type, this is not the whole picture. Typically, money, as well as other resources, is in a constant flow among entities of different types. Individuals pay companies for goods and services. Companies pay wages to individuals and taxes to governments, as well as paying other companies for goods and services. Responsibility is of course not traded like money and other resources, but it does seem to us that it makes sense to allow for the same kind of flexibility when discussing the distribution of responsibility.

3 Distribution principles

Our main method for shedding some philosophical light on the question of individual responsibility for guiding the development of AI in a desirable direction is to see what happens when we apply some of the most influential principles for distribution of responsibility in climate change ethics and ask, how can we apply them to our main question, and what implications will they have if we do? We will discuss each of the five principles for distribution in light of these two questions.

The literature is full of theories about distribution. As stated above, we find the discussions about distribution of responsibility for climate change mitigation and adaptation to be a useful starting point for our discussion. This is a discussion that has gone on for a while, that has engaged many philosophers as well as non-philosophers and produced a

large treasure trove of literature, and that is directly policy relevant. We have therefore chosen to focus our investigation on theories for distribution that are influential in these discussions (see e.g. [40, 41, 60]). Each of these basic principles comes in different variants. In some cases, but not all, they also come in both consequentialist and deontological versions. Some also come in different versions depending on whether we talk about the distribution of something good or of something bad. There are also, of course, different mixes of these five basic principles.

The five basic principles for distribution that we will apply are:

- Effectiveness (e.g., [69, 70])
- Equality (e.g., [71–74])
- Desert (e.g., [60, 61, 72, 75])
- Need (e.g., [71, 73, 76–79])
- Ability (e.g., [60, 61, 80])

Different ethicists, activists, planners, and decision-makers, as well as different people in general, seem to prefer different distribution principles [41], and it cannot be denied that there are good arguments for all of these principles. In this paper, we will not assess the general merits of each of the principles as such, but only their applicability to our particular problem. Ideally, it would be desirable to be able to distribute responsibility in a way that satisfies all of them. This is not always possible, however, even though, as we will see shortly, the relationships between the different principles, as well as between different versions of the same principle, are complicated. There is some overlap regarding which distributions different principles recommend in practice, and there are also cases where they limit or strengthen each other.

4 Effectiveness

Effectiveness is a very common basis for distributing responsibility, in practice as well as in economic and planning literature. It does not make much sense to argue that effectiveness has value as an end in itself. It does make sense, however, to claim that effectiveness has instrumental value in relation to most, if not all end values. This means if we interpret the other distribution principles in a consequentialist sense, that is, that the aim of these distribution principles is to distribute responsibility in such a way that it best promotes the values behind the principles (e.g., equality or need mitigation), effectiveness should probably at least be a factor in the distribution as such.

Even so, it can also be argued that effectiveness should not be the only basis for the distribution of responsibility. That the distribution is fair is commonly mentioned as an

important end value and not seldom as a moral requirement. There is also instrumental value attached to whether a distribution is perceived as fair. No matter how effective a distribution looks on paper, it will not work in reality if those who are affected by it perceive the distribution as being grossly unfair. This means, in order for a distribution of responsibility to work, it must be possible to justify that distribution to those affected by it, and especially to those who will have to shoulder a bigger responsibility, which probably implies that the chosen distribution needs to, at least, contain elements of several of the other distribution principles mentioned above.

If we accept effectiveness as one basis for the distribution of responsibility for AI development, where does that leave us concerning the responsibility of individuals? One thing we can probably say with a high degree of confidence is that effectiveness is closely connected to ability, and that in turn seems to imply that entities, such as big companies, influential international organisations and the governments of rich and powerful nations, should probably take on a higher degree of responsibility than most individuals. Since we can also find large differences in ability among individuals, it looks like we should probably also go for a highly differentiated responsibility among individuals. At least, that will be the first answer. As we will see soon, however, ability is not a fixed unit, something that will complicate things a bit.

Given this preliminary answer, we may also ask whether it is actually more effective to exclude individuals completely. In environmental ethics, there is an ongoing debate regarding the effectiveness of individual responsibility. The literature seems to be divided on the issue, and so is apparently the environmental movement and the population in general. We do obviously not have an answer regarding the corresponding question in relation to AI development, but we do want to raise awareness of the issue. It is an important question that needs to be discussed also in relation to AI development.

5 Equality

Equality is a popular basis for distribution (e.g., [58, 61, 71, 74, 77, 78, 81–87]). When discussing the distribution of responsibility in connection with climate change mitigation, Peter Singer suggests equality as the natural starting point that should only be diverted from for good reasons [73]. Is equality applicable to our question, and what implications would it have? First, we need to distinguish between, on the one hand, an equal distribution of the responsibility as such (deontological interpretation), and, on the other hand, distributing the responsibility for AI development in such a way that it promotes equal distribution of something else (consequentialist interpretation), for instance of wealth, or the risks and benefits produced by AI (see e.g. [88]) on the

difference between procedural and substantive fairness, see also [89] on what objects to be equally distributed).

If we aim at an equal distribution of the responsibility as such, the degree of equality that will be achievable will be limited by the degree of inequality in other areas (e.g., [90]), including unequal distributions of knowledge, power, and money. Considering the great differences among different actors when it comes to all of these things, absolute equality of responsibility seems not even to be an option. It might still be possible to increase the equality of responsibility to some degree, however, by amending or compensating for these other differences. If we find increased equality of responsibility a worthwhile goal, in itself or for some instrumental reason, it is therefore important to also address other inequalities, including economic inequalities, differences in power and differences in knowledge (especially knowledge about AI technology) (c.f. [91]). If we determine that a more equal distribution of responsibility is morally required but that other inequalities are in the way, it seems that we are also morally required to address these other inequalities. For instance, it could be reasonable that governments and other institutional agents have a responsibility to enable individuals to act responsibly (c.f., [31, 54]).

Let us take a look at the other interpretation of the principle of equality-based distribution, that is, that we should distribute the responsibility for AI development so as to promote a more equal distribution of something else (maybe in particular, the distribution of risks and benefits from AI).

The development of AI is expected to have an immense impact on the question of equality in several ways. AI has the potential to increase or decrease the gaps among groups of people in many areas of life, such as income, health, power—you name it. Whether AI will widen or close these gaps is essentially up to us and we believe that this is one of those areas where the path we choose now at the beginning of the AI development is very important for directing the long-term development. If some get a strategic advantage now, it is highly plausible that this advantage will just grow bigger with time. Wealth tends to create more wealth. Technological advantages tend to create bigger technological advantages.

What is the practical relation between the deontological and the consequentialist interpretations of the equality-based distribution principle? It does not automatically follow that an equal distribution of the responsibility as such is the best way of achieving a resulting equal distribution of something else. We can expect, however, that making more, and more different, groups of individuals involved in decisions about AI development will result in more perspectives being represented, and that in turn will probably lead to a development of AI where at least the direct benefits and risks from AI, and maybe also more indirect risks and benefits following in the tracks of the development of AI, will be more equally distributed. So even

though a completely equal distribution of responsibility does not seem to be a realistic option, a more equal distribution of responsibility for the development of AI would probably help push the development towards a more equal distribution of the risks and benefits of AI. This is probably true both if we talk about the distribution of responsibility among individuals and if we talk about the distribution between individuals and other actors. The benefits of an equal distribution of responsibility as a means to a more equal distribution of other things obviously have to be balanced against the benefits of involving experts (in AI but also in other relevant areas). Knowledge is not a fixed unit, however, and it is not a necessary truth that all experts have to come from groups that already have a strong influence on the development of AI.

We will get back to the question of knowledge and other abilities later, but we need to look more closely at the relation between influence and responsibility. In the reasoning above, we assume that increased responsibility for members of underrepresented groups will also increase the influence of these individuals on the development of AI. The idea behind this assumption is that if we accept that someone has a certain responsibility, then we also have to accept that she has a corresponding level of influence over the matter. It makes little moral sense to say that “this is your responsibility”, and in the same breath state that “you have no say in this matter”.

It can even be claimed that if we actively assign a certain responsibility to someone else, we also have to actively concede to her the corresponding degree of influence over the matter necessary to live up to this responsibility. An annual meeting of an organisation that decides that the board is responsible for making sure the economy is in order but also decides that the board does not have the right to decide about the economic activities of the organisation would be acting both immorally and irrationally.

If we accept that ethics, as well as reason, demands that assigning responsibility to someone implies an obligation to grant her the corresponding influence, then assigning less responsibility to members of underrepresented groups can be used as an excuse to deny them the corresponding influence, while assigning more responsibility to members of the same groups is actually a way of empowering the members of these groups (c.f., [92]). These double edges of responsibility illustrate how distributing responsibilities is an act of power [31, 93, 94]).

6 Desert

Using desert as a basis for distribution means one should get (or pay) what one deserves. Desert is discussed as a basis for distribution in many circumstances and of many different phenomena, positive as well as negative (e.g. [49, 58, 60, 61,

71, 72, 74, 75, 78–80, 84–88, 95–97]), but c.f. [98]. When desert is discussed in connection with forward-looking responsibility, it is often in the sense that by having, or being assigned, forward-looking responsibility, one deserves some compensation (e.g., [71, 99], but c.f. 100). This is not the focus of this paper, however. Our focus is instead on whether desert can be a good reason for assigning responsibility and what implications that would have.

Whether we want to ground a distribution on desert in terms of guilt or merit usually depends on what it is we are about to distribute and whether that is considered to be desirable or not. Typically, distribution of responsibility is treated as a distribution of something negative (e.g., [101]). On one hand, this makes sense. Having responsibility can be perceived as a rather heavy burden. It is also associated with certain costs in that one probably needs to put some time into thinking about how to live up to the responsibility, and it may demand that one adapts one’s behaviour. When we talk about backward-looking responsibility, it is also often connected with blame or punishment if you are found responsible for something negative (e.g., [102]). Even though we here talk about forward-looking responsibility, we cannot totally get around the issue of backward-looking responsibility since the two are often connected (e.g., [48, 103]). Having forward-looking responsibility to do or not do something can easily be transformed into a backward-looking responsibility if you fail in your forward-looking responsibility, and backward-looking responsibility can in turn easily be converted into forward-looking responsibility if your failure results in a mess that you will be responsible for clearing up (c.f., [32]). Either way, we can conclude that forward-looking responsibility can be a burden.

On the other hand, responsibility can also be something positive. It can feel nice to know that your decisions matter. It can be seen as an honour to be assigned responsibility, and as we have seen, responsibility is closely connected with influence, which may feel nice and can also be instrumentally valuable even beyond its usefulness to fulfil the responsibility in question (c.f., [104]).

Considering that desert is about distributing something based on what the potential recipients have already done (good or bad), we might ask whether desert can at all be considered as a basis for determining forward-looking responsibility. As we noted above, however, the distribution of forward-looking responsibility is quite often based on backward-looking responsibility. The responsibility to clear up a mess, for instance, is about what we should do, but it is often thought to depend on one’s guilt in creating the mess ([51]; c.f., [105] on the distinction between objective and subjective guilt). It is thus clearly possible to base forward-looking responsibility on desert. The question is, is it applicable to the particular situation discussed here? That is, is desert applicable to the issue of individual responsibility to lead the development of

AI in a desirable direction? Here, it is arguably not a matter of cleaning up an existing mess. It is more strongly future-oriented than climate change mitigation that is about handling a mess that we to a large extent have already created (e.g., [106, 107]). Though there is already a fair amount of mess to clear up regarding our present use of AI, that is not the focus of this paper. Our focus is entirely on the responsibility for guiding the future development of AI in a beneficial direction.

One possible way of reasoning here that would make the desert principle more future-oriented, is to talk about *risks* for future harm and *potential* for future benefits (c.f., [40, 108]), rather than about harms or benefits that have already materialised. So, instead of saying that some individuals have deserved more responsibility than others for the development of AI because they have created harm or benefits as such, we could point out that they have created, or are in the business of creating, risks and potential for future benefits.

If we accept desert as a basis for distributing responsibility in connection with AI development, and that guilt and merit can be construed in terms of the creation of risks or potential benefits as a basis for distributing responsibility, then we should assign greater responsibility to entities that have created, or are in the process of creating risks or possibilities for future benefits.

In practice, this probably means that those who work with AI, or AI applications, have a greater responsibility for the development of AI since if you work with these things, you create risks or potential for benefits to a larger degree than those who do not work in these fields. The focus on desert based on the production of risks and benefits with AI will therefore in practice often coincide with the existing idea of grounding responsibility on control [109], (c.f., [32]).

In practice, this probably means that companies and governments will have more responsibility than most individuals. It also seems to indicate that those individuals who work with AI or AI applications in their professional lives also have a greater responsibility as professionals. Does this also mean that they have a greater responsibility for how they act as consumers and citizens? This sounds like a question that needs to be looked into in future research.

We can also notice a potential conflict here between desert-based and equality-based distribution since assigning responsibility on the basis of being in the business of creating risks or potential benefits of AI would increase the responsibility, and therefore, as we have seen, the influence of those who are already influential in this area.

7 Need

What about need as a basis for distributing responsibility? Need-based distribution makes intuitive sense for the distribution of resources and is often promoted in that context

(e.g., [78, 110]). It makes good ethical sense that those who are starving should get more food than those who are already full. In climate change ethics, need is sometimes used as a basis for responsibility distribution in the sense “everyone for herself”. That is, if you are more vulnerable to the effects of climate change, it is up to you to mitigate and adapt [71, 79]. More often, however, need is used as a negative basis for responsibility, in the sense that those who are already burdened by need in a general sense (that is, being poor) should not have to also shoulder the responsibility for climate change mitigation and adaptation [73, 77, 78].

How can we use need for guiding the development of AI in a desirable direction?

First, we need to define ‘need’ in a way that makes sense in this context. How to define ‘need’ in general and in connection with the distribution of resources has been intensely discussed [110]. We are not going to contribute to that general debate. Instead, we make the limited suggestion that in this particular context, it makes sense to interpret ‘need’ in terms of vulnerability to the effects of AI. This way, the principle will be immediately relevant to the question of AI development, but it is still within the general scope of need-based distribution as this principle is commonly used, including how it is often used in environmental ethics, which allows us to make use of experiences from discussions in that area.

Need-based distribution can, just as equality-based distribution, refer to the distribution itself or to the result of the distribution. The former interpretation means, in our case, that the responsibility for the development of AI should be distributed based on how vulnerable different entities are, while the latter means that the responsibility should be distributed in such a way that it results in a more favourable, especially a safer, outcome for those who are more vulnerable.

What is the best way of distributing responsibility for the development of AI if we want the outcome to be particularly advantageous to those most vulnerable?

The most famous answer to this question is perhaps the statement “From each according to his ability, to each according to his needs” [111], see also [112, 113]. The idea that those who are more able to contribute should do so, and that this should benefit those in more need, is often used in climate ethics (sometimes under different names) (e.g., [73, 75, 77–79]). When applied to the question of responsibility for AI development, this would mean that we should distribute the responsibility for AI development in such a way that it best helps those most vulnerable. Marx’s principle could then be translated into the statement that “AI should benefit those most vulnerable, and the development of AI should be guided in that direction by those most able to do so”. If we accept the

consequentialist interpretation of the need-based distribution principle, this makes good sense, though it will of course also be vulnerable to the usual critique of Marx's principle, including the objection that it is neither fair nor productive to “punish” the most able. As we have seen, however, responsibility cannot just be seen as a cost but also as something positive, which should take some of the edge off that objection in our particular context.

Effectiveness and ability are thus important parts of a distribution of responsibility aiming at helping the most vulnerable. As we noted in connection with the consequentialist version of the equality-based distribution, it is probably also important to involve those whom we want to benefit in the actual decisions. It is thus probably also important that those most vulnerable should have at least some influence over the development, which brings us back to the importance of the connection between responsibility and influence. That is, assigning more responsibility to those who are more vulnerable will be a way of putting pressure on everyone to also concede more influence on members of this group.

Need-based distribution of responsibility is often given a negative interpretation. That is, the more need one experiences (in this case, the more vulnerable one is), the less responsibility one should have to shoulder. In connection with climate change mitigation and adaptation, it is often stated that countries with a low level of economic development should shoulder less, or none, of the mitigation and adaptation costs (e.g., [58, 74, 77, 84, 86, 114, 115]). This makes good sense if we talk about the responsibility to cut down on emissions or pay for adaptation measures, but is it a good idea when we talk about guiding the development of AI in a desirable direction? Cutting down emissions and paying for adaptation is easy to interpret as costs. As has been pointed out several times in this paper, however, responsibility may well be perceived as positive as well as negative. This is not least true in the case we discuss here. To influence the development of AI can no doubt be costly in terms of time, engagement and even money, if it, for instance, involves buying more expensive tech products to avoid supporting companies that use biased algorithms, etc. There are also clear advantages, however. By being part of the process of guiding the development of AI, one also gets to influence this development. Assigning lower responsibility for members of vulnerable groups, therefore, seems risky in the sense that it will probably also limit their influence. It might even be used as an excuse to deliberately limit the influence of the most vulnerable.

What does this mean in practice for the application of a need-based distribution of responsibility among individuals and between individuals and other entities regarding the development of AI? Both individuals and collectives are of course affected by AI, but since individuals are typically more vulnerable, this seems to point in favour of at least

some individual responsibility and more so for more vulnerable individuals.

8 Ability

What about ability as a basis for responsibility for the development of AI? Ability is an important basis for the distribution of responsibility in different contexts, for instance the responsibility between system designers and individuals for traffic safety [37] or for the responsibility to cut down greenhouse gases or paying for adaptation [60, 71, 74, 78, 79, 80, 86]. In these cases, ‘ability’ is used in a wide sense, including for example knowledge, money, control and political influence. Can ability in this sense be applied to the question of responsibility for guiding the future development of AI in a beneficial direction?

A common objection to an ability-based distribution is that it does not seem fair that those who have acquired certain skills, should be “punished” by having to take on a bigger responsibility. It is hard to deny that there is some intuitive support for this objection. The objection assumes, of course, that responsibility is predominantly a negative commodity. As we have seen, this is far from the whole truth, which makes the objection less relevant in our context.

A good argument in favour of ability as a basis for the distribution of responsibility has to do with effectiveness. For the consequentialist interpretations of the equality- and need-based distribution principles, that is if we care more about the outcome of the distribution than about the distribution itself, this is probably important. If we want to distribute the responsibility for AI development in such a way that, for instance, AI develops towards a more equal distribution of the risks and benefits of AI, or in such a way that it will help, or at least not worsen, the situation for the most vulnerable individuals, it is probably a good idea to give more responsibility to those most able to achieve these ends. The trick will of course be to identify the relevant abilities, and the individuals who possess them, which might not be the same individuals for both tasks. It may also take a wider definition of what counts as an expert than what is traditionally the case when it comes to AI. It may for instance be that knowledge about the technical aspects of AI is not the only area of importance. Knowledge of ethics, political science, development studies, etc., could be just as important.

A word of warning seems to be in place, however. We have already noted that an important feature of responsibility is the influence that has to come with it for the responsibility to make operational as well as moral sense. In the case of ability, this means we also have to be careful to assign too much responsibility, and thus too much influence, to a particular group. In connection with equality-based and need-based distributions of responsibility, we were discussing the

assignment of more responsibility to presently underrepresented or vulnerable individuals to empower the members of these groups. In connection with ability-based distribution of responsibility, the situation is largely the opposite. Those with more knowledge, resources and power can be expected to already have significant influence. Even though assigning more responsibility to this group probably provides considerable advantages in terms of effectiveness, there is also a risk that it will increase an already uneven distribution of influence.

Another aspect that cannot be overlooked when we talk about ability in connection with responsibility is the almost universally accepted principle that ought implies can [116]. You cannot have the responsibility to do something you cannot do. This means that whatever other basis we want to use for distributing responsibility, ability also has to play a part in the distribution, if nothing else, as a limiting factor in the sense that no matter how valuable it would be from some other perspective to assign responsibility to X, we cannot do it if X does not have what it takes to live up to that responsibility.

For our investigation, this seems to indicate that those with little knowledge, small resources or limited power and control cannot have a lot of responsibility for guiding the development of AI in a certain direction.

In reality, things are a bit more complex, however. So far in this section, we have talked about ability as if it were a fixed unit. People have the abilities they have. What about when ability is not a fixed unit? Usually, it is not. Usually, ability can change. This means we could at least ask the question: If we have independent moral reasons for wanting to increase X's responsibility for Y but X lacks the ability to do that, does this imply an additional moral responsibility on the part of X to increase her ability to take on a larger responsibility, and/or a moral responsibility on others to help increase X's ability to take on a larger responsibility for Y? If there, for instance, are strong egalitarian reasons for wanting more individuals from underrepresented groups to take on a greater responsibility for guiding AI development in a desirable direction, do less able members of these groups have an additional responsibility to increase their relevant abilities, including their knowledge about AI and other relevant fields, and do every able individual and institution have a responsibility to help with this? It seems to us that this indeed makes sense. If we attach high value to, or see a moral requirement for, for instance, a more equally distributed responsibility or for an increased responsibility for those most vulnerable to the effects of AI, then maybe all of us have to accept a responsibility to increase the ability for those who need it to be able to live up to this responsibility.

In the section about need-based distribution, it was concluded that for it to make sense, ethically as well as rationally, to assign responsibility to someone, that responsibility

must be matched by the necessary influence. In this section, we have concluded that if we have independent reasons to assume that a particular individual, group of individuals, or more individuals in general, should have more responsibility for the development of AI, everyone has a corresponding responsibility to increase, or help to increase the ability necessary to take on that responsibility.

Before we close this section about ability as a basis for responsibility to guide the development of AI, let us, in true philosophical tradition, see what happens if we try to take it to its extreme. What if we turn the ought implies can-statement around and say that can implies ought (c.f., [31])? Is it a reasonable and justified move, and what would it mean? Obviously, it cannot stand by itself as a general rule. That we can do something is not sufficient for saying that we have to do it. This is particularly important when we talk about technology, where one of the most important lessons to always remember is that just because one can do something does not mean that one should (see e.g., [117]).

In our case, however, we discuss something that we assume we agree should be done (guiding AI development in a desirable direction) and the question only regards who should do it. Maybe, in this particular context, it actually makes sense to say that can implies ought? That is, anyone who is able to help with the important task of guiding the AI development in a desirable direction should help to the extent of their abilities. This does make sense to a certain degree. We still have to balance the importance of the task with the cost for the individual, though, to avoid that those most able will have to carry an unreasonably large part of the cost and enjoy an unreasonably large influence.

9 Conclusion

The main purpose of this investigation was to throw some philosophical light on the question “how should the responsibility for guiding the development of AI in a desirable direction be distributed among individuals and between individuals and other actors?”.

To answer this question, we departed from five principles for distribution that are influential in discussions about climate change ethics: effectiveness, equality, desert, need and ability. We applied them one by one on our question of forward-looking responsibility for the direction of AI development to see how it would work out and what implications it would have.

We found that it was possible to apply all five principles on this question, either directly, or in some cases, after some adaptation. Several of the distribution principles were overlapping in their answers to our question, while in some cases, they contradicted each other, and in many cases, they strengthened each other and when combined they could

sometimes answer our question in novel and constructive ways. We also found that several of the principles had more than one possible interpretation that could be applied to our question.

To make desert work as a basis for forward-looking responsibility, we defined guilt in terms of the creation of risks for future bad effects, and merit in terms of the creation of potential for future benefits. This led to the conclusion that those who have the most influence over the development of AI today also have a special responsibility for guiding the development in a beneficial direction. This is not, as such, a new idea. The novelty lies in that desert can be used as an argument for this idea when interpreted to be about risks and potentials instead of about harms and benefits that have already materialised.

Equality-based distribution can be interpreted to mean that the distribution of responsibility itself should be equal, or that the distribution of responsibility should be made in such a way that it results in an equal distribution of something else. We found that with the latter interpretation, effectiveness and ability will be important distribution principles in the sense that those responsible for the development of AI must be good at guiding the development of AI in a direction that results in an equal distribution of wellbeing, and in particular in an equal distribution of the harms and benefits from AI. We also found, however, that if we aim at an equal distribution of the benefits and risks of AI, it is probably also important to strive for an equal distribution of responsibility as such. This is because a more equal distribution of the influence over the development of AI will increase the chance that the risks and benefits produced by AI will be more equally distributed and equal distribution of influence is intimately connected with an equal distribution of responsibility. The basis for this conclusion was another important finding, namely that responsibility requires influence, which in turn means that redistribution of responsibility requires redistribution of influence.

A more equal distribution of responsibility is thus an important tool for achieving a more equal distribution of the risks and benefits of AI, and therefore also for equality in general (considering the great influence AI is expected to have over all aspects of our lives).

To achieve a more equal distribution of responsibility we probably also have to address other (e.g. social, economic, educational) inequalities that hinder underrepresented groups from taking on a larger responsibility for the development of AI.

Need-based distribution of responsibility can just like equality-based distribution be interpreted in a deontological and a consequentialist sense, where the former means that the distribution as such should be based on need, while the latter means that responsibility should be distributed in such a way that it favours those who are in most need, interpreted

as those most vulnerable to the effects of AI. The deontological interpretation also comes in two versions, stating that individuals with more need should have more responsibility or stating that individuals with more need should have less responsibility.

In climate change ethics, the idea that those in more need should have less responsibility to contribute to mitigation and adaptation is very influential. When applied to our question, however, we found that limiting the responsibility of those with more need, defined as those who are less affluent in general terms, or more vulnerable to the effects of AI, implies a risk that the influence of these groups will also decrease. Increasing the responsibility of those with more need, on the other hand, is a way of strengthening the influence of these groups over the development of AI. This follows from our finding that if we acknowledge that someone has a certain degree of responsibility, we are required by ethics as well as by reason to also accept that she must have the corresponding degree of influence to be able to live up to the responsibility, and to help the vulnerable groups take on this responsibility.

This conclusion points at our discussion about how responsibility can be both a burden and an opportunity to influence and control, but also at a more general aspect of distribution of responsibility as an act of power. If those who are more vulnerable have more influence over the development of AI, it will increase the probability that AI will develop in a direction that is more helpful, or at least less risky, for those who are more vulnerable.

Effectiveness-based distribution of responsibility was found to have a large instrumental value almost independently of what the aim of the distribution is. Even so, other considerations will probably have to play a role too. If we take a deontological position on ethics, it will be important that the distribution as such is fair, which means some of the other distribution principles (e.g., equality, desert or need) will also have to be considered. Fairness also has an instrumental value since a distribution that looks efficient on paper will not really be efficient if those who are affected by it do not perceive the distribution as fair.

We also noted that if the aim of our distribution is to promote equality or alleviate need (i.e., if we accept the consequentialist interpretations of equality-based or need-based distribution), effectiveness probably demands a fairly equal distribution of responsibility as such, or a distribution that assigns some significant degree of responsibility to those with more need.

Finally, we found that the almost universally accepted principle that ought implies can, means that the differences in ability among actors will be a limiting factor for most other distributions. In addition, we found that ability, in a wide sense that includes knowledge, money, control, power, etc., is an important part of an effective distribution, and thus

also of the consequentialist interpretations of equality- and need-based distributions. We also found that even though it is generally true, not least in connection with emerging technologies, that just because one can do something does not mean one should, in the special case where we agree of what should be done (in this case guiding the development of AI in a beneficial direction) and only inquire who should do it, it does make sense to say that can implies ought, at least to a degree.

We also found, however, that ability is usually not a fixed unit. This means that if we find that, for instance, equality or benefitting those who are most vulnerable is morally required, and that a more equal distribution of responsibility or increased responsibility for those most vulnerable, is important to achieve these ends, then we all also have a moral responsibility to increase the ability among those who need to be assigned more responsibility for the development of AI.

As for which actual distributions are suggested by the different principles, the effectiveness-, desert-, and ability-based principles all seem to point towards more responsibility for the “big” actors, that is, for companies, nations and institutions over individuals, and for individuals that are already more influential, than for individuals who are presently less influential. Since we also found, however, that efficiency to a degree depends on acceptability, and ability is not a fixed unit, these two distribution principles are in fact also compatible with more inclusive distributions.

The deontological versions of equality- and need-based distribution seem to point towards more responsibility for actors that today are less well represented, and for including individuals as well as non-individuals.

The consequentialist versions of the equality- and need-based distribution principles are the most complex when it comes to concrete suggestions. To achieve the goals of these principles, it is probably important to include the deontological interpretations of each principle to some degree, but also to include effectiveness and ability as basis for the distribution of responsibility. In addition, it is essential to account for the relation between responsibility and influence, and for the possibility and obligation to improve upon ability as needed, calling attention to the responsibility of governments and other institutional agents to enable individuals to act responsibly. As we discuss in our analysis, broad inclusion of different groups and individuals is crucial to bring up more, and more diverse, perspectives on direct and indirect risks and potential benefits of AI, which will be necessary to distribute forward-looking responsibility for the development of AI in a way that is both effective and fair.

While guiding the development of AI in a desirable direction is a goal at a high level of abstraction, as mentioned above, more specific goals may be required for responsibility assignments to be effective.

Author contributions Persson contributed knowledge about ethical and philosophical theory and the main part of the writing. Hedlund contributed knowledge about political and social science theory and part of the writing. The ideas and analysis presented in the paper emerged from discussions among the authors where both authors contributed equally.

Funding Open access funding provided by Lund University. The research presented in this paper was funded by the Marianne and Marcus Wallenberg Foundation. Grant number 2018.0020.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bakke, E.: Predictive policing: the argument for public transparency. *NYU Ann. Surv. Am. Law* **74**, 131–171 (2018)
2. Barrett, L.: Reasonably suspicious algorithms: predictive policing at the United States border. *NYU Rev. Law Soc. Change* **41**, 327–365 (2017)
3. Häggström, O. (2021) *Tänkande maskiner - Den artificiella intelligensens genombrott*. Fri tanke förlag. ISBN: 9789189139978
4. Rose, D., Hanheide, M., Pearson, S. (2021, June 23). Robot farmers could improve jobs and help fight climate change—if they're developed responsibly. *The Conversation*. Retrieved June 24, 2021, from <https://theconversation.com/robot-farmers-could-improve-jobs-and-help-fight-climate-change-if-theyre-developed-responsibly-162718>
5. Sharma, G.D., Yadav, A., Chopra, R.: Artificial intelligence and effective governance: a review, critique and research agenda. *Sustain. Fut.* (2020). <https://doi.org/10.1016/j.sftr.2019.100004>
6. Janssen, M., Kuk, G.: The challenges and limits of big data algorithms in technocratic governance. *Gov. Inf. Q.* **33**, 371–377 (2016). <https://doi.org/10.1016/j.giq.2016.08.011>

7. Osoba, O.A., Welsler, W., IV.: An intelligence in our image: the risks of bias and errors in artificial intelligence. *Rand Corpor* (2017). <https://doi.org/10.7249/RR1744>
8. Ranschaert, E.R., Duerinckx, A.J., Algra, P., Kotter, E., Kortman, H., Morozov, S.: Advantages, challenges, and risks of artificial intelligence for radiologists. In: Ranschaert, E.R., Morozov, S., Algra, P.R. (eds.) *Artificial Intelligence in Medical Imaging—Opportunities, Applications and Risks*, pp. 329–346. Springer (2019)
9. Bostrom, N.: *Superintelligence—Paths, Dangers*. Oxford University Press, Strategies (2014)
10. Georges, T.: *Digital soul: intelligent machines and human values*. Hachette (2004)
11. Russell, S.: *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin Random House (2019)
12. Tegmark, M. (2017). *Life 3.0*. Knopf (US) Allen Lane (UK). p. 280. ISBN 978-1-101-94659-6
13. Cellan-Jones, R. (2014, December 2). Stephen Hawking warns artificial intelligence could end mankind. *BBC News*. Retrieved June 15, 2021, from <https://www.bbc.com/news/technology-30290540>
14. Clifford, C. (2018, March 14). Elon Musk: ‘Mark my words—A.I. is far more dangerous than nukes’. *CNBC*. Retrieved June 15, 2021, from <https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html>
15. Cuthbertson, Anthony (2020, July 27). Elon Musk claims AI will overtake humans ‘in less than five years’. *Independent*. Retrieved June 15, 2021, from <https://www.independent.co.uk/life-style/gadgets-and-tech/news/elon-musk-artificial-intelligence-ai-singularity-a9640196.html>
16. David, J. E. (2017, August 14). Elon Musk issues a stark warning about A.I., calls it a bigger threat than North Korea. *CNBC*. Retrieved June 15, 2021, from <https://www.cnbc.com/2017/08/11/elon-musk-issues-a-stark-warning-about-a-i-calls-it-a-bigger-threat-than-north-korea.html>
17. Santos R. A. (2018, March 15). Stephen Hawking warned about the perils of artificial intelligence – yet AI gave him a voice. *The Conversation*. Retrieved June 16, 2021, from <https://theconversation.com/stephen-hawking-warned-about-the-perils-of-artificial-intelligence-yet-ai-gave-him-a-voice-93416>
18. Shead, S. (2020, May 13). Elon Musk has a complex relationship with the A.I. community. *CNBC*. Retrieved June 15, 2021, from <https://www.cnbc.com/2020/05/13/elon-musk-has-a-complex-relationship-with-the-ai-community.html>
19. Clifford, C. (2017, July 24). Facebook CEO Mark Zuckerberg: Elon Musk’s doomsday AI predictions are ‘pretty irresponsible’. *CNBC*. Retrieved June 15, 2021, from <https://www.cnbc.com/2017/07/24/mark-zuckerberg-elon-musks-doomsday-ai-predictions-are-irresponsible.html>
20. Collins, K. (2018, May 25). Elon Musk is ‘exactly wrong’ on AI, says Google’s Eric Schmidt. *CNET*. Retrieved June 16, 2021, from <https://www.cnet.com/news/elon-musk-is-exactly-wrong-on-ai-says-googles-eric-schmidt/>
21. Pinker, S. (2018, February 24). The dangers of worrying about doomsday. *The globe and mail*. Retrieved June 16, 2021, from <https://www.theglobeandmail.com/opinion/the-dangers-of-worrying-about-doomsday/article38062215/>
22. Shead, Sam (2018, September 28). Elon Musk ‘Nuts’ To Call For AI Regulation, Says Facebook’s Chief AI Scientist. *Forbes*. Retrieved June 15, 2021, from <https://www.forbes.com/sites/samshead/2018/09/28/elon-musk-nuts-to-call-for-ai-regulation-says-facebooks-chief-ai-scientist/>
23. Chan, A., Okolo, C. T., Ternner, Z., Wang, A. (2021). The limits of global inclusion in AI development. *Association for the Advancement of Artificial Intelligence*, <https://arxiv.org/abs/2102.01265>
24. Robinson, S.C.: Trust, transparency, and openness: How inclusion of cultural values shape Nordic national public policy strategies for Artificial Intelligence (AI). *Technol. Soc.* (2020). <https://doi.org/10.1016/j.techsoc.2020.101421>
25. Spatola, N.: The citizen at the centre of ethics. *Nat. Mach. Intell.* **2**(2), 85–85 (2020). <https://doi.org/10.1038/s42256-020-0150-0>
26. Caruana, R., Crane, A.: Constructing consumer responsibility: exploring the role of corporate communications. *Organ. Stud.* **29**(12), 1495–1519 (2008). <https://doi.org/10.1177/0170840607096387>
27. Connolly, J., Prothero, A.: Green consumption: life-politics, risk and contradictions. *J. Consum. Cult.* **8**(117), 116–145 (2008)
28. Evans, D., Welch, D., Swaffield, J.: Constructing and mobilizing ‘the consumer’: responsibility, consumption, and the politics of sustainability. *Environ. Plan. A Econ. Space* **49**(6), 1396–1412 (2017)
29. Geels, F.W., McMeekin, A., Mylan, J., Southerton, D.: A critical appraisal of sustainable consumption and production research: the reformist, revolutionary and reconfiguration positions. *Glob. Environ. Change* **34**, 1–12 (2015). <https://doi.org/10.1016/j.gloenvcha.2015.04.013>
30. Isehour, C.: Can consumer demand deliver sustainable food? Recent research in sustainable consumption policy & practice. *Environ. Soc.* **2**(1), 5–28 (2012). <https://doi.org/10.3167/ares.2011.020102>
31. Nihlén Fahlquist, J.: Moral responsibility for environmental problems—individual or institutional? *J. Agric. Environ. Ethics* **22**(2), 109–124 (2009)
32. Van de Poel, I., Sand, M.: Varieties of responsibility: two problems of responsible innovation. *Synthese* **198**, 4769–4787 (2018)
33. Laukyte, M.: Artificial agents among us: should we recognize them as agents proper? *Ethics Inf. Technol.* **19**, 1–17 (2017). <https://doi.org/10.1007/s10676-016-9411-3>
34. Snapper, J.W.: Responsibility for computer-based errors. *Metaphilosophy* **16**(4), 289–295 (1985)
35. Sullins, J.P.: When is a robot a moral agent? *Int. Rev. Inf. Ethics* **12**(6), 23–36 (2006)
36. Hedlund, M.: Epigenetic responsibility. *Med. Stud.* **3**(2), 171–183 (2012)
37. Nihlén Fahlquist, J.: Responsibility ascriptions and vision zero. *Accid. Anal. Prev.* **38**, 1113–1118 (2006)
38. Van de Poel, I., Royakkers, L., Zwart, S.D.: *Moral Responsibility and the Problem of Many Hands*. Routledge (2015)
39. Knaggård, Å., Persson, E., Eriksson, K.: Sustainable distribution of responsibility for climate change adaptation. *Challenges* (2020). <https://doi.org/10.3390/challe11010011>
40. Persson, E., Eriksson, K., Knaggård, Å.: A fair distribution of responsibility for climate adaptation –translating principles of distribution from an international to a local context. *Philosophies* **6**(3), 68 (2021). <https://doi.org/10.3390/philosophies6030068>
41. Persson, E., Knaggård, Å., Eriksson, K.: Public perceptions concerning responsibility for climate change adaptation. *Sustainability* **13**, 12552 (2021). <https://doi.org/10.3390/su132212552>
42. Cane, P.: *Responsibility in Law and Morality*. Hart Publishing (2002)
43. Karp, D.J.: The responsibility to protect human rights and the RtoP: prospective and retrospective responsibility. *Glob. Responsib. Prot.* **7**(2), 142–166 (2015)
44. Kutz, C.L.: Shared Responsibility for Climate Change: From Guilt to Taxes. In: Nollkaemper, A., Jacobs, D. (eds.) *Distribution of Responsibilities in International Law*, pp. 341–365. Cambridge University Press (2015)
45. Kornhauser, L.A.: Incentives, Compensation and Irreparable Harm. In: Nollkaemper, A., Jacobs, D. (eds.) *Distribution of Responsibilities in International Law*, pp. 120–152. Cambridge University Press (2015)

46. Nollkaemper, A., Jacobs, D.: Introduction: Mapping the Normative Framework for the Distribution of Shared Responsibility. In: Nollkaemper, A., Jacobs, D. (eds.) *Distribution of Responsibilities in International Law*, pp. 1–35. Cambridge University Press (2015)
47. Miller, D.: Global justice and climate change: how should responsibilities be distributed? *Tanner Lect. Hum. Val.* **28**, 117–156 (2009)
48. van de Poel, I.: Moral responsibility. In: van de Poel, I., Royakkers, L., Zwart, S.D. (eds.) *Moral Responsibility and the Problem of Many Hands*, pp. 12–49. Routledge (2015)
49. Anton, A.L.: *Moral Responsibility and Desert of Praise and Blame*. Lexington Books (2015)
50. Miller, S.: Collective Responsibility. *Public Aff. Q.* **15**(1), 65–82 (2001)
51. Miller, D.: Distributing responsibilities. *J. Polit. Philos.* **9**(4), 453–471 (2001)
52. Honoré, T.: *Responsibility and fault*. Hart Publishing (1999)
53. Abbot, C.: Bridging the gap—non-state actors and the challenges of regulating new technology. *J. Law Soc.* **39**(3), 329–358 (2012). <https://doi.org/10.1111/j.1467-6478.2012.00588.x>
54. Pettit, P.: Responsibility incorporated. *Ethics* **117**(2), 171–201 (2007)
55. Caney, S.: Cosmopolitan Justice, Responsibility, and Global Climate Change. In: Gardiner, S.M., Caney, S., Jamieson, D., Shue, H. (eds.) *Climate Ethics—Essential Readings*, pp. 122–145. Oxford University Press (2010)
56. Droz, L.: Environmental individual responsibility for accumulated consequences. *J. Agric. Environ. Ethics* **33**(1), 111–125 (2020). <https://doi.org/10.1007/s10806-019-09816-w>
57. Erskine, T.: “Coalitions of the Willing” and the Shared Responsibility to Protect. In: Nollkaemper, A., Jacobs, D. (eds.) *Distribution of Responsibilities in International Law*, pp. 227–264. Cambridge (2015)
58. Gardiner, S.M.: *A Perfect Moral Storm: The Ethical Tragedy of Climate Change*. Oxford University Press (2011)
59. Grasso, M.: A normative ethical framework in climate change. *Clim. Change* **81**, 223–246 (2007). <https://doi.org/10.1007/s10584-006-9158-7>
60. Hayward, T.: Climate change and ethics. *Nat. Clim. Change* **2**(12), 843–848 (2012). <https://doi.org/10.1038/nclimate1615>
61. Jamieson, D.: Adaptation, Mitigation, and Justice. In: Gardiner, S.M., Caney, S., Jamieson, D., Shue, H. (eds.) *Climate Ethics—Essential Readings*, pp. 263–283. Oxford University Press (2010)
62. Lang, A.F., Jr.: Shared Political Responsibility. *Distribution of Responsibilities in International Law*, pp. 62–86. Cambridge University Press (2015)
63. Lucas, J.R.: *Responsibility*. Oxford University Press (1995)
64. Pierik, R.: Shared Responsibility in International Law. In: Nollkaemper, A., Jacobs, D. (eds.) *Distribution of Responsibilities in International Law*, pp. 36–61. Cambridge University Press (2015)
65. Manos, R., Drori, I. (eds.): *Corporate Responsibility: Social Action, Institutions and Governance*. Palgrave Macmillan (2016)
66. Portney, P.: Corporate social responsibility: an economics and public policy perspective. In: Hay, B.L., Stavins, R.N., Vietor, R.H.K. (eds.) *Environmental Protection and the Social Responsibility of Firms*, pp. 107–131. RFF Press (2005)
67. Klabbbers, J.: *International Law*. Cambridge University Press (2013)
68. Neuhäuser, C.: Structural injustice and the distribution of forward-looking responsibility. *Midwest Studies in Philosophy* **38**(1), 232–251 (2014). <https://doi.org/10.1111/misp.12026>
69. Barry, B.: *Theories of Justice*. University of California Press (1989)
70. Garvey, J.: *The Ethics of Climate Change: Right and Wrong in a Warming World*. Continuum International Publishing Group Ltd. (2008)
71. Ringius, L., Torvanger, A., Underdal, A.: Burden sharing and fairness principles in international climate policy. *Int. Environ. Agree. Polit. Law Econ.* **2**, 1–22 (2002)
72. Shue, H.: Transboundary damage in climate change criteria for allocating responsibility. *Distribution of responsibilities*. In: Nollkaemper, A., Jacobs, D. (eds.) *Distribution of Responsibilities in International Law*, pp. 321–340. Cambridge University Press (2015)
73. Singer, P.: One Atmosphere. In: Gardiner, S.M., Caney, S., Jamieson, D., Shue, H. (eds.) *Climate Ethics—Essential Readings*, pp. 181–199. Oxford University Press (2010)
74. Vanderheiden, S.: *Atmospheric Justice: A Political Theory of Climate Change*. Oxford University Press (2008)
75. Caney, S.: Environmental Degradation, Reparations, and the Moral Significance of History. *J. Soc. Philos.* **37**(3), 464–482 (2006). <https://doi.org/10.1111/j.1467-9833.2006.00348.x>
76. Caney, S.: Justice and the distribution of greenhouse gas emissions. *J. Glob. Ethics* **5**, 125–146 (2009)
77. Gardiner, S.M.: Ethics and global climate change. In: Gardiner, S.M., Caney, S., Jamieson, D., Shue, H. (eds.) *Climate Ethics—Essential Readings*, pp. 5–35. Oxford University Press (2010)
78. Grubb, M., Sebenius, J., Magalhaes, A., Subak, S.: Sharing the Burden. In: Mintzer, I.M. (ed.) *Confronting Climate Change: Risks, Implications and Responses*, pp. 305–322. Cambridge University Press (1992)
79. Shue, H.: Subsistence emissions and luxury emissions. In: Gardiner, S.M., Caney, S., Jamieson, D., Shue, H. (eds.) *Climate Ethics—Essential Readings*, pp. 200–214. Oxford University Press (2010)
80. Page, E.: Climatic justice and the fair distribution of atmospheric burdens: a conjunctive account. *Monist* **94**, 412–432 (2011)
81. Atuguba, R.A.: Equality, non-discrimination and fair distribution of the benefits of development. In: Marks, G., Stephen, P. (eds.) *Realizing the Right to Development: Essays in Commemoration of 25 years of the United Nations Declaration on the Rights to Development*, pp. 109–116. United Nations Publications (2013)
82. Blomfield, M.: Global common resources and the just distribution of emission shares. *J. Polit. Philos.* **21**(3), 283–304 (2013). <https://doi.org/10.1111/j.1467-9760.2012.00416.x>
83. Bowie, N.E.: Equality and distributive justice. *Philosophy* **45**(172), 140–148 (1970)
84. Brooks, E., Davoudi, S.: Climate justice and retrofitting for energy efficiency. *disP Plan. Rev.* **50**(3), 101–110 (2014). <https://doi.org/10.1080/02513625.2014.979048>
85. Markandya, A.: Equity and distributional implications of climate change. *World Dev.* **39**(6), 1051–1060 (2011). <https://doi.org/10.1016/j.worlddev.2010.01.005>
86. Moellendorf, D.: Treaty norms and climate change mitigation. *Ethics Int. Aff.* **23**(3), 247–266 (2009). <https://doi.org/10.1111/j.1747-7093.2009.00216.x>
87. Nicklisch, A., Paetzel, F.: Need-based justice and distribution procedures: the perspective of economics. In: Traub, S., Kittel, B. (eds.) *Need-Based Distributive Justice: An Interdisciplinary Perspective*, pp. 161–190. Springer (2020)
88. Temkin, L.: Justice, equality, fairness, desert, rights, free will, responsibility, and luck. In: Knight, C., Stemplowska, Z. (eds.) *Responsibility and Distributive Justice*, pp. 51–76. Oxford University Press (2011)
89. Hansson, S.O.: Equity, equality, and egalitarianism. *Arch. Philos. Law Soc. Philos.* **87**(4), 529–541 (2001)
90. Roemer, J.E.: A pragmatic theory of responsibility for the egalitarian planner. *Philos. Public Aff.* **22**(2), 146–166 (1993)

91. Graafland, J.J.: Distribution of responsibility, ability and competition. *J. Bus. Ethics* **45**, 133–147 (2003). <https://doi.org/10.1023/A:1024188916195>
92. Ulbert, C.: In search of equity: Practices of differentiation and the evolution of a geography of responsibility. In: Ulbert, C., Finkenbusch, P., Sondermann, E., Debiel, T. (eds.) *Moral Agency and the Politics of Responsibility*. Routledge (2017)
93. Bexell, M., Jönsson, T.: *The Politics of the Sustainable Development Goals: Legitimacy, Responsibility, and Accountability*. Routledge (2021)
94. Ulfert, C.: In search of equity: Practices of differentiation and the evolution of a geography of responsibility. In: Ulfert C, Finkenbusch P, Sondermann E & Debiel T (eds.). *Moral Agency and the Politics of Responsibility*, London: Routledge, pp. 105–121 (2017)
95. Caruso, G.D., Morris, S.G.: Compatibilism and retributivist desert moral responsibility: on what is of central philosophical and practical importance. *Erkenn* **82**, 837–855 (2017). <https://doi.org/10.1007/s10670-016-9846-2>
96. Cupit, G.: Desert and responsibility. *Can. J. Philos.* **26**(1), 83–99 (1996)
97. King, M.: Moral responsibility and merit. *J. Ethics Soc. Philos.* **6**(2), 1–20 (2012)
98. Abad, D.: Desert, responsibility and luck egalitarianism. In: Vincent, N.A., et al. (eds.) *Moral Responsibility—Beyond Free Will and Determinism*, pp. 121–140. Springer (2011)
99. Olsaretti, S.: Distributive justice and compensatory desert. In: Olsaretti, S. (ed.) *Desert and Justice*. Oxford University Press (2003)
100. Spitzley, J.: The future of moral responsibility and desert. *Rev. Philos Psychol.* **12**(4), 977–997 (2021)
101. Regan, M.M.: Sharing financial responsibility of river basin development. *J. Farm Econ.* **40**(5), 1690–1702 (1958)
102. Corlett, J.A.: Making more sense of retributivism: desert as responsibility and proportionality. *Philosophy* **78**(304), 279–287 (2003)
103. Vargas, M.: Moral influence, moral responsibility” in *Essays of Free Will & Moral Responsibility*, Trakakis, N., Cohen, D. (eds.). Cambridge Scholars Publishing, pp. 90–122 (2008)
104. Bartling, B., Urs F.: Shifting the blame: On delegation and responsibility. *Rev.Econ. Stud.* **79**(1), 67–87 (2012)
105. Greenspan, P.S.: Subjective guilt and responsibility. *Mind* **101**(402), 287–303 (1992). <https://doi.org/10.1093/mind/101.402.287>
106. Jagers, S.C., Duus-Otterström, G.: Dual climate change responsibility: on moral divergences between mitigation and adaptation. *Environ. Polit.* **17**(4), 576–591 (2008). <https://doi.org/10.1080/09644010802193443>
107. Page, E.A.: Distributing the burdens of climate change. *Environ. Polit.* **17**(4), 556–575 (2008). <https://doi.org/10.1080/09644010802193419>
108. Nilsson, A.K.: *Enforcing environmental responsibilities—a comparative study of environmental administrative law*. Uppsala University (2011)
109. Fischer, J.M., Ravizza, M.: *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press (1998)
110. Traub, S.: Perspectives for a theory of need-based distributive justice. In: Traub, S., Kittel, B. (eds.) *Need-Based Distributive Justice: An Interdisciplinary Perspective*, pp. 1–20. Springer (2020)
111. Marx, K. (1875) *Kritik des Gothaer Programms*. Letter to the Social Democratic Workers’ Party of Germany (SDAP)
112. Gallego, B., Lenzen, M.: A consistent input-output formulation of shared producer and consumer responsibility. *Econ. Syst. Res.* **17**(4), 365–391 (2005). <https://doi.org/10.1080/0953531050283492>
113. Lamont, J.: Problems with effort-based distribution principles. *J. Appl. Philos.* **12**(3), 215–229 (1995). <https://doi.org/10.1111/j.1468-5930.1995.tb00134.x>
114. Baer, P., Athanasiou, T., Kartha, S., Kemp-Benedict, E.: Greenhouse development rights—a framework for climate protection that is “more fair” than equal per capita emissions rights. In: Gardiner, S.M., Caney, S., Jamieson, D., Shue, H. (eds.) *Climate Ethics—Essential Readings*, pp. 215–230. Oxford University Press (2010)
115. Füssler, H.: How inequitable is the global distribution of responsibility, capability, and vulnerability to climate change: A comprehensive indicator-based assessment. *Global Environ Change* **20**(4), 597–611 (2010)
116. Kant, I.: *Critique of Pure Reason*. Cambridge University Press (1999)
117. Hedlund, M.: *Demokratiska genvägar: expertinflytande i den svenska lagstiftningsprocessen om medicinsk genteknik*. Lund University (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.