Philosophy & Technology (2022) 34:1819–1846
https://doi.org/10.1007/s13347-021-00490-3

RESEARCH ARTICLE

Cultivating Moral Attention: a Virtue-Oriented Approach to Responsible Data Science in Healthcare

Emanuele Ratti[1] · **Mark Graves**[2]

Received: 24 May 2021 / Accepted: 21 October 2021 / Published online: 2 November 2021

Abstract

In the past few years, the ethical ramifications of AI technologies (in particular data science) have been at the center of intense debates. Considerable attention has been devoted to understanding how a morally responsible practice of data science can be promoted and which values have to shape it. In this context, ethics and moral responsibility have been mainly conceptualized as compliance to widely shared principles. However, several scholars have highlighted the limitations of such a principled approach. Drawing from microethics and the virtue theory tradition, in this paper, we formulate a different approach to ethics in data science which is based on a different conception of "being ethical" and, ultimately, of what it means to promote a morally responsible data science. First, we develop the idea that, rather than only compliance, ethical decision-making consists in using certain moral abilities (e.g., virtues), which are cultivated by practicing and exercising them in the data science process. An aspect of virtue development that we discuss here is moral attention, which is the ability of data scientists to identify the ethical relevance of their own technical decisions in data science activities. Next, by elaborating on the capability approach, we define a technical act as ethically relevant when it impacts one or more of the basic human capabilities of data subjects. Therefore, rather than "applying ethics" (which can be mindless), data scientists should cultivate ethics as a form of reflection on how technical choices and ethical impacts shape one another. Finally, we show how this microethical framework concretely works, by dissecting the ethical dimension of the technical procedures involved in data understanding and preparation of electronic health records.

Keywords Data ethics · Electronic health records · Microethics · Virtue ethics · Data science

Emanuele Ratti and Mark Graves contributed equally to this article.

✉ Emanuele Ratti
mnl.ratti@gmail.com

Extended author information available on the last page of the article

1 Introduction

The rapid emergence of data science as a field and the pervasive impact of AI and machine learning tools have sparked intense debates on the ethical consequences of how data is used for modeling and prediction. In particular, requirements for responsible data science have been central for the development of data and AI ethics as a discipline. There have been efforts to formulate higher-level principles (such as beneficence or fairness), which should guide research and development of AI tools in a direction consistent with societal values. This way of understanding the discipline has been named *hard ethics* (Floridi, 2018) or *macroethics* (Bezuidenhout & Ratti, 2020), and it plays a fundamental role in informing conversations about governance and regulation of AI more in general as a policy goal (Floridi, 2018). Under this rubric, there has been a proliferation of private and public initiatives for formulating the correct principles (Floridi & Cowls, 2019; Jobin et al., 2019; Saltz & Dewar, 2019). But is this approach effective in making data scientists, and others constructing AI systems within society, "more ethical" and more responsible? While this principled approach has proven effective as a macroethics, a number of issues have emerged when those principles have been implemented in the actual practice of data science. The process described as "from *what* to *how*" (Morley et al., 2020)—*from the principles to their implementation*—has proven to be more difficult than expected. Higher-level principles, given their generality, are very difficult to be located in actual processes characterizing the daily routine of data scientists.[1] But while these problems have now been clearly identified, the solutions proposed by many scholars implicitly stem from the same roots characterizing the problem itself. As Hagendorff (2020b) has noticed, proposed solutions to the "from what to how" problem (Dignum, 2018; Floridi, 2018; Morley et al., 2020) assume that making principles more precise and more narrowly identifying the locus where to apply them will solve the problem. But this class of solutions remains "principled" nonetheless, and this is problematic because of the ineffectiveness of principles and rules in fostering better behavior by themselves (McNamara et al., 2018; Kelly, 2018). In addition to other problems identified by Hagendorff (2020b), in this paper, we want to focus on an implicit conception of "ethics" underpinning these "from what to how" frameworks. Making the principles more precise so that data scientists will be able to unambiguously apply them to their work hides a conception of "ethics" as rule-based and hence as a process that can be automated. Moreover, this rule-based/automation-seeking conception assumes that "being ethical" is to be understood as mere compliance. Within this perspective, compliance can be mindless such that no one is really interested in whether data scientists "learn ethics" and become (more or less) independent ethical agents, as long as they are compliant (Kelly, 2018). In this view, ethical decision-making is improved only by making sure that data scientists are exposed to the right principles or rules, and that they know exactly where to

[1] A specific characterization of these problems has been already done elsewhere (Bezuidenhout and Ratti 2020; Hagendorff 2020a, 2020b; Morley et al., 2020; Mittelstadt 2019).

apply them. This view of ethical decision-making also reflects on what a responsible data science looks like, i.e., compliance- and rule-based.

In this paper, we formulate a different approach to ethics in data science which is based on a different conception of "being ethical" and, ultimately, of what it means to promote morally responsible data science. Rather than mere compliance, ethical decision-making consists in using certain moral abilities (e.g., virtues), which are cultivated by practicing and exercising them in the data science process. Therefore, while we agree with the "from what to how" literature that we should embed ethical decision-making in the actual data science process, we understand this idea of "embedding" as the cultivation of moral abilities in the daily activities of data scientists. *We consider our approach a "microethics"* (Bezuidenhout & Ratti, 2020; Hagendorff, 2020a; Komesaroff, 1995), and we conceive it as a form of ethical training and exercise in which the goal for a data scientist is to learn how to identify the ethical relevance of his/her day-to-day activities. Rather than "applying ethics" (which can be done also mindlessly), data scientists should cultivate ethics as a form of reflection on the subtleties of their technical choices: our goal is to provide a framework to do this. As an example, we show our approach in action in the analysis of electronic health records (EHRs), in particular by describing the dense microethics that emerges even in mundane choices that data scientists face in data understanding and data preparation.

The structure of our paper is as follows. In Section 1.1, we address some preliminary concerns about principles, rules, and character development. In Section 2, we introduce our conception of ethical training which, rather than "applying ethics," aims at cultivating ethical dispositions or moral abilities. These moral abilities (e.g., virtues) should be cultivated in the same way the technical skills necessary to practice data science are learnt (Annas, 2011). This implies that, as skills are learnt by exercising and practicing, so should moral abilities/virtues. In this paper, we focus on a preliminary moral ability that we call *moral attention*. This is the ability to understand how the factors of a situation have ethical relevance and to imagine the ethically relevant consequences of intervening on some of those factors. After having clarified the virtue-theory nature of this training, in Section 3, we specify in detail in what sense technical choices in data science can be "ethically relevant." To operationalize the notion of "ethical relevance," we adapt the capability approach (in particular Nussbaum, 2006 and Ruger, 2010), though only in a procedural and heuristic way. In Section 4, we show how moral attention can be exercised and practiced in distinct stages of the data science pipeline, in particular data understanding and preparation for analysis of EHRs. Section 4 makes clear how rich a microethics can be, thereby making the case for a systematic development and application of our framework for "ethical training" and fostering a more responsible data science. We conclude the article with Section 5, where we provide suggestions for future work, and for understanding the relation between the ethical training we have described and participatory approaches to technology design.

1.1 Rules, Principles, and Compliance

Before we proceed, it is important to qualify more precisely in which sense we think about "rule-based" ethics, and the differences with respect to an approach aimed to cultivate virtue "ethics.'

In the previous section, we have talked about principles, and then we have raised some considerations about "rule-based" ethics. But it is important to emphasize that there is a difference between rules and principles. As succinctly suggested by Zwolinski and Schmidtz (2013), rules function as trump cards, in the sense that "[i]f we have a rule, and can believe with complete confidence that the rule ought to be followed, and if we ascertain that a certain course of action is forbidden (…) that settles it" (p 222). Once we are told that a rule is the right one for the situation, and the rule is completely unambiguous, then there is no reason to discuss anything or to think about the moral saliency of a situation: you just follow the rule, and that is it. This conception of "following the rule"—which has some internal problems—is also connected to the idea that rules are comforting, because "it has the feel of relieving us of moral responsibility" (Zwolinski and Schmidt 2013, p 223). When we follow a rule, we are not really taking responsibility for what we are doing, because there is nothing we can do but follow the rule.[2] Principles are different; they work like weights rather than trump cards, in the sense that they may orient our actions, but there is always plenty of deliberation in understanding how to operationalize and contextualize them, especially when it looks like they are in contrast one with the other. This is why acting on principles leaves us "with no doubt as to who is responsible for weighing them" (Zwolinski and Schmidt 2013, p 223). Therefore, deliberating with principles can hardly be done mindlessly, unlike rules. In other words, "applying" principles effectively requires some abilities or skills in dealing with moral issues. But if this is the case, then why in the previous section we have said that the principled approach suffers from the problems of a rule-based ethics? This is because the idea of solving the "from what to how" problem by specifying more precisely the principles is based on the idea that the shortcomings of principles are addressed by transforming them into rules, in order to guide data scientists at every step of the way. This promotes the comfort of an ethics based on rules, which can be mindless and can remove (most of) responsibility from data scientists themselves. Prioritizing mindless compliance is (in our opinion) one fundamental flaw of this view—it does not address character, and hence, it does not provide a robust "infrastructure" for ethical action, nor does it lead to autonomous ethical agents.[3]

[2] One may also point out, as a reviewer did, that even though in a rule-based approach one is not taking the responsibility of deciding what to do, at the same time he/she is taking the responsibility of doing it. While this is certainly true, it is still a lot that the subject is not taking responsibility for.

[3] It should be noted that the "compliance" paradigm implied by a rule-based ethics has some other problems, as Kelly (2018) argues. First, it fosters a legalistic interpretation of rules "that equates ethical action with adherence to a formal understanding of the rule" (pp 69–70), which in turn promotes loophole reasoning. But in order to avoid loophole reasoning, Kelly continues, the rules are made even more complex, making ethical action sometimes requiring legal training. But the worst consequence is that compliance is perceived as an external constraint, thereby making ethics itself something external to

These considerations alone should be enough to argue that this is not the direction we want responsible data science to go.

But let us assume for the time being that there is nothing wrong with these features of rule-based ethics as applied to data science. After all, one may say, if we have the rules, then the "from what to how" problem will be solved. But the idea that we can have all the rules covering every possible and conceivable situation is controversial. First, there is the normative problem of deciding the exact interpretation of the principles in order to turn them into more precise rules. Given that general principles can sometimes have conflicting interpretations (Binns 2018), it is not clear who gets to decide in which way a principle should be turned into a precise rule of conduct. Second, and most important, it is very hard (if not impossible) to come up with a list of rules that can cover all possible situations (Allen et al. 2000)—there are just too many possible scenarios. We can certainly think about a long-term project in which we compile a long list of rules covering all possible situations a data scientist will find him/herself in, but this will not change the problem that data scientists will be often in situations where principles could not be readily applied to a situation. The only way to overcome the shortcomings of a situation where one does not know what to do with principles is to cultivate abilities to recognize and understand, on a case-by-case basis, how the principles can be applied (in case we still want to focus on principles)—i.e., cultivate moral abilities. This is why we need to cultivate ethics as a form of reflection, rather than as an effort to compile lists of rules.

2 Good Data Scientists: Skills and Virtues

In this section, we introduce our conception of what an ethical training is.

Data scientists learn how to become data scientists not just by reading textbooks, but by developing skills related to that profession. These skills are connected to the practice of coding, to the understanding of statistics, to the ability to sufficiently grasp relevant aspects of a new domain to model them, to interpret results and communicate them as actionable insights, etc. We develop skills by learning how to do the things that a skilled person in a relevant context would do, and we learn how to do such things by practicing: "we learn an art of craft by doing the things that we shall have to do when we have learnt it" (NE, 1103a). In this context, we learn how to code by practicing a lot with coding.

But it is not just a matter of brute practicing. In a complex task, one needs guidance to practice the right skills, in the right way, and in order to strengthen the needed skills; otherwise, one's practice may reinforce many subtler bad habits other than the few obvious good ones. We need to learn from someone who already knows how to do the things related to that particular skill—in other words, we need

Footnote 3 (continued)
practice, and sometimes even alienating. With these remarks, we are not saying that compliance should be abandoned, but that it should not be all that matters.

a *role model* or an *exemplar*. But learning from a role model and developing skills "involve more than copying a role model" (Annas, 2011, p 17). Rather, it is a process that involves a virtuous sort of docility or, to use Annas' expression, the "drive to aspire," which is an active disposition and good will to learn in the right way and understand what to follow of the role model.[4] This is not just passive; in fact, it requires a constant dialogue with the exemplar who provides reasons why certain things are done in certain ways. Many of the skills of data scientists, as well as biologists, experimental physicists, chemists, etc., are acquired in this way. While the "theoretical" part of studying foundational concepts is done earlier in undergraduate education, practicing scientists or data scientists learn the "skills" of their disciplines in flexible and creative ways only later in graduate school or in a professional context.[5] This is not to say that in undergraduate education there is no place for practice. However, especially at the very early stage of scientific education, the training of scientists looks like a "dogmatic initiation" (to use Kuhn's expression), where prospective scientists are taught to perform highly successful and established techniques or experiments of a specific discipline. But these are learnt in the kind of *mindlessly-copying-the-role-model-way* that Annas criticizes. The skills that make a practicing scientist are learnt by practicing ad nauseam in conjunction with learning from a role model (such as a more experienced scientist).

The recent push towards more ethical and responsible AI, machine learning, and data science can be interpreted as suggesting that data scientists should not merely become technically good data scientists, but also *morally good or responsible*. However, the literature is surprisingly silent on how one becomes a morally/responsibly good data scientist. *Here, we suggest that, as data scientists become skilled data scientists by cultivating certain technical abilities, they can become morally good data scientists by cultivating certain moral abilities*. This analogy is supported by the well-established tradition of virtue ethics that, while it separates the nature of technical skills and virtues, recognizes a similarity in the way they are acquired (Annas, 2011). Ideally, the moral abilities that data scientists cultivate will develop into full-blown virtues, but they may not be as strong as some conceptions of virtues imply.[6]

[4] It is important to be more precise about our use of the term "docility." We use this term in the connotation of the scholastic tradition (docilitas), as a twofold "virtue": an openness to learn which is balanced by a critical examination on what is taught. It is interesting to note that the role "docility" in science has been analyzed in recent scholarship (see, for instance, Bezuidenhout et al. 2019). Understood in this way (and not in the pejorative, contemporary connotation), docility is indeed compatible with the "drive to aspire," as a disposition to learn in the right way, by avoiding mindlessly copying.

[5] Even a data scientist who lacks direct mentorship by a role model likely has admired exemplars, teachers, senior colleagues, and the collective knowledge of blog posts, video tutorials, and skill-oriented websites, such as stackexchange.com.

[6] In order to better understand this point, the scope of our ethical training, and the analogy with technical training, it is important to spend a few words on what is meant here by "virtues." In a very preliminary sense, virtues are usually understood as excellences. An excellence is "any stable trait that allows its possessor to excel" (Vallor 2016, p 17)—e.g., skills are excellences. In Aristotle, an excellence is a long-lasting attribute in virtue of which something or someone is good or things go well (Russell 2015). Excellences in ethics—moral virtues—are stable traits and long-lasting ways at being good with respect to how we act and live with other people. Now, we are aware of how controversial it is opting for one notion of virtue rather than another. For instance, it is usually said that virtues are dispositions, but this is both vague and incomplete. It is vague because it may make us think that a disposition to follow moral

In other words, one may be happy to have data scientists cultivate dispositions that are almost virtues, or even just cultivate components typical of character virtues.

The similarity between how skills and virtues are acquired suggests a picture of ethical training that is based *not only* on ethical theories, stand-alone ethics courses, or compliance to principles or rules. Rather, it implies that being ethical is a difficult achievement. Moreover, in the context of data science, it suggests that *ethics is taught and cultivated in the same way in which technical skills are taught*, in particular by practicing in the context where they are needed, and by having exemplars showing you how and why we should make certain choices rather than others. Ethics is embedded in data science, not something external to it (Bezuidenhout & Ratti, 2020), and data scientists should be provided with exercises, heuristics, and novel problems, in order to increase their skills as well as their virtues. Therefore, "learning ethics" is the process of cultivating virtues or moral abilities for the data science context[7]; it "requires time, experience, and habituation to develop it" (Annas, 2011, p 14), and the result is "the kind of actively and intelligently engaged practical mastery that we find in practical experts such as pianists and athletes[8]" (Annas, 2011, p 14).

Our approach has its roots in the virtue ethics tradition, in the sense that we do not ask "what do data scientists should do specifically in this and that particular situation?", but rather "how do data scientists become morally 'good' data scientists?". This question is "virtue-oriented," because the way it is approached is based on a particular answer to the more fundamental question "what does it mean to be ethical?" which is "it means cultivating some traits or abilities (or virtues) to act morally in a given situation." It remains virtue-oriented even if here we do not discuss full-blown virtues, but only moral abilities, because the attitude is virtue-oriented. But at this point, one may again argue in favor of a rule-based type of ethics. After all, our approach does not directly solve the "from what to how" problem because,

Footnote 6 (continued)
rules counts as a virtue, and this is surely not the case. As Annas argues, a virtue is not a natural disposition such as a static tendency (e.g., a glass disposed to break under certain circumstances); rather, it is an active disposition, in the sense that it is being disposed to act in certain ways. A virtue is also a reliable disposition, in the sense one's acting out of virtue does not do it by accident. But as a disposition is also characteristic, in the sense that it is part of a person's character. Moreover, while virtue is the result of habituation, it is nothing like a mindless routine behavior. What we want to say is that any data scientist, in order to be a good data scientist, ideally needs to cultivate some of these active, reliable, and characteristic dispositions to live well with other people which, from the point of view of data scientists qua data scientists, is to make sure that algorithmic systems not only do not harm, but that they also promote flourishing—and this should be the goal of any ethics training. Exactly which specific virtues a data scientist needs to be a good data scientist is an open-question—for instance, Hagendorff (2020b) mentions some of them. We understand that this is a strong conception of virtue; here we will focus (Section 3) especially on a moral ability which may not entirely fit the bill of a full-blown virtue, but that we think is necessary for the other virtues to function properly.

[7] This is something explicitly recognized in Gogoll et al. (2021), when they say that we should treat ethical thinking as a skill, and that "it needs to be practiced and shaped by the software developers (…) This approach would lead to ethical empowerment" (p 20).

[8] These features of virtue-acquisition apply also if we opt not for full-blown virtues, but for relatively stable moral abilities.

a common criticism goes, virtue ethics does not provide any specific rule to follow, and hence no precise algorithm for the "how" part of the problem. But, again, it is impossible to have rules that will cover all possible situations a data scientist will face. Maybe understanding "how" in this automated and mechanical way is misleading in the first place. In virtue ethics terms, the question "how do we design these algorithmic systems ethically?" can also be understood as "how do we make sure that the next generation of data scientists have the skills and abilities to develop ethical algorithmic systems?". This is a more fundamental take on the "from what to how problem," that only a virtue ethics perspective can address.

Although the analogy between technical and ethical training is intuitively convincing, there are three problems with this smooth picture. First, the mere analogy does not specify what is the target of "ethics," what is ethically relevant, and which virtues or moral abilities we should cultivate. Second, even if we had all this information, the analogy by itself would not tell us clearly what to do—how to operationalize those ideas is not clear. Finally, in the context of data science, we lack a fundamental aspect of the cultivation of virtues that is present in the cultivation of skills, namely, the exemplars. The first two problems will be addressed in the next section. Here we want to conclude this section with a few words on the third. As Julia Annas (2011) richly describes in her account of how skills are acquired, the role of exemplars is fundamental because "what is conveyed from the experts to the learner will require giving the reasons" (p 19). This is not a negligible aspect because "giving reasons" or "explaining" "enables the learner to go ahead in different situations and contexts" (p 19). While we have moral exemplars in our world, we probably do not have moral exemplars tailored for the technosocial aspects of data science, which is plagued by a complexity described by Shannon Vallor (2016) as "technosocial opacity." We have plenty of "moral experts" in the sense of "macroethics," but we have already emphasized how difficult it is to embed their insights in the actual practice of data science. Ideally, we would need "expert" data scientist practitioners who already help other data scientists to develop their technical skills, and that at the same time they explicitly do the same for the moral aspects of their job. This is why we think that data ethics should stem necessarily from practicing scientists, rather than external moral philosophers or moral experts.[9] But we are not aware of the existence of such a tradition of data scientists that teach ethics as a daily activity to do *not* on top of other "technical" things, but rather as integrated to the technical acts. Even if we do not have exemplars that we can easily identify, we still think that we can provide a heuristic that can stimulate data scientists to go in the right direction.

[9] This does not mean that "moral experts" cannot function as catalysts for moral self-cultivation—ethicists can be useful in a maieutic process where data scientists learn how to be "ethical.".

3 Microethics: Ethical Relevance and the Capability Approach

In the previous section, we have clarified a general idea of "ethical training," which is based on the analogy between skills and moral abilities, and the way they are acquired. However, we have emphasized some problems, which will be addressed in 3.1 (what is ethical relevance, and which virtues or moral abilities are important to cultivate) and 3.2 (how we operationalize the notion of ethical relevance). Next, we connect what we have formulated in 3.1. and 3.2 to a complete formulation of a microethics approach for the data science context (3.3).

3.1 Ethical Relevance

The fact that decisions can have "ethical relevance" is central to our approach to ethics based on the analogy between virtues and skills. But what does this mean exactly? The answer to this question is connected to the very meaning of an ethical training. Van Wynsberghe and Robbins (2019) lament that, at least in the case of robotics, ethical issues tend to flatten into issues about safety. In other cases, ethics is merely about compliance (Hagendorff, 2020b). But recent literature on Fair-ML (Fazelpour & Lipton, 2020) grasps an important aspect of ethical relevance, namely, that data scientists with their work can shape other people lives' in significant ways. This is the aspect we want to develop further.

Ethics is about how one ought to live his/her own life, and one's conceptions of "living well." Whatever "good life" one sees fit for him/herself, it will consist in certain plans and goals and it will be realized by instantiating certain patterns of actions or behaviors. These patterns of behavior have ethical relevance because they are constitutive of the good life itself.[10] Vallor (2016) makes a similar argument when she explains the relation between ethics and technology more in general, by saying that artifacts and technologies afford specific patterns of thought, valuing, and behavior (p 2), and for this reason shape decisions on how to realize those life plans constituting the good life. Therefore, the ethical training of data scientists we refer to is a training that can help them to appreciate the way their small technical acts (e.g., the way they clean data, the algorithm they choose) can potentially shape data subjects' patterns of behavior, thought, and valuing to the point that can impact data subjects' own perception of what it means to live well. It has been shown that data science plays a huge role in shaping how data subjects, to use Nussbaum's expression, "adjust their preference to what they think they can achieve[11]" (2006, p 73). Because of this "power" of data science, a moral skill or ability that data scientists should develop is *moral attention* (see for instance Vallor, 2016), which is the ability to recognize the ethical relevance of a situation by imagining the way one's own actions will shape other people's actions and thoughts. We say that "moral attention" is a moral ability or a moral skill on purpose. We recognize that if we stick to the

[10] Shaping other people's pattern of behavior means impacting their autonomy, but we do not want to characterize it necessarily as negative.

[11] See, for instance, Susser et al. (2019).

strong definition of virtue that we have reported in the previous section, moral attention would not completely fit the bill of a virtue. This is because typically virtues allow agents to reason, choose, and act accordingly, and moral attention may not be effective in choosing and acting, but only in reasoning. However, it is an ability related to the moral dimension of our lives, it can be characteristic and reliable, and it is acquired in the same way skills and virtues are. But even if it is not a full-blown virtue, we see it nonetheless as strictly connected to other virtues, and to the cultivation of virtues.[12] Without the moral ability of being able to identify the ethical relevance of a situation, other virtues may not even be perceived as being relevant to choose and act accordingly, because the situation would not be considered as in need of moral deliberation in the first place. Therefore, we can say that moral attention is a necessary condition for well-functioning virtues. And because moral attention is a necessary condition for the other virtues to be cultivated, we say that it is a minimal ability that data scientists should develop in order to start their ethical training. We leave the question of which other virtues should be cultivated on top of this moral ability for another article.

3.2 The Capability Approach as a Heuristic to Identify Focal Loci of Ethical Relevance

While the way we have defined "ethical relevance" is useful, it is excessively broad. We impact and shape lives of other people constantly because we are all part of a "common life" (Walzer, 2006). How do we identify those impacts that really matter? We believe that the *capability approach* can provide a useful heuristic to identify ethically relevant and *crucial* impacts, and hence, it is a framework that can be used to develop and cultivate moral attention.

The capability approach is a political and economic program proposed by Sen (1985), which aims to delineate a framework to make comparisons of life quality. Introducing in detail this approach goes way beyond the scope of this paper, so we will just focus on a few key aspects. Here, we use the approach especially in the formulation made by Nussbaum (2006). The cornerstone of the capability approach, which differentiates it from other approaches based on cost–benefit analysis, is that individuals "should have access to the necessary positive resources, and they should be able to make choices that matter to them" (Alkire, 2005, p 117). This seemingly straightforward idea provides the foundation of a distinction that makes the capability approach unique: the distinction between functionings and capabilities. Functionings are "the various things a person may value doing or being" (Sen 1999, p 157), from being nourished to being able to participate in political activities. Alkire claims that functionings are constitutive of a person's being, and for this reason the capability approach can appreciate all changes in a person's life, "from knowledge to relationships to employment opportunities" (p 119). But measuring life's quality only on the basis of achieved functionings—as other consequentialist approaches

[12] This aspect is also emphasized by Vallor (2016), when she says that moral attention is part of a process of moral self-cultivation.

do—is partial. In order to get a comprehensive picture of a person's quality of life, we should also consider the freedom that an individual has in deciding which path to pursue. The question to ask is not only what a person has done, but rather *what a person is able to do and to be*. According to Nussbaum, "the crucial good societies should be promoting for their people is a set of opportunities, or substantial freedoms, which people then may or may not exercise in action: the choice is theirs" (2011, p 18). Capabilities are a range of potential functionings that are feasible for a person to achieve. By distinguishing between functionings and capabilities, the capability approach is explicitly pluralist about values and, at least in principle, avoids paternalism,[13] since the goal of a policy based on this approach should be to expand "people's capabilities and not force people into certain functionings" (Oosterlaken, 2015, p 224). When a person is stimulated to expand capabilities rather than being forced into functionings, that person has the freedom to instantiate those patterns of behaviors and thoughts that he/she thinks constitute the good life and living well.

Nussbaum compiles a list of central capabilities stemming from an idea of what it means for a human to function well. These include life; bodily health; bodily integrity; senses, imagination, and thought; emotions; practical reason; affiliation; being able to live with concern for nature and other species; play; and control over one's environment (for details on this list, see Nussbaum, 2006, pp 76–78). The distinction between functionings and capabilities should be also complemented with the idea that availability of resources is not a sufficient condition for increased capabilities or functionings. There are other factors—such as personal, social, and environmental factors—that determine "the degree to which a person can transform a resource into a functioning" (Robeyns, 2016, p 406), and they are called "conversion factors."

Technology and capabilities are tightly connected (Oosterlaken, 2015; Coeckelbergh, 2010, 2011). As Vallor says, technologies "invite or afford specific patterns of thoughts, behavior, and valuing" (2016, p 2). In terms of capabilities, we can say that technologies can shape both the internal characteristics of individuals and those environmental factors that, combined together, constitute combined capabilities, which filter what can be transformed into functionings. Technology is so important for capabilities not only because we use technical artifacts which sometimes shape what we can or cannot do, but also because those artifacts are embedded in the same sociotechnical systems in which we are embedded, and they arguably shape those systems. For instance, "ICTs [i.e. information and communication technologies] change the ways in which governments and politicians go about their daily business, which may in turn have consequences for an individual's capability to have control over his/her political environment" (Oosterlaken, 2015, p 229). In general, using the capability approach in the context of ethics of information technologies can be potentially very fruitful, given that it "allows to highlight how information technologies shape what people are (or will) actually be able to do" (Coeckelbergh, 2011, p 81).

[13] There is indeed a debate on Nussbaum's capability approach which sees her views as paternalistic, despite the efforts (see Cenci and Cawthorne 2020).

What does this have to do with ethical relevance, ethical training, and the cultivation of moral attention? If actions have ethical relevance in the way defined in Section 3.1, then it is like saying that actions can potentially shape capabilities, in the sense of shaping what people can choose to be or do (i.e., their substantial freedoms). There is overwhelming evidence that the work of data scientists can potentially shape capabilities, in particular by training algorithmic systems for automated tasks in ways that can potentially filter what data subjects can do, be and have access to (Vold & Whittlestone, 2019).[14] Given the overwhelming number of ways in which data scientists impact other people's lives, we propose to use Nussbaum's list as a way to identify crucial loci of impacts that can make an important difference in terms of social justice. In other words, *data scientists' decisions have ethical relevance anytime those decisions impact the substantial freedoms implied by Nussbaum's basic capabilities and the way data subjects can possibly exercise them*. Therefore, the capability approach is used as a heuristic to restrict the scope of what is crucial within what is ethically relevant. However, here the approach itself is not used to decide which course of action is morally the best, at least not at the stage we are discussing it, so rather than an ethical theory, we consider our use of capabilities as a mere approach, which is open to becoming a theory in different ways.

Another way of putting this is to say that we use the capability approach merely as a tool to stimulate and habituate ethical thinking in data scientists' everyday activities. This excludes normative commitments to the evaluation of specific capabilities. There may be other goals in using this approach, which may imply normative choices. For instance, one can use the capability approach as a theory to guide the deliberation between different stakeholders in the design of a piece of technology, such as an algorithmic system. But our "pedagogical goal" (i.e., using capabilities to habituate data scientists to identify what is ethically relevant) is not in contrast with a more "deliberative" approach. First, the two approaches do not exclude one another. Second, we think that habituating data scientists to ethical thinking may make the hypothetical deliberative process as a whole more effective, because data scientists will be more sensitive to many of the issues discussed. In order to explain

[14] In particular, data scientists will impact "conversion factors" or, at least, whether some factors become conversion factors (more on this in Section 4). A famous example of how small technical choices have ethical relevance for capabilities and conversion factors is the study by Obermeyer et al. (2019). In this article, the authors analyze the performances of a typical commercial risk-prediction tool that is used by large health systems and applied to roughly 200 million people in the USA to target high-risk individuals. The aim of the algorithm is to identify those who need additional attention and resources. The problem with this tool, they say, is that it uses healthcare cost as a proxy for health, and hence as a proxy to identify those who need more attention. In particular, the algorithm used demographics, insurance type, diagnosis codes, procedure codes, medication, and detailed costs to predict the appropriate label: this means that "the algorithm's prediction on health needs, is, in fact, a prediction on health costs" (Obermeyer et al. 2019, p 450). But this seemingly uncontroversial predictive proxy generates interesting consequences: only those who have access to proper healthcare in the first instance can be recognized as needing more attention. This means that those who cannot have access to proper healthcare will be ignored by the tool. Therefore, the technical act of choosing certain features to train the algorithm rather than others has transformed some personal, social, and environmental circumstances into conversion factors enabling or disabling the very possibility of making choices pertaining to bodily and mental health. The ripple effect is significant.

this idea better, let us discuss two other works where the capability approach has been used in relation to ethics of science and technology.[15]

In (2020), Cenci and Cawthorne applied the capability approach to solve some specific problems of value-sensitive design (VSD). They recognize that embedding values in the design of technologies can be problematic, because often a specific normative dimension is adopted, and it is usually the one of the most powerful stakeholder, thereby resulting in a paternalistic VSD implementation. However, analyses of how technologies impact the well-being of different stakeholders show that many times there is a tradeoff between different values, and that adopting substantial ethical theories may undermine the endeavor of reconciling seemingly incommensurable values. Cenci and Cawthorne support the idea that ethical proceduralism offers a significant alternative to handle problems that have challenged a VSD based on substantive ethical theories. In order to address problems about pluralism and paternalism, they propose a participatory process where different stakeholders can have their voice heard. They base their view on Sen's deliberative approach to capabilities which, according to them, differs substantially from "Nussbaum's deontological, expert-led, over-specified, complete and perfectionist list of ten basic capabilities" (p 2649). Despite our use of Nussbaum's framework, our goal and Cenci and Cawthorne's goal are compatible. There is indeed a debate on Nussbaum's alleged shortcomings such as paternalism, ethnocentrism, and inattention to contexts (see for instance Claassen 2014). We do not aim to discuss these issues here, but our use of Nussbaum's list is immune to all these shortcomings attributed to her approach. First, we are neutral with respect to Nussbaum's quasi-Kantian claim that sees the notion of human dignity as providing a foundation for the approach. We do not need a strong substantive foundation of the approach for the way we use it, because what we take from it is just the connection between what people can do or be, and how these can impact conceptions of "living well" and the good life. The approach adds depth to the notion of "ethical relevance" by specifying concepts such as internal capability, combined capability, substantial freedom, and functioning. Second, we are also immune to the widely discussed issue of which capability to select. The list is a great starting point for data scientists to cultivate the moral ability or skill of moral attention, given that the ten "central human capabilities" are all moral entitlements, and hence, they have ethical relevance. In other words, we just use the approach to stimulate data scientists to recognize the ethical relevance of their own technical choices, but we leave to them and to their value judgments to decide positive and negative connotations: we want to avoid any form of nudging or paternalism. In fact, we are using Nussbaum's list in a rather instrumental and minimalist way: this list is just a heuristic, in the sense that it can be understood as a way to navigate a complex conceptual space to identify what is ethically relevant and what is not. The positive and negative connotations of shaping capabilities in one way or another is a normative question that should be taken up in an ideal subsequent deliberative process—to which data scientists should participate—among stakeholders affected by the design of the algorithmic system. But the cultivation of moral

[15] We thank a reviewer for raising this point and suggesting a few readings that we missed.

attention via the heuristics of capabilities is also important for that deliberative process. While the participatory design envisioned by Cenci and Cawthorne implies that other people than data scientists should participate in the design process in order to have all voices heard (and we agree), it is important to foster and stimulate the cultivation of a moral sensibility that can predispose data scientists to a more fruitful dialogue both at the epistemic and moral level, which eventually can improve the deliberative process that Cenci and Cawthorne delineate—and moral attention can be important from this point of view. Using the list of capabilities as a tool to solicit ethical questions about the impact of data scientists' technical choices is not only an important starting point; it is also essential, because in the public deliberative process, the ethical relevance of many small technical choices and aspects of algorithmic system construction may be opaque to the relevant stakeholders, and hence, data scientists must be put in the position to recognize them. To sum up, we embrace the procedural perspective of Cenci and Cawthorne, but we see Nussbaum's list as a way to make the procedure more precise and effective (at least long-term) by habituating single data scientists to reason explicitly in ethical terms.

We get to similar conclusions by analyzing an article which is in opposition to Cenci and Cawthorne's work. In (2020), Jacobs implements the capability approach in VSD, but she decides to opt for Nussbaum's approach rather than Sen's procedural capability account as Cenci and Cawthorne. Her choice is motivated by the fact that VSD lacks substantive ethical commitment. While Cenci and Chawthorne solve this issue by proposing a procedural approach, Jacobs argues explicitly for a substantive ethical commitment, which, in her opinion, Nussbaum's theory provides. While Jacobs and the present work emphasize Nussbaum's list, our use is, again, instrumental and not substantive. In a sense, Nussbaum's approach can form, in our opinion, the backbone of a procedure that can inform ethical decision-making and ethics pedagogy, which is also something that Jacobs and others recognize (pp 3369–3370). But when we stay at the level of habituating data scientists to ethical reasoning, we are not bound up to use Nussbaum's approach in a substantive way, as Jacobs does. When scholars like Jacobs say that capabilities, and central capabilities in particular, "should be brought to at least a threshold level (…) to lead a dignified life" (Jacobs, 2020, p 3372), our use of Nussbaum's approach does not address which threshold is necessary: we just say that for the purpose of habituating to ethical reasoning, recognizing that capabilities are connected to leading a dignified life is all that matters.

3.3 A Microethical Approach to the Practice of Data Science

Now we have all the ingredients to specify what a microethics approach (Komesaroff, 1995) to data science and its ethical training is. This approach has several components.

First, there is a general view that the aim of ethical training is to cultivate moral excellences in identifying the ethical dimension of a situation. Ethical training happens along the same lines and in the same context in which data scientists learn the skills necessary to become skilled data scientists—ethics is *not*

external to practice. As a data scientist learns the necessary skills for data science by practicing and by learning from role models, so a data scientist becomes a *good* data scientist by practicing and by learning from role models about the ethical subtleties of the data science profession.

The second component is a description of the moral abilities (or those components of character virtues) virtues data scientists should cultivate. We have focused our attention on moral attention, which is a moral ability for identifying and grasping the ethical relevance of a situation. We have defined a decision as ethically relevant when it has profound impacts on one's ability to realize one's conception of "living well" and "good life." We have made this insight more precise by connecting it to the capability approach: when something is ethically relevant, it impacts one or more of the central capabilities identified by Nussbaum. We use the approach as a way to identify proxies for ethical relevance.

This approach is a microethics because it prioritizes the minutiae of the small, daily, technical acts of data scientists and their ethical relevance. In other words, one cultivates moral attention by systematically asking questions about the relevance of each technical process or operation (e.g., training algorithms, cleaning data sets) for the basic capabilities of data subjects. The approach is not principled because we do not apply general concepts to technical decisions. Rather, the idea is to reflect on how data science tools impact data subjects' ability to exercise substantial freedoms and to shape the data science process accordingly. A rich microethics emerges in the sense that technical mundane decisions and ethical considerations shape one another.

4 Cultivating Moral Attention in the Practice of Data Science

In this section, we show how the heuristic of asking questions about capabilities can be put fruitfully at work in the context of healthcare data science. The goal is to give an idea of how ethically rich and dense data science turns out to be when the ethical relevant aspects of the technical process are identified through our microethical lens. In what follows, we first contextualize our approach in health care (4.1), then we present an idealized framework illustrating the data science process (4.2), and show how seemingly mundane technical decisions can impact significantly various capabilities (4.3). In particular, we will show in detail how in the stages of data understanding and preparation data scientists can use questions about capabilities (and about the factors normally shaping them) as a proxy to embed ethics in the practice of data science and to become familiar with ethical issues in their work. While we will briefly comment on other phases of the data science process, our choice of focusing on data understanding and preparation is motivated by the fact that in the literature this stage has been largely ignored by those seeking to promote a more ethical data science, probably because from a macro-ethical perspective it looks like a stage with little ethical relevance.

4.1 Contextualizing Capabilities in Healthcare

It is important to emphasize that, while we provide a general heuristic, the nitty–gritty of how the heuristics is applied and developed will depend on the specific context. In other words, although Nussbaum's ten capabilities (2006) are broadly relevant for moral attention within any data science project, several of them appear particularly likely to arise in specific domains. For instance, life and health are affected by many data science applications to healthcare; bodily integrity and control over one's environment would arise easily in the legal domain; ability to have concern for other species impacts environmental and agricultural projects; and capability to use one's senses, imagination, and thought is directly affected by education and other cultural projects; and so on. The remainder of the paper will focus on the healthcare domain, and thus, *health* will be particularly relevant. In particular, we will characterize the cultivation of moral attention in this context as a twofold process.

First, data scientists should pay attention to how impairing health can downstream impair other important capabilities. Life and bodily health are certainly important both as functionings and capabilities, because possessing them can have important impacts on the other capabilities. One may also say that they are the necessary condition for having at least the possibility of the substantial freedoms of the other capabilities. In cultivating moral attention, an important component of asking questions about capabilities requires understanding the relation between health as a capability and other capabilities. This is especially because limited health may impact some conversion factors that will likely impact other capabilities as well.

The second part of the process is to cultivate and exercise moral attention to understand health as a capability. According to the capability approach, there is more to health than just health itself. Ruger (2010) dissects the health capability as being composed by *health functioning* and *health agency*. Health agency is an instance of agency freedom—the foundation of the capability approach—and is defined by Ruger as "individuals ability to achieve health goals they value" (2010, p 42). In Ruger's view, health agency requires a number of "conversion factors" that allow subjects to realize whatever valuable health goals they want to realize. At the individual level, conversion factors include health knowledge, health-seeking skills, self-governance, and effective health-decision making (for a complete list, see Ruger, 2010). At the social level, health agency requires external factors, such as social norms, social networks and capital, group membership, material circumstances, access to health services, etc. This is because health agency "is dependent on how one's external environment enhances or detracts from an individual" (Ruger, 2010, p 43). As examined below, some of these internal and external factors affecting individuals can be identified and/or inferred by electronic health records (EHRs). Therefore, cultivating moral attention requires that data scientists pay attention to how the mundane technical decisions obfuscate or hide some of these factors.

4.2 The Data Science Process

To organize the data science process, we use a linearized, stepwise characterization common to introductory textbooks and language of the practice, acknowledging that the process is iterative and often cyclic (see Fig. 1). We organize the linearized process into seven stages grouped into three phases. The early phase of a data science project consists of stages for (1) problem understanding and definition and (2) data acquisition. The middle phase consists of (3) data understanding and preparation, (4) data analysis and modeling, and (5) validation and interpretation of the model. The late phase consists of (6) communication and deploying the results and (7) evaluating feedback on the solution. The middle phase receives the most technical emphasis, while the early and late phases are more dependent upon the domain and broad social context. Defining and understanding the problem may require looking ahead and attempting preliminary solutions, which can drive an iterative or cyclic process, and intrinsic uncertainty may necessitate moving forward through stages anticipating the need to backtrack later. We clarify the detailed process among the numerous ways data science—as a young, rapidly expanding field—is conceived to identify specific places where moral attention adds technical as well as ethical value to data science, and in the following section, we discuss data understanding and preparation in detail in a healthcare context.

4.3 Microethics of Data Understanding and Preparation

As a working example, consider the problem of extracting text from Electronic Health Records (EHR) in order to construct features on Social Determinants of Health (SDoH) useful for prediction, cohorting, and possible intervention for patients with diabetes and/or cardiovascular disease. An appropriate dataset could be the EHR of patients in a national or large US healthcare system, such a hospital, Health Maintenance Organization (HMO), or Medicare Advantage program.

Data understanding is the process of describing and exploring the data and identifying data quality issues to understand its size, quantity, and accuracy. Data descriptions characterize the dimensionality and sparsity of the data records, and data exploration involves summary statistics, visualization, and other methods in an initially unstructured way to better understand the nature of the data set. Data quality problems include missing data, noise, artifacts, outliers, inconsistency, and duplicate data. *Data preparation* consists of cleaning the data followed by transforming and reducing it to prepare for modeling. Data cleaning can be an unanticipated, time-consuming aspect of the process and involves addressing the data quality problems: discarding or imputing missing data, smoothing out noisy data, removing artifacts and outliers, correcting inconsistencies, and removing duplicate data. Data transformation changes the data values, format, or structure in a way more amenable to the problem being addressed and includes normalization, standardization, and feature construction; and data reduction modifies the

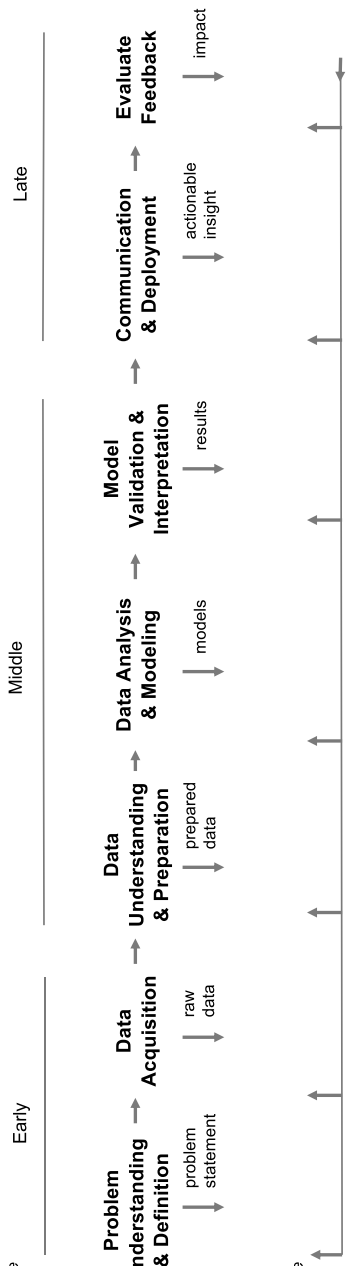

Fig. 1 Stages of the Data Science Process

quantity and/or structure of the data by sampling, selecting features, or applying dimensionality reduction methods, such as principal component analysis.

4.3.1 Data Understanding

Let us start by analyzing the microethics of *data understanding*. This process illustrates well the importance of a micro-ethics based on the capability approach in filling the gaps created by a duty-oriented professional ethic or regulatory compliance. If data fairness were incorporated in the data acquisition stage, and one successfully avoids data uses prohibited by regulations such as HIPAA and GDPR, then from a principled/rule-based perspective, the data understanding and preparation stage is ethically relatively straightforward. One should be sufficiently conscientious about potentially conflicting demands implicit in understanding the data to prepare it accurately and unbiasedly for an anticipated modeling approach in a timely manner, so the data used to generate the model still corresponds closely to the source, and one should be a good steward of corporate and/or client resources and accountable to deadlines. However, most of the time at this stage there are numerous mundane decisions, which may still have significant ethical and other sociotechnical impacts. A micro-ethics approach to data understanding requires not only understanding the data statistically but also in its sociotechnical context. Moral attention can help identify geographic, biomedical, behavioral, and social factors that may influence understanding the structured data, and which incidentally are also "conversion factors" that can impact capabilities (Ruger, 2010). Geographic neighborhoods (such as US zip codes) can be a proxy for a variety of socioeconomic factors, race/ethnicity demographics may indicate known genetic allele difference associated with disease propensity across populations, occupation may suggest environmental toxicities or stressors, and variations in healthcare systems may correlate with quality of care. In other words, the idea is that data scientists should develop an ability to identify some of the factors that can function as "conversion factors" affecting capabilities and hence having ethical relevance.

But what point of the process demands the greatest attention? And to which factors? Moral attention should be exercised especially in the most mundane decisions. For instance, in understanding unstructured EHR data, one might discover that 80% of the data is exported cleanly from an electronic medical records (EMR) system and 20% requires optical character recognition (OCR) of scanned or faxed documents. A technically prudent and accountable decision might be to set aside the potentially very time-consuming OCR documents until later and proceed with the EMR-exported 80%. However, if the OCR documents predominantly come from small clinics in underserved neighborhoods where SDoH factors could significantly differ from those receiving care in the EMR-enabled health system, then project, professional, and ethical requirements would demand at least understanding the OCR records. Therefore, in focusing on EMR data for matters of convenience, *one should ask why certain data sets require OCR rather than EMR processing*. This is a question that, if asked with capabilities (and conversion factors) in mind, can help data scientists to cultivate moral attention. First, let us start with the relation between health agency and the choice of

going through EMR or OCR data. If data sets are in OCR because they come from underserved areas, then one may negatively affect health agency of those who live in that area, because superficially skipping OCR data will mean excluding them from the "big data loop" (Lerman, 2013), with bleak consequences for the healthcare they can possibly access. Being excluded from the loop means that no one will notice the particular conditions or lack of healthcare affecting specific areas, which will become increasingly invisible. Therefore, attention should be paid to some of the conversion factors that shape health agency that Ruger identifies (2010) and that can be found in data sets or inferred from them. For instance, one can check if there are differences in EMR and OCR with respect to economic circumstances, such as, income and employment status, facilities, and resources of neighborhoods, social security of the macrosocial environment, access or presence of barriers to health facilities, and general effectiveness of those facilities. Clearly, conversion factors at the individual level (e.g., self-governance, health-seeking skills) cannot be accessed by data scientists, but external conversion factors identifiable at the social level are less opaque and can be inferred by looking carefully at data sets. These considerations are not important only for health agency per se—impact on a health capability can have downstream consequences also *for other capabilities* (Ruger, 2010). In fact, health as a capability becomes even a conversion factor affecting the other capabilities. Poor or absent quality of care impacts capabilities of *affiliation.* This is because of both the material difficulties of engaging in social interaction in conditions of precarious health, as well as the psychological detrimental effects on our social skills. As a consequence, *control over one's environment* is impacted as well (especially in the material sense meant by Nussbaum). Therefore, even a seemingly neutral and innocent technical decision of spending more time in understanding why some EHR data requires OCR can have substantial ethical relevance upon multiple capabilities.

Another important aspect of data understanding is connected to *missing data*. Ethically relevant factors can affect data missing or appearing as outliers, and thus, technical decisions deserve an additional layer of attention, which is genuinely moral. Including or excluding data can directly affect the capabilities of others. In understanding missing data, one generally distinguishes between three mechanisms of addressing missing data (Rubin, 1976):

1. Missing completely at random (MCAR), which does not affect statistical analysis
2. Missing at random (MAR), which although not actually random, can be statistically treated that way using other known variables
3. Missing not at random (MNAR), which is missing due to reasons the variable is measuring.

Consider an intervention to improve access to nutritional food among patients with diabetes. To measure the effect of the intervention, one might choose participants who regularly have fasting glucose tolerance test results, as changes in those results could indicate whether the intervention affected blood sugar. Even though some participant's glucose tests may not be obviously "missing," if they

have less frequent testing due to other health conditions or unrecognized social factors, such as healthcare access or financial constraints, they could be excluded from the intervention due to the same underlying reasons that reduce their access to nutritious food and precipitate their diabetes, i.e., data MNAR. If data is MAR, one can use statistical methods to address the omission, making sure that no inadvertent inequities are introduced. In deciding that some data are MAR instead of MNAR, one may compare subjects with and without data and establish that the data missing are not such because of conversion factors enabling health agency. In other words, we explore the data set and systematically look for absence or presence of conversion factors enabling (or disabling) health agency. For instance, if those with fewer tests have similar health conditions, healthcare access given those conditions, and socioeconomic contexts of those with more tests, then you can conclude that the unknown reasons for fewer tests lie outside the scope of the intervention because those people excluded appear to have similar relevant capabilities (e.g., bodily health, bodily integrity, access to health care) to those included, and thus are MAR instead of MNAR. This systematizes an aspect of data science otherwise idiosyncratic and is where Ruger's list of conversion factors shaping health agency can be useful to cultivate technical and moral attention: a data scientist will look exactly for those factors in the data set. In order to say that some data are MAR, one wants to ensure the missingness only depends upon known variables which have and/or depend on known capabilities, and that determination requires domain knowledge and has no purely statistical solution. If the data already exists, then it can be addressed statistically; for example, if some healthcare payers or providers incentivize greater or lower number of tests, one can incorporate the payer-provider in analysis or model (after verifying that the healthcare access is not related to transportation or financial factors likely to affect the access to nutritious food). In other cases, the data may not appear as variables in the acquired data, but examining others of Ruger's external factors may suggest acquiring additional variables that would improve the analysis or modeling both technically and ethically. One can hypothesize that it is a matter of healthcare access and consider the factors that would affect such access, such as finances or transportation. Assuming financial data is not directly available, one can use zip code or other area-based proxies; and for transportation, one might calculate distance between the participant and providers who have given the diabetes diagnosis using driving distance and public transportation travel time and then address, e.g., using representative samples across those variables as needed. However, since the study involves nutritional food, then missingness due to access to nutritional food would be *MNAR*, and one could use something like transportation distance to the nearest grocery store with a large fresh food selection or other variables from Food Access Research Atlas (ERS, USDA, 2021) to determine how much that affects the available data. Therefore, moral attention here does not mean only imagining the moral ramifications of not having the data points; moral attention here requires a genuine understanding of the moral dimension of missing data by using domain knowledge to connect data explicitly to external factors influencing health agency. But as in the case of OCR data, concluding that data are MCAR or MAR when in fact they are missing *not* at random

can also have consequences *on other capabilities* other than health. By impacting bodily health, it also has detrimental effects, down the line, for the capabilities of senses, imagination, and thought in the sense of "[b]eing able to have pleasurable experiences and to avoid beneficial pain" (Nussbaum, 2006, p 76). Moreover, affiliation and control over one's material environment are impacted negatively. Therefore, a routine and technical procedure such as accounting for missing data not only can have ethical ramifications, but it requires the exercise of moral attention and the application of the heuristic of the capability approach in order to make the right technical decision. Some of these factors might also be discovered by attending closely to social or political factors, but moral attention to capabilities also orients how one uses those factors to maintain health agency, and thus the moral principle of autonomy.

4.3.2 Data Preparation

In preparing data, one engages in technical decisions concerning data cleaning, transformation, and reduction. *Data cleaning* involves addressing the data quality problems discovered in data understanding, such as the missing data previously discussed. *Data transformation* changes the values, format, or structure of data and provides ample opportunities to expand or limit the capabilities of others. One often needs to normalize, scale, standardize, or bin the data for analysis and visualization, which can obscure heterogeneity. For example, binning body mass index (BMI) may simplify analysis for predicting diabetes or cardiovascular disease, but incorporating sex and ethnic differences in creating BMI bins can yield better predictive models and thus improve access to appropriate interventions.[16]

One may also need to split or combine features, and feature engineering, in particular, lends itself to moral attention of capabilities. *Feature engineering* transforms the raw data into features that better represent the data for modeling. One of the benefits of a practiced or exemplary moral attention would be the ability to engineer features that correspond well to some of Ruger's external factors that shapes health agency, functioning and, down the line, other important capabilities. Typically, one might need a variety of measures to ascertain the nuances of some of the external factors, and considering the existing features in those terms can add insight into data understanding. One may also want to split features based upon those factors affecting the health capability. For example, distinguishing smart phone access from other mobile phone access can indicate significant differences in access to healthcare information. Within feature engineering, thinking about the capabilities upon which different possible features might depend can increase awareness of the broader context. Although access to healthcare information should be identified in SDoH analysis based solely on domain knowledge, it can arise through moral attention

[16] Although differences in sex and ethnicity may not drive different medical treatments, taking them into account still may result in better predictive models, as they may serve as a proxy for social and cultural factors (Laxy et al., 2018) or help account for demographic differences in self-report accuracy (Richmond et al., 2015).

elsewhere, e.g., the ability to "convert" awareness of body changes to knowledge of medical "symptoms" and then to a timely decision to seek medical care.

Transformation for unstructured data (e.g., texts) often involves tokenization and possible normalization (e.g., spell correction or lemmatization) and mapping to standardized medical language (ontologies). Text data may have additional nuance not captured in structured data and provides additional opportunities to create features oriented toward capturing those factors that can influence health agency. For example, patient interviews can be searched for terms related to social support, even if some specific questions about that were not explicitly asked. Although patient medical and behavioral histories and other reports have information essential for determining SDoH, differences in external factors between patients may also affect the text and the extraction of relevant factors. For example, non-native language speakers may use a more limited vocabulary; those with perceived or actual healthcare treatment disparities due to gender and ethnicity may have less detailed (or filtered) reports; staff in less rewarding healthcare systems may have more spelling errors due to apathy, poorer education, or more stressful workloads; and patients may omit relevant factors due to embarrassment or to appear dutiful or compliant. For the data scientist, asking explicit questions on the external factors influencing health agency opens up the possibility for identifying omitted statements, such as a no mention of social support in an extended interview. In the overall data science process, interpreting or evaluating the model may suggest returning to the data transformation step and revising transformations to improve the representation of patient capabilities.

During *data reduction*, sampling can exacerbate or address issues in representation; e.g., oversampling underrepresented groups can improve model performance across the targeted population compared to the population of the acquired data. In *feature selection*, one may attend to what external factors, if not other capabilities other than health, each feature measures: income indicates increased purchasing power; zip codes can reflect environmental stressors, healthcare access, and other qualities of life; mobility restrictions can affect transportation options; and the presence of expensive laboratory tests not mandated by a clear diagnostic indication may suggest greater access to healthcare. The goal of a microethical approach to moral attention is to see through the data to the social conversion factors affecting health agency and the capability of health sufficiently to enact at least the principle of non-maleficence. *Dimension reduction* (e.g., principal component analysis) can have similar issues as feature engineering, but with greater obfuscation and less explainability and transparency, and any required standardizing of features can mask variations related to unrecognized capabilities.

After the data understanding and preparation stage, we can look back at the prior stages to see if revisions are necessary. In the problem statement, the interventions were implied to be beneficial medical interventions, but the same processes could be used to identify individuals with high medical cost for possible cost reduction interventions. In those cases, the technically good data science practices would more readily diverge from the morally good ones, though moral attention may help minimize harm or suggest a creative solution to meet economic and moral needs. More subtly, even clearly beneficial interventions may be constrained for a cohort

of patients by cost or number of available healthcare workers, in which case moral attention may identify multiple courses of action to be considered depending upon how the ethical tradeoffs are weighed. Varying emphasis on biomedical ethical principles, such as minimizing non-maleficence or maximizing beneficence could respectively lead to a large number of minimal interventions or a smaller number of more effective interventions, and characterizing possible outcomes depends upon synthesizing specific information in the cohort only accessible via data science methods. Looking ahead in the data science process to data analysis and modeling also orients data understanding and preparation toward different possible ends, and highlights some challenges to moral attention caused by the frequent use of newly developed, complex modeling methods (such as recently, multimodal, deep learning) in highly-specialized, novel domains (such as, new CRISPR techniques for intervention). In these situations, the data scientist may be among the first to even consider such applications, so no prior ethical rules would exist. In these cases, moral action may require habituated practice at moral attention to evaluate the ethical tradeoffs among the extensive novel technical choices.

5 Microethics and Participatory Design

So far, we have described the dense microethics emerging from the ethical training to develop moral attention shaped by the capability approach. However, one can raise one issue against the project as a whole. Is this microethical training even realistic? In order for the data scientist to be able to develop moral attention more effectively in this context, he/she has to know a lot about SDoHs, and in general about the social and political dimension of healthcare. This is an important criticism that allows us to contextualize the work of data scientists in a wider process that, we think, should characterize the construction of algorithmic systems that also contribute to sociotechnical systems.

Indeed, we do think that data scientists should also have a preliminary and basic training in the particular field or discipline to which their work is applied. It is important that data scientists working with biomedical data have basic knowledge of the social dimension of the health care aspects shaped with their work. If working in education, then the data scientist needs to be initiated to basic notions of how a school system works, what do teachers do, the struggle they encounter, etc. Moreover, we must emphasize that data science should not be considered as extremely portable, in the sense that algorithmic systems should be designed with knowledge of the environments in which they will work, and not by abstracting too much from those environments because of technical convenience. And in order to do that, data scientists are required to know something more than just coding or statistics.

However, we should also be clear on another aspect. Data scientists cannot know everything, and also they may not be in a position to decide how to weigh the different ethical considerations that may emerge via the cultivation of moral attention. Data scientists should be humble, and recognize that they cannot design algorithmic systems only by themselves. This consideration has an interesting consequence, namely, that the ethical training we describe is just a first step towards ethical

algorithmic systems. Because algorithmic systems are going to shape capabilities of a number of individuals, relevant stakeholders should participate in a deliberation on how exactly those systems are going to be designed, which is a consideration already emerged when discussing Cenci and Cawthorne (2020) in Section 3.2. Moreover, in addition to stakeholders, social scientists should participate as well, because even if data scientists may have some background knowledge of, say, how healthcare systems work and how they impact individuals, a much more robust expertise is needed to appreciate the subtle ways in which capabilities are shaped.[17] Proposals to see algorithmic systems "as being embedded in a larger context of institutional or organizational norms and standards that safeguard the interests and goods of those it serves" (von Eschenbach, 2021, p 14) go in this direction. But cultivating moral attention is nonetheless fundamental, because by developing a moral sensibility, data scientists can readily connect ethical issues with technical choices in the ideal deliberative process between stakeholders and social scientists that we have just sketched. For instance, the social scientist may emphasize some important aspects of SDoHs, but it is the data scientist that is more likely to make the direct connection between those SDoHs and some seemingly technical and neutral aspects of the data science process that the social scientist may overlook. In other words, the data scientist must be ready to understand how ethical/social issues and technical aspects shape one another, and developing moral attention will facilitate this process. This is surely a topic for another article, but we wish to highlight the place of the ethical training we have described here in a much bigger process of algorithmic training.

6 Conclusion

In this paper, we have envisioned the nature of ethical training in data science as an exercise aimed at progressively cultivating some abilities. These moral abilities will be learnt in the same way the technical skills necessary in the data science context are learnt. We have grounded our microethics approach in virtue ethics. The approach is "micro" in the sense that it aims to identify the ethical relevance of every single mundane technical choice of data scientists, rather than merely understanding if a technical process is compliant with general ethical principles. In order to be able to identify the ethical relevance of technical choices, we have used the capability approach as heuristic that can help data scientists to familiarize with ethical problems raised by their tools.

This article is only an introduction to our microethical approach. We will need to elaborate a much more systematic pipeline that embed specific questions about capabilities (or at least more specific themes) in the different phases of the data science process. Depending on the context of the data science process, one will have to make extensive research on conversion factors affecting capabilities in that context. What we have accomplished here is just illustrative: after formulating our reasons in

[17] On the roles that social sciences can play in this kind of situations, see Lohse and Canali (2021).

favor of our approach, in Section 4 we have shown how rich a microethics can be, and how ethics in general can be really hard and intricate.

Funding Open access funding provided by Johannes Kepler University Linz.

References

Alkire, S. (2005). Why the capability approach? *Journal of Human Development, 6*(1), 115–135. https://doi.org/10.1080/14649880520003427 5

Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12, 251–261.

Annas, J. (2011). *Intelligent Virtue*. Oxford University Press.

Aristotle. (2014). *Nicomachean Ethics* (C.D. Reeve, ed.). Hackett Publishing Company.

Bezuidenhout, L., & Ratti, E. (2020). What does it mean to embed ethics in data science? An integrative approach based on the microethics and virtues. *AI and Society*, 36, 939–953. https://doi.org/10.1007/s00146-020-01112-w

Bezuidenhout, L., Ratti, E., Warne, N., & Beeler, D. (2019). Docility as a Primary Virtue in Scientific Research. *Minerva*, *57*(1), 67–84. https://doi.org/10.1007/s11024-018-9356-2

Binns, R. (2018). *Fairness in Machine Learning: Lessons from Political Philosophy,* 1–11. http://arxiv.org/abs/1712.03586

Claassen, R. (2014). Capability paternalism. In *Economics and Philosophy 30*(1), 57–73. Cambridge University Press. https://doi.org/10.1017/S0266267114000042

Cenci, A., & Cawthorne, D. (2020). Refining value sensitive design: A (capability-based) procedural ethics approach to technological design for well-being. *Science and Engineering Ethics, 26*(5), 2629–2662. https://doi.org/10.1007/s11948-020-00223-3

Coeckelbergh, M. (2010). Health care, capabilities, and AI assistive technologies. *Ethical Theory and Moral Practice, 13*(2), 181–190. https://doi.org/10.1007/s10677-009-9186-2

Coeckelbergh, M. (2011). Human development or human enhancement? A methodological reflection on capabilities and the evaluation of information technologies. *Ethics and Information Technology, 13*(2), 81–92. https://doi.org/10.1007/s10676-010-9231-9

Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *In Ethics and Information Technology, 20*(1), 1–3.

ERS/USDA (2021). Economic Research Service, US Department of Agriculture. *Food Access Research Atlas*. https://www.ers.usda.gov/data-products/food-access-research-atlas/

Fazelpour, S., & Lipton, Z. C. (2020). Algorithmic fairness from a non-ideal perspective. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 57–63*,. https://doi.org/10.1145/3375627.3375828

Floridi, L. (2018). Soft ethics and the governance of the digital. *Philosophy and Technology, 31*(1), 1–8. https://doi.org/10.1007/s13347-018-0303-9

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*, 1–13. https://doi.org/10.1162/99608f92.8cd550d1

Gogoll, J., Zuber, N., Kacianka, S., Greger, T., Pretschner, A., & Nida-Rümelin, J. (2021). Ethics in the software development process: From codes of conduct to ethical deliberation. *Philosophy and Technology*. https://doi.org/10.1007/s13347-021-00451-w

Hagendorff, T. (2020a). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*,. https://doi.org/10.1007/s11023-020-09517-8

Hagendorff, T. (2020b). *AI virtues -- The missing link in putting AI ethics into practice*. http://arxiv.org/abs/2011.12750

Oosterlaken, I. (2015). Human capabilities in design for values: A capability approach of "design for values." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains,* 221–250. Springer Netherlands. https://doi.org/10.1007/978-94-007-6970-0_7

Jacobs, N. (2020). Capability sensitive design for health and wellbeing technologies. *Science and Engineering Ethics, 26*(6), 3363–3391. https://doi.org/10.1007/s11948-020-00275-5

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kelly, T. (2018). *Professional ethics - a trust-based Approach*. Lexington Books.

Komesaroff, P. (1995). From bioethics to microethics: ethical debate and clinical medicine. In P. Komesaroff (Ed.), *Troubled Bodies - Critical Perspectives on Postmodernism, Medical Ethics, and the Body*. Duke University Press.

Laxy, M., Teuner, C., Holle, R., & Kurz, C. (2018). The association between BMI and health-related quality of life in the US population: Sex, age and ethnicity matters. *International Journal of Obesity, 42*(3), 318–326. https://doi.org/10.1038/ijo.2017.252

Lerman, J. (2013). Big Data and its exclusions. *Stanford Law Review, 66*, 55.

Lohse, S., Canali, S.. (2021). Follow the science? On the marginal role of the social sciences in the COVID-19 pandemic. *European Journal for Philosophy of Science*

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? *ESEC/FSE 2018 - Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 729–733. https://doi.org/10.1145/3236024.3264833

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 501–507.

Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. "From what to how. an overview of ai ethics tools, methods and research to translate principles into practices." *Science and Engineering Ethics*, no. 0123456789. Springer Netherlands. https://doi.org/10.1007/s11948-019-00165-5.

Nussbaum, M. (2006). *Frontiers of justice - disability, nationality, species membership*. Harvard University Press.

Nussbaum, M. C. (2011). *Creating capabilities: The Human Development Approach*. Harvard University Press.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Richmond, T. K., Thurston, I., Sonneville, K., Milliren, C. E., Walls, C. E., & Austin, S. B. (2015). Racial/ethnic differences in accuracy of body mass index reporting in a diverse cohort of young adults. *International Journal of Obesity, 39*(3), 546–548. https://doi.org/10.1038/ijo.2014.147

Robeyns, I. (2016). Capabilitarianism. *Journal of Human Development and Capabilities, 17*(3), 397–414. https://doi.org/10.1080/19452829.2016.1145631

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592.

Ruger, J. P. (2010). Health capability: Conceptualization and operationalization. *American Journal of Public Health, 100*(1), 41–49. https://doi.org/10.2105/AJPH.2008.143651

Russell, D. (2015). Aristotle on cultivating virtue. In N. Snow (Ed.), *Cultivating Virtue - Perspective from Philosophy, Theology, and Psychology* (pp. 17–48). Oxford University Press.

Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: A systematic literature review and proposed project framework. *Ethics and Information Technology, 21*(3), 197–208. https://doi.org/10.1007/s10676-019-09502-5

Sen, A. (1985). *Commodities and Capabilities*. North-Holland.

Sen, A. (1999). *Development as Freedom*. Anchor Books.

Susser, Daniel, Beate Roessler, and Helen Nissenbaum. (2019). Technology, autonomy, and manipulation. *Internet Policy Review* 8 (2), 1–22. https://doi.org/10.14763/2019.2.1410.

Vallor, S. (2016). *Technology and the virtues - a philosophical guide to a future worth wanting*. Oxford University Press.
van Wynsberghe, A., & Robbins, S. (2019). Critiquing the reasons for making artificial moral agents. *Science and Engineering Ethics,* 25, 719–735. https://doi.org/10.1007/s11948-018-0030-8
Vold, K., & Whittlestone, J. (2019). Privacy, autonomy, and personalised targeting: rethinking how personal data is used. In C. Véliz (Ed.), *Report on Data, Privacy, and the Individual in the Digital Age*.
von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy and Technology*. https://doi.org/10.1007/s13347-021-00477-0
Walzer, M. (2006). *Just and Unjust War*. Fourth Edition, Basic Books.
Zwolinski, M., & Schmidtz, D. (2013). Environmental Virtue Ethics. In D. Russell (Ed.), *The Cambridge Companion to Virtue Ethics,* 221–239. Cambridge University Press.

Authors and Affiliations

Emanuele Ratti[1] · **Mark Graves**[2]

[1] Institut für Philosophie und Wissenschaftstheorie, Johannes Kepler University Linz, Linz, Austria

[2] Parexel AI Labs, San Francisco, CA, USA