AI & SOCIETY
https://doi.org/10.1007/s00146-020-01112-w

OPEN FORUM

What does it mean to embed ethics in data science? An integrative approach based on microethics and virtues

Louise Bezuidenhout[1] · **Emanuele Ratti**[2]

Received: 2 July 2020 / Accepted: 4 November 2020

Abstract
In the past few years, scholars have been questioning whether the current approach in data ethics based on the higher level case studies and general principles is effective. In particular, some have been complaining that such an approach to ethics is difficult to be applied and to be taught in the context of data science. In response to these concerns, there have been discussions about how ethics should be "embedded" in the practice of data science, in the sense of showing how ethical issues emerge in small technical choices made by data scientists in their day-to-day activities, and how such an approach can be used to teach data ethics. However, a precise description of how such proposals have to be theoretically conceived and could be operationalized has been lacking. In this article, we propose a full-fledged characterization of 'embedding' ethics, and how this can be applied especially to the problem of teaching data science ethics. Using the emerging model of 'microethics', we propose a way of teaching daily responsibility in digital activities that is connected to (and draws from) the higher level ethical challenges discussed in digital/data ethics. We ground this microethical approach into a virtue theory framework, by stressing that the goal of a microethics is to foster the cultivation of moral virtues. After delineating this approach of embedding ethics in theoretical detail, this article discusses a concrete example of how such a 'micro-virtue ethics' approach could be practically taught to data science students.

Keywords Data science · Microethics · Virtue ethics · Teaching ethics · Embedded ethics

1 Introduction

As our world becomes increasingly digital, we are starting to ask questions about how we, as a society, construct, utilize and live in a digital world. In addition to emerging patterns of online social behavior, it is recognized that the design of digital infrastructures and algorithms, and the use of data pose can pose serious ethical challenges (Williams et al. 2018; Zliobaite 2017; Dressel and Farid 2018).

These ethical discussions have coalesced under the loose heading "digital/data ethics". In recent years, a wide range of courses have been developed to educate researchers, as well as the public, about the ethics of AI, machine learning, algorithm design, and digital behavior.[1] Many of these courses focus on the broader challenges of accountability, privacy, and fairness posed by the emerging digital landscape.

The dominance of a "big picture" focus in digital/data ethics instruction is important for raising awareness of the nascent trends and problems. Nonetheless, there are growing reservations as to how effective such courses are for preparing data science practitioners to work ethically in their daily activities within these digital environments (Grosz et al. 2019). These reservations are linked to a broader discussion within ethics pedagogy about the limits of higher level case studies as a tool for ethics instruction (Troug et al. 2015). Such concerns highlight that these case studies can leave students without a clear understanding of individual responsibility and ethical daily practice (Chen Forthcoming).

Louise Bezuidenhout and Emanuele Ratti have contributed equally.

✉ Emanuele Ratti
mnl.ratti@gmail.com

[1] Institute for Science, Innovation, and Society, University of Oxford, Oxford, UK

[2] Institute of Philosophy and Scientific Method, Johannes Kepler University Linz, Linz, Austria

[1] A list of courses offered around the world can be found here https://docs.google.com/spreadsheets/d/1jWIrA8jHz5fYAW4h9CkUD8gKS5V98PDJDymRf8d9vKI/edit#gid=0. Accessed March 22nd, 2020.

Published online: 08 December 2020

Taking such concerns onboard, some have proposed that we should find ways to align ethics closer to the practice of data science. This idea is sometimes expressed as 'embedding ethics', and it has been proposed both in teaching as well as in research and development (Grosz et al. 2019; McLennan et al. 2020). However, while there are some examples of curricula following this intuition, a full-fledged account of what are the foundations of embedding ethics, to our knowledge, is missing. Moreover, while these attempts to embed ethics are commendable, it is not clear whether these approaches are mutually exclusive with approaches focusing on the 'bigger picture'. This paper proposes a novel framework articulating the idea of 'embedding ethics', and from this framework it derives a pedagogical approach that connects the 'big picture' of digital/data ethics to the routines of daily practice. Using the emerging model of "microethics" (Komesaroff 1995; Truog et al. 2015; Hagendorff 2020), we propose a way of teaching daily responsibility in digital activities that is connected to (and draws from) the higher level ethical challenges discussed in digital/data ethics. We ground this microethical approach into virtue theory, by stressing that the goal of a microethics is to foster the cultivation of moral virtues. Our approach is 'neutral' with respect to the actual ethical content—we aim at formulating a methodology. In order to show more concretely how the exact ethical content can be included in specific cases, the paper then goes on to discuss some examples of how such a "micro-virtue ethics" approach could be practically taught to data science students. The paper concludes with a brief discussion on how such an approach may be expanded beyond data science to teach responsible and responsive digital citizenship to more general audiences.

Before starting, we hasten to add that this article is addressed to those who teach data ethics to data science students and/or data scientists training in professional contexts. While we think that, with minor revisions, the framework we develop can be used also in more traditional computer science courses, we have formulated our views with the concerns of data science curricula in mind. Our framework can also be used to embed ethics in actual data science research (and both authors are working on this aspect), but this will require additions to contextualize it within the specific contexts in which data scientists end up working. This is not a minor point: additions may be different depending on the context, which, in the case of data scientists, may range from medicine to retail.

2 Digital ethics as a macroethics and its problems

A growing amount of scholarship examining the online environment has highlighted a range of social and ethical challenges emerging from digital spaces. These include individual misbehaviors caused by erosion of empathy, promotion of narcissistic behavior, internet addiction, etc. (Vallor 2016). What such issues illustrate is the urgent need for a comprehensive understanding of socially responsible behavior that enables individuals to function and flourish online, while protecting the individuals, structures and systems around them. In particular, there is a need for a robust interpretation of "digital citizenship" that takes into account the novelties of the online environment as compared to traditional spheres of action. What is needed are better descriptions of how individuals can act responsibly in spaces that are removed from traditional societal structures and mechanisms of control.

Perhaps even more challenging to formulations of digital citizenship, however, is the recognition that the online environment is highly dynamic. The same people who are users of digital structures are also contributing to their evolution. Indeed, the emergence of tools such as machine learning mean that user behavior dynamically changes the tool in question through feedback and evolution. A considerable amount of research already details these problems, such as those examining the operation of search engines and the perpetuation of biases (Bozdag 2013). Moreover, examples such as Cambridge Analytica highlight how these processes can be used to manipulate the behavior of users (Zuboff 2015; Susser et al. 2019). A number of scholars have already started trying to bring together these disparate ethical considerations into comprehensive narratives. Floridi (2018), for example, distinguishes between digital governance, digital regulation, and digital ethics. Digital governance relates to the procedures and practices for establishing and implementing policies, as well as the creation of codes of conducts and practice, while digital regulation refers to the evolving system of rules and laws enforced through social and governmental institutions. Digital ethics plays a role in both. According to Floridi, it can play a significant role in shaping both governance and regulation by providing guidance on principles that fosters more just digital environments that align with features of 'the good society'. Here we will focus especially on the part of digital ethics that deals with data science—what has been called *data ethics*.

2.1 Data ethics

In large part, discussions on responsible individual behavior online within data ethics have focused on understanding the ethical outcomes of data use and the design of algorithms. Approaches vary from utilitarian discussions on the societal impact of algorithm design to the deontological development of principles to guide action. The latter in particular has attempted to address ethical and societal issues connected to the digital environment through the formulation of ethical principles to guide the innovation and the use of digital

tools. Websites such as Algorithm Watch offer an (almost) up-to-date list of initiatives proposing frameworks or principles.[2] The proliferation of scholarship on aspirational, guidance and enforceable codes of conduct has been welcomed as a positive contribution to AI regulation and governance (and data science in particular).

In an attempt to focalize data ethics discussions, Floridi and Cowls (2019) have proposed that the number of ethical principles in use should be reduced. They suggest that the identification of a set of common principles will inform ongoing attempts of digital governance and regulation, and it can constrain the ability of corporations to embrace expedient relativism in their interpretations of ethics. In particular, they identify five principles that seem to be common to many relevant initiatives. These are beneficence, nonmaleficence, autonomy, justice, and explicability (which includes intelligibility and accountability). This, they acknowledge, aligns AI ethics (and as a result also data ethics) with the principlist approach in biomedical ethics (Beauchamp and Childress 2009), and less with the rich tradition of ethics of technology and computer/information ethics. The ethics discourse around the AI revolution (including data science) is thus emerging with a specific character. It is increasingly aiming to deliver an abstract and general evaluation of what is right and wrong, and to identify common shared principles that loosely guide grand projects of regulation and governance, as well as individual behavior.

This move towards principlism is not without its critics (Mittelstadt 2019; Whittlestone et al. 2019), and an increasing number of scholars are raising concerns. Given that the principlist approach has been developed in the medical context, its content has been shaped along those lines. Some criticisms are geared especially towards the shape of this particular content: data scientists are not physicians, and the ethical content of principlism may not be adequate to properly cover the issues emerged in the data science context. However, we are more interested in other issues which are connected to two key areas: the level at which the discourse is situated ("applicability") and the problems associated with pedagogy ("teachability").

2.2 Applicability

Digital ethics—and data ethics is no exception—is currently dominated by what has come to be called *macroethics or hard ethics* (Floridi 2018). This approach attempts to integrate the disparate areas of infrastructure design, deployment, and the use by taking a broad view of the online environment. This approach links to the growing number of centers and courses focusing on internet and society. These centers (and the courses that they offer) focus on internet studies, intersecting with key fields like human–computer interaction and science, and technology studies.

The scope covered by macroethics, together with its alignment with the social studies of digital environments/cultures, can make it difficult to locate the individual within ethics discussions. Indeed, how individual responsibility plays out in spaces in which disparate technologies, platforms, stakeholders, practices and discourses are co-evolving is extremely complex. As a result, much of macroethics discourse focuses on key themes, such as identity and subjectivity, social exclusion and inequality, politics and democracy, globalization and development, privacy and surveillance.

In discussing these themes, macroethics often uses higher level case studies from thematic areas, such as social media, big data, citizen journalism, digital culture, the creative industries, internet governance, and digital rights. These include examples of clear-cut ethics violations, such as the controversy surrounding Cambridge Analytica's involvement in the US elections (Susser et al. 2019). They also include examples of multifaceted, multistakeholder problems, such as the integration of algorithmic bias in search engines (Bozdag 2013). These case studies are variously presented using both deontological and utilitarian ethics, but are united through their focus on the higher level outcomes and the impact of these outcomes on society. Rarely, if ever, do they specifically focus on individual actions, collaborative negotiations and decision-making practices.

The use of high-level case studies thus presents various problems. First, while the principlist approach implicit in the use of high-level case studies works well for analyzing these large issues, understanding them from an individual perspective is more difficult. Many of these case studies either do not describe individual action (focusing on companies, multi/national structures), describe intentionally maleficence actions, or reduce individual action to yes/no decisions (i.e., to use or not use a platform). The nebulous position of the individual within these issues, and the reliance on higher level principles, thus reduces discussion on individual ethics and agency to a reduced range of positions. These can be detailed as follows in Table 1:

Moreover, while individuals are able to engage with the case studies and discuss the ethical implications in general, the link between these ethics and their personal experiences and daily activities is far from certain. Indeed, most digital activity is repetitive and relatively mundane, and users unlikely to be engaged with the action spaces in which most of these case studies play out.

As a result, macroethics discussions often limit individual responsibility to the avoidance of obviously unethical behavior, such as theft, harm, violation of privacy. This leaves the responsibility—and agency—for the ethical issues described

[2] https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/. Accessed January 13, 2020.

Table 1 Challenges of using macro-level case studies

Type of Case studies	Position of individual	Possible student response
Multi-stakeholder case study	No description of individual action	"I could not influence such a situation"
Evidence of maleficent actions	Individual actor perpetrating harm	"I'd never do that"
Outside of field of reference	Reduced range of actor decisions	"I'd never be in a position to do that/ the right course of action"

in the case studies to large corporations and governments, as they fight for control over algorithms, data distribution and re-use. Thus, while the individual user is recognized to be a contributor to the dynamic digital evolution, there is little guidance on how they can influence their immediate online environment towards more ethical futures. In other words, macroethics provides few hints on how to apply ethical principles in concrete situations.

These problems have been noted in the literature. For instance, Morley et al. (2019) argue that, while macroethics gives a justification of "why" individuals should be concerned about AI ethics (and hence data ethics), it does not provide an easy pathway from "why" to "how" they should be engaged. Floridi recognizes this problem of applicability, by stressing that it is "not just *what* ethics is needed but also *how* ethics can be effectively applied and implemented in order to make a positive difference" (2019, p. 185). Nonetheless, as highlighted again by Morley et al. (2019), "[t]he gap between principles and practice is large" (p. 7), since efforts in data ethics do not specify to practitioners where and how the principles should be implemented exactly. This is a problem that also hampers codes of conduct—shaped in a principled way—in the computational sciences with the result of being ineffective in practice (McNamara et al. 2018). When one attempts to applies those principles in specific contexts, what emerges is that much of the macroethical work on data ethics "has been completed in the *abstract*, independent of concrete cases" (Kitto and Knight 2019, p. 2856).

Similar voices of concern come from Haggendorff (2020), who claims that "[u]ltimately, it is a major problem to deduce concrete technological implementations from the very abstract ethical values and principles". Madaio et al. (2020) add that "the abstract nature of AI ethics principles *[including data ethics]* makes them difficult for practitioners to operationalize" (p. 1). On a related note, Vakkuri et al. (2020) claim that "[d]evelopers struggle to implement abstract ethical guidelines into the development process" (p. 1). The problem of 'deducing concrete technological implementations from principles' or 'operationalizing principles' has two parts. First, principles are not rules, which are precise and neat. As Zwolinski and Schmidtz say "[w]here rules function in our reasoning like trump cards, principles function like weights" (2013, p. 222). They can be weighed one against the other, in the sense "principles can weigh against X without categorically ruling out X" (p. 222), and "[q]uestions of weight and priority must be assessed in specific contexts" (Beauchamp 2015, p. 406). Yet, people expect principles to be like rules. The second part of the problem is that those principles can be understood in radically different ways, sometimes mutually exclusive. This creates confusion in understanding which version of the principles we should apply (Binns 2018). These issues have motivated new proposals aimed at 'embedding ethics' in the practice of data science (Grosz et al. 2019; McLennan et al. 2020). The idea behind this intuition is that we should find ways to move ethics closer to the actual practice of data science, so that data scientists will be able to visualize what part of their job has ethical relevance.

A final set of issues associated with macroethics that exacerbate the applicability problem relates to its scope. The focus on general principles means that it rarely engages with the diversity of roles that individuals play within the digital landscape (e.g., data producer, data engineer, data analyst, machine learning engineer, general user). The diversity of the digital landscape itself makes it difficult to translate the macroethical concerns into rules (the "how") that apply "across the board" to daily individual activities. Similarly, it does not respond to recent sociotechnical scholarship on digital landscapes. It is therefore ill placed to address questions of landscape boundaries, such as whether it includes the data, the technical infrastructure, the companies operating online, the online communities, etc. Related questions of whether the digital landscape is solely located online, or whether it extends to the physical world through its interconnectedness with sociotechnical landscapes.

2.3 Teachability

The gap between principles and concrete technological implementations has consequences on the *teachability* of macroethics to students or training of professionals. If there is not a connection between individual technical choices and ethical relevance (i.e., if ethics is not embedded in the actual practice of data science), then it is difficult to deliver modules that shows the relevance of ethics for the tasks of data scientists. The difficulties of teaching macroethics are even more evident when considered from the point of view of its strong links to other dominant pedagogical strategies

within bioethics, namely biomedical ethics and Responsible Conduct of Research (RCR), which seem to suffer from analogous problems. Indeed, data management, data sharing and responsible online behavior are often incorporated into RCR teaching in universities across the globe.

When considered in light of the problems of developing an individual ethics that accounts for daily actions (as outlined in Sects. 2.1, 2.2), it is unsurprising that this *bioethicization/RCRization* can be viewed as problematic. Biomedical ethics, in particular, has been heavily criticized for its reliance on extreme and unrealistic moral dilemmas and famous controversies where the application of principles is more straightforward. This has led to concerns that the full spectrum of ethical nuances encountered in the medical profession are unlikely to be fully addressed. Komesaroff (1995), for example, suggested that the structure of prevalent bioethical discourse constrained the way topics are taught, most notably in the form of a dilemma. Ethical issues are positioned within a demarcated theoretical field that postulates choices from a range of pre-established possibilities, with clear attractive and unattractive connotations. This, in turn, restricts the scope of its subjects, by emphasizing topics more prone to be expressed in the form of "extreme dilemmas" such as euthanasia, autonomy and paternalism. Truog et al. (2015) emphasized that most educators largely rely on a case-based method for teaching ethics, and that these case studies tend "to focus on extreme or unusual situations [and] controversies that generate media attention" (p. 11). This focus is not helpful in educating medical students to identify other subtle and highly contextualized ethical issues. It is precisely this lack of contextual guidance that Komesaroff laments when he suggests that medical ethics ignores the subtle nature of doctor–patient interaction, its social context, and all ethical issues underneath this endless negotiation. Multimodal communication, such as the choice of words, inflexions and gestures, all have ethical relevance in shaping the doctor–patient relation but are largely ignored by most bioethics training (Komesaroff 1995). RCR trainings have been plagued by similar problems (Chen 2020). There are growing concerns that the vocational nature of RCR has been replaced educational approaches that foster rule following, compliance and avoidance of recognized misbehaviors rather than aspiring to excellence.

Data ethics modules do not have yet a precise identity, but a list of courses discussing ethical issues related to AI and data science[3] shows that many courses are shaped along the lines of the characteristics of bioethics/RCR that we outlined above. Teaching courses in this way reflects a macroethical approach that simply imports in pedagogy the same issues of applicability outlined above (McNamara et al. 2018; Madaio et al. 2020; Vakkuri et al. 2020).

However, if data ethics has yet to find its identity in terms of pedagogical strategies, we want to avoid that it inherits all these problems. In what follows, we will focus on the issues around teachability, especially in the context of teaching data ethics to students in data science, by proposing a new approach based on the integration of microethics within a virtue ethics framework. Integrating microethics and virtue ethics, we argue, provides solid foundations for embedding ethics in the practice and in the teaching of data science.

2.4 The need for a new approach

While macroethics provides an important perspective on the "big picture" of digital evolution, it thus struggles to address the questions that affect individuals in their daily activities. It would seem that what is needed instead is a way of fostering mindfulness, social responsibility and care that directly relates to the daily engagement of individuals with the digital landscape. In the rest of this paper, we develop such an approach by focusing on data scientists, and in particular our target are students in data science curricula. However, our approach can be extended to data science professionals or researchers with little adjustments. We make use of "data scientist" broadly to refer to any individual with a level of computing/programming expertise whose daily activities involve working with data analysis or processing. These data scientists work in a wide range of disciplines, institutions, and make use of a plethora of different data types. Nonetheless, they are united by the scope and focus of their daily actions and the types of computational tools they use.

Before proceeding, it is important to note that our criticism of macroethics relates to its use as the *sole* means of ethics instruction. There is undoubted value in using macroethics case studies as a means of outlining the ethics of the emerging digital landscape. Nonetheless, as a means of teaching responsible daily research conduct to data scientist, we believe that macroethics needs to be blended with another approach that highlights the ethical import of daily actions.

3 Microethics and virtue ethics

Data ethics instructors have to address a number of different challenges. Lesson content must be contextually/content appropriate for different actors. It must provide students with an understanding that they have agency within this complicated landscape to act in responsible ways, and the ability to effect positive change in the individuals and infrastructures around them. It must make students aware of the high-level

[3] https://docs.google.com/spreadsheets/d/1jWIrA8jHz5fYAW4h9CkUD8gKS5V98PDJDymRf8d9vKI/edit#gid=0. Accessed September 12th, 2020.

discussions that are informing regulation, governance and investment provide important framing tools, but not rely on high-level case studies for instruction. Indeed, any case study employed must demonstrate to students how to "grappl[e] with ethical questions as they arise in the daily course of social life" (Komesaroff 1995, p. 65), and to identify pathways of ethical practice that foster responsible conduct and moral development.

This seems overwhelming. However, if one examines normal digital behavior—online activity, coding, data management—a common element becomes apparent. Digital activity comprises discrete and repetitive actions that cumulatively produce effect. The following sections demonstrate how this characteristic can be turned to good pedagogical effect. An approach to embedding ethics in data science is proposed that is grounded in two existing ethical traditions: virtue ethics and clinical microethics. The focus of this paper is teaching, but integrating virtue ethics and microethics in the context of data science is useful also for training of professionals, and it can be in principle extended to the use of data science in research and development.

3.1 Repetitive actions: virtue ethics

In the discussion above, we highlighted the difference between individuals understanding the ethical issues associated with the digital environment, and individuals who have internalized the ethics training and embody a core set of values in their daily practices. As described in the discussion above, we believe that it is only through the cultivation of the latter that we can not only grow an ethics of practice in digital/data activities, but also contribute to the ethical evolution of the dynamic digital environment. Although a deontological approach to data ethics is limited in achieving these goals, other ethical theories, such as virtue ethics, provide a valuable alternative approach.

3.1.1 Why virtue ethics?

In recent decades, an increasing number of ethicists are proposing approaches focused on character development and the acquisition of various skills related to ethical reasoning. For instance, in the context of data science, a striking proposal comes from the program *Embedded EthiCS* developed at Harvard University (Grosz et al. 2019). The program is motivated by the dissatisfaction towards traditional standalone methods of teaching applied ethics. As we stressed the importance of showing how ethics is interconnected with the daily activities of the digital environment, so Embedded EthiCS "employs a distributed pedagogy that makes ethical reasoning an integral component of courses throughout the standard computer science curriculum" (Grosz et al. 2019). For instance, one goal of the program is to help students to familiarize them "with a variety of concrete ethical issues and problems that arise across the field". They propose to do this by exposing students to "repeated experiences of reasoning through issues and communicating their positions effectively" (Grosz et al. 2019). In this way, they think, students will develop "ethical reasoning skills".

In this context of attention to character development, virtue ethics and its attention to the cultivation of moral excellences has had a revival. An excellence is "any stable trait that allows its possessor to excel" (Vallor 2016, p. 17). Especially in Aristotle, an excellence is a long-lasting attribute in virtue of which something or someone is good or things go well. For instance, being an excellent guitar player means not only being good at playing guitar once, but it is being good at playing it in stable and long-lasting ways. There are many excellences, and some of them are named by Aristotle as 'skills'. 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. The fact that a virtue is long-lasting is important to guarantee that it is a feature of a person "as a whole, and not just any old feature, but one that is persisting, reliable, and characteristic" (Annas 2011, p. 9). Virtues include features, such as justice, courage, faith, and hope. These enable individuals to determine the 'right action' within a specific context and to act consistently across many different activities and contexts.

While it is tempting to think about virtues as special, they are, in fact, very mundane and apply to very basic daily activities and social interactions. According to Russell (2015), "built into the very idea of what a virtue is are certain ideas about how such a thing develops" (p. 17). The cultivation of moral virtues (i.e., ways of being good at living with other individuals) come from specific *ethoi*. An *ethos* is simply a process that makes an action familiar. When we learn how to use a piece of software, such as R, all the commands appear very complex, but after a process of habituation it just comes naturally, even after years an individual will make use of the commands automatically. *Ethos*, in its essence, is a training—and we transform an attribute into something more stable and long-lasting by virtue of practice.

Even though Aristotle distinguishes sharply between skills and virtues, he nonetheless recognizes that virtues and skills are cultivated in similar ways, as also implied by our analogy with learning R. Therefore, we get better at being friends, at helping others, at empathizing with others only by being often in situations that require friendship, help other people, honesty, etc. The practice transforms an attribute in a stable disposition concerning our affective nature.[4] However, as Annas rightly emphasizes, habituation is not

[4] Another important aspect of cultivating virtues is the presence of moral exemplars.

mindless habit or 'routine'—anytime a virtuous person act virtuously he/she actively and intelligently finds the right course of action. Aristotle is not really specific in describing this process of habituation. Shannon Vallor (2016) describes in detail a process of habituation as moral self-cultivation by articulating different virtue traditions, most notably the Aristotelian, Confucian, and Buddhist. Among the different phases, the process includes the development of *moral attention* and *appropriate extension of moral concern*. We will get back to these virtues, but for the time being it suffices to say a virtue ethics approach to data ethics should foster the cultivation of such qualities.

From what we just said, it would appear that a virtue ethics approach may be well-suited to digital/data ethics through its focus on individual character development, individual responsibility for actions, and the acquisition of virtues through repetitive actions.[5] The general idea is that as data scientists acquire familiarity with the technical aspects of the digital environment by exercising daily their technical skills, analogously they can acquire a familiarity with the ethical subtleties of the digital environment by incorporating in their daily activities an attention to the ethical dimension of those infrastructures.

3.1.2 Teaching virtue ethics

Despite the promising emphasis on repetition and cultivation of moral abilities, virtue ethics is often criticized as being difficult to teach. It places considerable emphasis on the identification of exemplars—virtuous individuals who serve as models for behavior—as a means of observing and emulating virtuous behavior. However, how to identify exemplars and what do with them is a highly contentious issue.

Moreover, while virtue ethics emphasizes the importance of individual action and assessment of a specific and often multifaceted situation, it does not do a very good job in specifying the boundaries of a situation. This leads to the temptation to use case studies from macroethics that are "situations of crisis", where one virtue is obviously foregrounded and with exemplars whose lives and actions have little reference to the lived experiences of the student (Pennock 2019). This means that students struggle to see how they can emulate an 'Aristotelean approach' to ethical behavior as they have (1) no contact to the exemplars that are foregrounded, (2) no understanding of the granularity of what constitutes a "situation", and (3) no instruction in how to preserve the unity of the ethical self in the variety of different situations and roles that they occupy in daily life.

What is needed is an approach that provides case studies that describe daily interactions, and foreground exemplars that are relatable to individual students. In attempting to describe such case studies, we integrate a virtue theory approach with another approach called *microethics*.

3.2 Discrete actions: microethics

Microethics has been developed in the medical context, by stressing the importance of 'micro-decisions' in this environment and their ethical relevance. This approach to medical ethics was motivated in the following way. Ethics is about how we ought to live our lives. How we decide to do that depends on our long-term life's projects, which can be realized only through our actions. News received in the medical context are likely to change our life's projects and how we think that our conditions will constrain the way we realize them. The manner of delivering a medical news may have an impact on how we conceptualize our condition, how vulnerable we think we are because of it, and hence our future actions. Because these medical micro-decisions and the way they are delivered thus shape our patterns of behavior, the conduct of physicians in these exchanges is ethically relevant.

The case-based method of teaching ethics does not teach future physicians to pay attention to the ethical relevance of these seemingly mundane interactions. Microethics is proposed as a way to grapple with the ethical relevance of these micro-decisions. In the medical context, it can be conceived as an ethics of relations and of communication that should foster the development and cultivation of what Truog et al. call *moral imagination*, which is "the ability to recognize the range of options available in how communication occurs and how decisions are made and the ability to appreciate the ethical valence" (p. 12).

Micro-decisions and concrete scenarios make an important contribution to virtue ethics, as they foreground the boundaries of specific situations. For instance, deciding to use the word 'condition' rather than 'disease' may make the difference with respect to how a patient will conceptualize her disease—'condition' may lead to a less dramatic internalization. Similarly, using the word 'baby' instead of 'fetus' in counselling a patient who is seeking an abortion, may indirectly lead the patient to conceptualize her condition in a different way than she did before deciding to undergo the procedure.

The microethics approach has had important consequences in the clinical setting, by assisting clinicians to help patients to make choices that "are as true as possible to the patient's authentic self" (Truog et al. 2015, p. 13). Nonetheless, its influence can potentially extend far beyond clinical

[5] The University of Notre Dame-based Social Responsibilities of Researchers (SSR) funded by NSF (Bourgeois Forethcoming) is an example of a project trying to incorporate these issues. This project aimed at proposing an alternative to RCR training, by specifically incorporating a virtue ethics perspective—an alternative needed because of the limitations about RCR that we have emphasized above.

Table 2 Illustrating the range of daily actions that a typical data scientist would engage in

Type of daily action	Level of influence: direct	Level of influence: distributed
Activities	Develop code Generate, analyze, reuse data Maintain detailed reporting of daily activities	Curate, disseminate data and code Dissemininate results in papers, presentations, etc
Relations	Educate students, support peers Maintain productive environment Uphold institutional commitments	Engage with disciplinary and data science community activities Engage with public on research/data science issues
Responsibilities	Uphold key regulations, licenses and legislation surrounding data and software Act according to RCR Notify relevant authority of issues of concern	Safeguard good practice in community through monitoring and engagement

settings. By emphasizing self-reflexiveness in daily activities, the microethics approach offers a way of developing a multifaceted awareness of the contexts of micro-decisions. In particular, it assists individuals in fostering an awareness of their own biases and preferences that may be sneaked into micro-decision-making—unpacking the complex web of power dynamics and contextual pressures that inform any decision. A general microethical approach is aimed at making ethical reasoning a familiar activity, and at developing an ethical sensibility.

4 Discrete and repetitive: a micro-virtue ethics for data scientists

In this section we outline a new approach to data ethics that combines the emphasis on individual character from virtue ethics with the concrete situatedness of microethics. We expand on how such an approach could look by focusing on the ethics training of data scientists.[6] Micro-decisions, in particular, demonstrate how daily events can be packaged into discrete instances of ethical reflection. Using repetitive micro-tasks—such as coding, clicking on content, engaging in chat forums—as a means of fostering virtues provides an important means of developing ethically-aware individuals.

However, the clinical setting and the data science settings are rather different, and the wholesale transposition of microethics within the data science context can be challenging. On the one hand, interactions with individual patients makes it easy to identify the boundaries of these micro-events. Moreover, the use of a deontological framework focuses primarily on maximizing beneficence towards patients, thus providing a unified and coherent ethical narrative. On the other hand, data science settings present extremely challenging environments for the application of models such as clinical microethics. The range of actors operating within the online environment, the relative banality of daily actions, the predominant lack of a significant "other" all make it difficult to see how such an approach may be used.

In response to these challenges we propose an adapted, hybrid version that includes elements of all the models discussed above. This version of 'digital micro-virtue ethics' is grounded in the concept of digital citizenship of contemporary virtue ethics (Bezuidenhout et al. 2020). It uses elements of the clinical micro-ethics model to provide a means of bounding daily activities, and providing a means of linking these daily activities to the "bigger picture" ethical conundrums. Here we focus specifically on data science, so 'digital micro-virtue ethics' is 'data micro-virtue ethics'. The object of data micro-virtue-ethics is to provide a model for fostering digital citizenship and the acquisition of virtues through the thoughtful enactment of routine data science practices. In the remainder of this section, we provide some key aspects of our data micro-virtue ethics in the context of the daily activities of data scientists.

4.1 Bounding daily actions

Key to our model of data micro-virtue virtue ethics is the recognition that the routine, repetitive actions that constitute daily data science activity not only have ethical import, but are also events that can provide ethical training. It is therefore necessary to recognize the types of actions (activities, relations and responsibilities) that individuals engage with on a day-to-day basis. Table 2 below illustrates the range of daily actions that a typical data scientist would engage in. The table makes a distinction between the first level, direct actions that are more likely to have a defined 'other', and the second level, distributed ones that have no single 'other'.

[6] We interpret data scientist as any individual with expertise in coding and/or data analysis who works regularly with data to analyse, visualise, curate and disseminate it.

Table 3 Actions have both ethical content and provide an opportunity for ethical training

Action	Ethical content questions	Opportunity for virtue acquisition
Developing code	Am I re-using others' code responsibly?	Deliberations about sharing fosters reflection on citizenship, responsibilities, generosity
	Am I providing the necessary credit for the work of others?	Deliberation about providing credit fosters reflection on gratitude and humility
	Am I going to share my code with other?	Deliberation about misuse fosters reflection on responsibility to community and future users
	Could my code be misused?	Deliberation about misuse fosters reflection on societal responsibilities

4.2 Foregrounding virtuous behavior

In our model, we propose that all of these actions have both ethical content and provide an opportunity for ethical training. An example of how this is envisioned in a RCR context is demonstrated in Table 3 below.

As can be seen from Table 3, a single daily action provides rich opportunity for ethical reflection and virtue acquisition. It is important to recognize that daily data science activities also provide an important additional resource that can be used to support ethical development, namely online communities. Coding activities, as discussed in Table 3, rarely occur in isolation. Individuals are likely in regular contact with the forums and communities that have evolved around coding repositories (such as Zenodo), collaborative coding environments (such as GitHub), and open software (such as R). Interaction with these different forums not only socializes individuals to expected behavior, but allows individuals to identify community leaders that can act as exemplars and guide their daily activities. The nonhierarchical nature of these forums thus enables individuals to actively engage and interact with the individuals they identify as exemplars.

Maximizing the positive impact of each action will eventually lead to the cultivation of certain moral abilities that underpin digital citizenship, most notably what have been called *moral attention* and *appropriate extension of moral concern* (Vallor 2016). Here our model of data micro-virtue virtue ethics relies on the virtue tradition where 'moral abilities' are associated with the vocabulary of virtue theories, especially in the Aristotelian tradition. Shannon Vallor (2016) articulate a discourse on these moral abilities as practices of self-cultivation, while here we think about them both as practices and as virtues.

Moral attention refers to a form of moral perception, in the sense of being able to "discern and attend to those features of a particular situation that are most salient for the purpose of ethical judgement" (Vallor 2016, p. 99). In other words, a person who cultivates moral attention correctly identifies the moral dimension of facts—"a type of sensitivity to changes in one's moral environment" (p. 100). Incidentally, this may be very important in a data science context for a data scientist. We have in mind three dimensions where moral attention can make a difference in how data science tools are designed and implemented.

First, in training algorithms, a data scientist may make use of data sets or tools that have moral relevance, and a training in imagining how a tool or a procedure will affect from a moral point of view the recipients of the procedure is a way to cultivate moral attention. For instance, choosing which features to prioritize to return outputs such as credit score, recidivism risk, or insurance premiums, may incorporate factors beyond the control of individuals (i.e., factors for which an individual may not be responsible), and that may well be proxies for racial and sexist prejudices of all sorts (Martin 2019; Zliobaite 2017).

Second, it is the very goal that a data science system achieves that can be sometimes morally problematic. For instance, data scientists stress the importance of predictive accuracy of certain tools, but in some cases what we 'predict' in the future is just a repetition of past injustices. This happens when we apply data science tools within the justice systems to predict things such as probability of recidivism (Angwin et al. 2016), and we create dangerous feedback loops where the predictive success is created by the algorithm itself by constraining the autonomy of data subjects (O'Neil 2016).

Finally, moral attention does not stop to moral consequences. Rather, there should be attention towards the moral assumptions that implicitly drive one's technical choices. Small acts/choices are informed by an *ethos*, which is a substitute word for background experiences, values, commitment to social and cultural norms, and habits of the agent. The ethos materializes in the way the data scientist, for instance, trains the algorithm, selects data sets, etc. Algorithms are trained with data sets in order to produce a certain outcome. There is a narrative that says that algorithms are just about maximizing efficiency and accuracy (Martin 2019). Even though this is seen as 'neutral', the emphasis on 'efficiency' and 'accuracy' is already a sign of a particular ethos. This and similar cases show that algorithms can incorporate ethical beliefs and, led to the extreme consequences, they can reinforce

Table 4 Complicated contexts and power dynamics framing each action provides important means of connecting daily digital practices to the 'big picture'

Action	Contextual considerations	Ethical considerations	'Big picture' ethical conundrums
Developing code	Individual, institutional, community ownership	Who (should) own the code?	Open Science/open software
	National and international copyright and patent	How does one manage conflicting priorities?	Limits of ownership of digital artifacts
	Coding community norms and requirements	How does one manage conflicting priorities to self and communities?	Free and open source software
	Social norms and requirements	How does any code contribute contribute to the broader body of the digital landscape?	FFP (fabrication, falsification, plagiarism) in research

existing social and cultural norms. Data scientists should be equipped to anticipate these issues naturally.

The *appropriate extension of moral concern* is another aspect of this process of moral self-cultivation that is important for a micro-virtue ethics approach to data science. This is defined as the "ability to expand one's basic attitude(s) of moral concern (…) to the right beings, at the right time, to the right degree, and in the right manner" (Vallor 2016, p. 110). This emphasizes another aspect of digital micro-virtue ethics, which is the ability to identify the relevant stakeholders and not just focusing on the immediate recipients of the machine learning system. This also implies that, sometimes, we identify relevant stakeholders but we think that there are no particular moral concerns attached to them (Robbins 2019). With this ability, we direct moral attention and concern to those we think deserve it. Please note that the point is not to identify in a univocal way who deserves moral attention or not. Given our different ethoi, it is likely that individuals will extend moral concerns in radically different ways. The issue here is to make sure that this process is transparent, and that the data scientist knows that it is part of the practice of data science to include it. For instance, when we have to decide which features to consider in training an algorithm, it is important to identify who we are leaving out (Lerman 2013)—in other words, who our tool will not target because he/she does not fall under a certain category.

4.3 Recognizing the multiplicity of roles

Another important aspect that data scientists have to be trained to recognize is the complexity of the socio-technical system they work in. This means emphasizing the different roles within a system, and the different moral concerns that each role may raise. Tomsett et al. (2018) elucidates the structure of what they call a *machine learning system*, which is defined as "one or more machine learning models,[7] the data used to train the model(s), any interface used to interact with the model(s), and any relevant documentation" (p. 9). In such a system, they distinguish different actors involved: creators (owners and/or implementers), operators, executors, decision-subjects, data-subjects, examiners. There may be different data scientists in the same machine learning system who fulfill different roles, or the same data scientist may occupy different roles in different daily activities. These roles could be as operators, implementers, but examiners as well. It is not our goal here to say precisely which roles the data scientists may possibly have, but just to say that data scientists may be different types of agents, and depending on the type of agents that they are, they may undertake different actions, and have different communities and stakeholders to consider.

4.4 Linking micro-events to the "big picture"

Critically examining the actions presented in Tables 2 and 3 draws attention to the complicated contexts in which they occur. The enactment of these actions is influenced by the social/political/physical world in which the individual, as well as the distributed and global digital environment. Drawing attention to the complicated contexts and power dynamics framing each action provides an important means of connecting daily digital practices to the 'big picture'. Returning again to the example of coding, it is possible to frame this connection in the way outlined in Table 4.

Approaching 'big picture' ethical discussions from our approach offers two important benefits. First, it enables the individual to see how their actions are linked to the 'big picture' conundrums. This foregrounds how each individual—as a digital citizen—has both the agency and the responsibility to safeguard the digital community that is enacted in their daily activities. Second, this approach highlights that even the smallest of actions has the potential to have important consequences. This makes it difficult for the individual to engage in self-centred misbehaviors. However, the important aspect to emphasize is that, in this way, our approach does

[7] They probably mean 'algorithm'.

connect naturally to the important issues raised in a macroethics context. In this sense, microethics does not exclude macroethics, but rather these two approaches complement each other.

5 Teaching data scientists using a micro-virtue ethics approach

The micro-virtue ethics approach proposed above offers a novel alternative for digital/data ethics pedagogy. Indeed, by combining virtue ethics and clinical microethics it offers a pathway for training data scientists to identify their responsibilities, activities, and relations qua data scientists within their daily activities. For instance, an important part is learning how to identify the stakeholders of their machine learning systems, with the aim of developing a sensibility how personal values can influence the design of algorithms and have moral consequences. In this way, they can recognize the issues raised in the macroethical context in the own sphere of influence. Developing this sensibility is akin to cultivating specific moral abilities such as 'moral attention' and 'appropriate extension of moral concerns'. Given that we do not aim to formulate any specific ethical content but only a method to develop a moral sensitivity, our approach can be applied to any environment where data science is used. This means that, while the method is portable virtually in any context, the precise ethical content of this 'moral training' (e.g., which moral aspects to emphasize, etc.) will have to be arranged according to the specificities of the context (e.g., military sector, business, medicine, etc.). For instance, we have shown how the ethical content can be conceptualized in the context of coding (i.e., Sects. 4.1, 4.2) or in professional data science (i.e., the three dimensions of moral attention in Sect. 4.2). In order to show even more concretely how our method can be applied, in the next section we describe the construction of one such course for data science students designed by one of the authors (LB).

5.1 The CODATA-RDA schools for research data science (SRDS)

The SRDS were founded in 2016 to provide data science training to early career researchers from low/middle-income countries (LMICs).[8] These 2-week residential schools provided a "broad and shallow" introduction to data science, as demonstrated by the curriculum in Fig. 1. In contrast to other data science training, the SRDS oriented the curriculum around "open and responsible research"—particularly Open Science, Responsible Conduct of Research and Responsible Digital Citizenship.

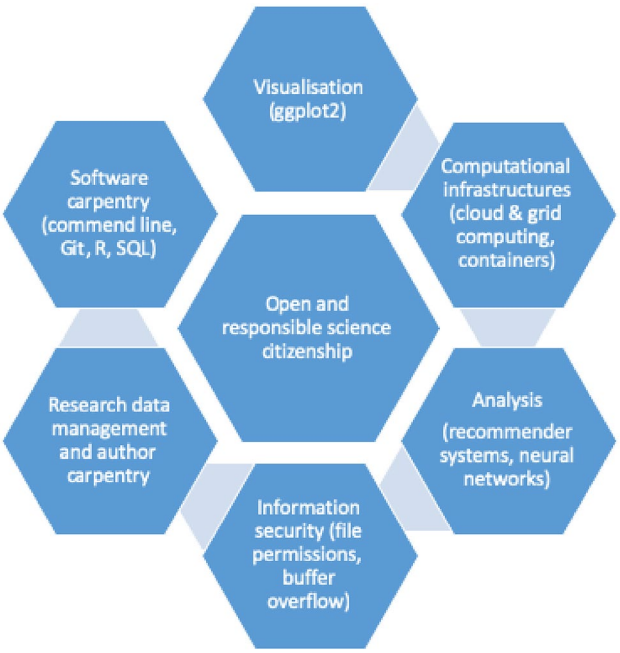

Fig. 1 Schematic diagram of SRDS curriculum, highlighting centrality of open and responsible science citizenship

One of the authors (LB) developed the curriculum for open and responsible research within the parameters of the curriculum. These parameters included students from diverse disciplines and nationalities, limited formal teaching time (3.5 h) and no expectation of prior ethical training from students.

The process of teaching open and responsible science citizenship to the data science students was taught in a number of different phases, as detailed by Fig. 2. First, the students were introduced to the ethics of data science on a macro-ethics level. Two formal lectures covered topics such as Open Science, Responsible Conduct of Research and ethical issues relating to data science, such as algorithmic bias, "infra-ethics" and the ethics of machine learning. In these lectures, key values such as justice, beneficence and nonmaleficence were highlighted to demonstrate the moral continuity within these different discussions.

The students at SRDSs came from many different disciplinary and national boundaries, and many had no prior ethics training. In order to make the instruction more accessible, the discussions about individual rights and responsibility were prefaced by an introduction to the concept of "data citizenship". This concept is based on an Aristotelian view of citizenship as ethical obligations arising out of social living (Aristotle 2014) as conceptualized through the dual lenses of Open Science and Responsible Conduct of Research (Bezuidenhout et al. 2020). This was found to

[8] https://codata-rda-datascienceschools.github.io/.

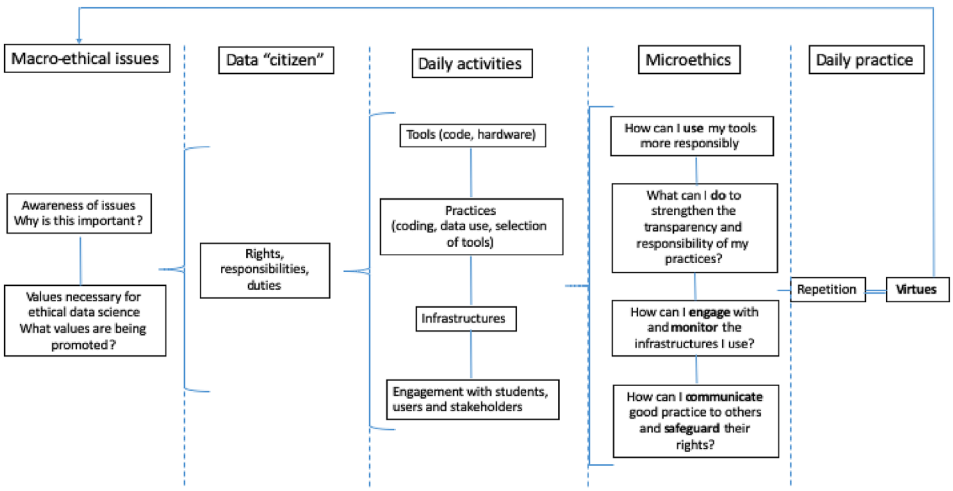

Fig. 2 Phases of teaching a micro-virtue ethics for data science

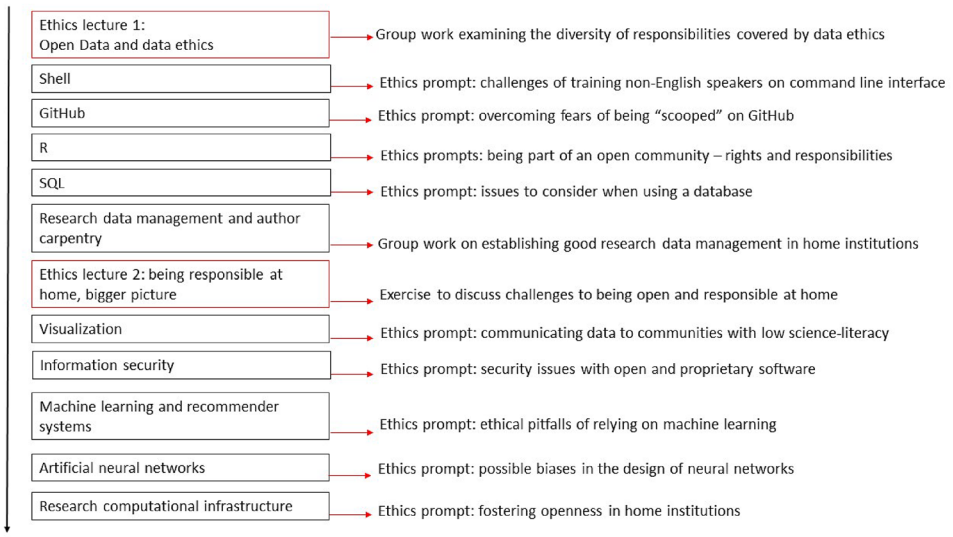

Fig. 3 Schema of ethics prompts used during SRDS. Full description of exercises available on GitHub

be a useful tool for introducing the reciprocal relationship of rights and responsibilities, and how being part of a community (such as the research community) made both of these an inalienable part of individual identity. This encouraged a practice-based perspective on ethics that is contextually informed (MacIntyre 2011).

A key objective of the ethics training at the SRDSs was to ensure that students understood that ethical practice is integrated into daily research activities, and is not a standalone subject to be visited occasionally.

In the next phase of the ethics instruction, students engaged in a range of "ethics exercises" that were explicitly linked to the computing modules listed in Fig. 1. These exercises were 15 min directed discussions on an ethics question linked to the practices learnt during the module. The ethics exercises are administered via a range of different modalities, including writing answers down on post-it notes, live voting and mind-mapping. These ethics exercises are specifically related to the content of the module completed, while linked to the broader ethical issues and digital citizenship concepts introduced in the lectures[9] (Bezuidenhout et al. 2020) (Fig. 3).

The object of the ethics exercises was to foster a critical reflexivity in daily practice. The exercises assisted students in seeing how their daily activities and decisions had impact on the broader issues discussed in the lectures. It helped them to identify the agency that they held to promote ethically positive practices, infrastructures and digital futures. It is anticipated that linking daily practice to ethical issues aids students in developing the moral imagination discussed above.

[9] A sample list of these exercises is available at https://codata-rda-datascienceschools.github.io/.

The combination of formal lectures and module-related ethics exercises was designed to achieve three key outcomes for students. First, to build their confidence in being able to engage with ethics discussions relating to data science. Second, to foster their understanding of their agency to act as an open and responsible science citizenship within their daily research activities. Third, to assist them in identifying areas for action as an open and responsible science citizen.

Other two aspects should be emphasized. First, it was intended that by understanding the connection between daily actions and the high-level issues, students would feel confident to engage in discussions on ethics. Second, by learning data science skills, the students assumed additional responsibilities assume community responsibilities, such as to their communities. This could be as exemplars for best practice, by surveilling emergent digital infrastructures, or by developing ethical practice within their research communities.

5.2 Future work

The example described in Sect. 5.1 is only the starting point of our micro-virtue ethics approach to data ethics. We want to develop and expand our approach in three directions.

First, we recognize that the limited space and time of a summer school makes our ideas on moral habituation difficult to be properly applied. It will be important in the future to organize longer courses where the space between the teaching of technical skills needed by data scientists and the attention to moral development is more evenly distributed. Ideally, the goal is to be able to integrate the strategies elucidated in Sects. 5 and 5.1 in the entire curricula of both undergraduate and graduate computer science majors. As already noticed at the beginning of this article, we are not alone in pursuing this goal. Currently, *Embedded EthiCS* program has integrated its strategies of teaching ethics in several courses within the computer science curriculum at Harvard. The pedagogy of this program is, as elucidated above, similar to the one we have outlined here, but its foundations are less clear, and there is no mention of virtues and/or microethics. Similarly, Marion Boulicalt and Milo Phillips-Brown pioneered a similar approach at MIT and, they say, they want to teach ethics as a skill[10] by explicitly referring refer to Aristotle's *techne*. We could not find exact indications of how their modules look like, but we just want to point out a couple of things. First, the explicit reliance on Aristotle's *techne* is puzzling. As it is widely known, in Aristotle practical knowledge includes *poiesis* (i.e., 'making', such as making a chair) and *praxis* ('acting', such as actions constitutive of the good life, human flourishing, etc.). These have two different goals, namely that "the end of production is something other than production [e.g., making a chair], while that of action is not something other than action, since doing well in action is itself action's end" (EN VI.5, 11140b). In this context, *techne* is what perfects *poiesis* by providing "the knowledge of a set of rules and standards that are applied in order to make a well-constructed and well-formed external product" (Ratti 2020, p. 166), while *praxis* is perfected by *phronesis*. Therefore, *techne* does not really deal with ethics, which is more the domain of *praxis* and *phronesis*. Second, appealing to skills and techne in this way comes with risks. It emphasizes the importance of a set of rules to achieve a certain goal. Independently of what Aristotle thought about these issues, this idea promotes a misleading picture of ethics: the literature on the problems of ethics as following a set of rules is just overwhelming. However, a more charitable interpretation would understand their claims within the analogy between virtues and skills formulated in great detail by Annas (2011).

Second, we recognize that a micro-virtue ethics approach to data ethics should not be limited to the education of data scientists. While in the case of data scientists it is striking how their technical choices have ethical ramifications, we should not underestimate the role of normal users in the digital environment. In other words, we envision a course in digital literacy, where users of search engines and digital platforms are habituated to consider the moral relevance of seemingly morally neutral acts in the digital environment.

Finally, as data types and practices vary across disciplines it is important that there is no "one size fits all" when it comes to data ethics. Indeed, certain practices or concerns will be highlighted according to the type of research being conducted and the data types produced. It is therefore necessary to develop a robust description of digital citizenship that suits these different contexts of application.

6 Conclusion

In this article, we have formulated a full-fledged framework to 'embed' ethics in the practice of data science which overcomes some of the limitations of a macroethical approach to data science. While the proliferation of macroethics initiatives is to be welcomed as a positive sign, we have identified some limitations of this approach. In particular, we have developed the idea that macroethics is difficult to be applied to the daily activities within the data science environment. Moreover, this problem of applicability is reflected also in the way data ethics is taught, especially data ethics to data scientists. Stand-alone courses based on the macroethical issues struggle to make a direct connection between the ethical issues and the daily practice of data science.

[10] https://shass.mit.edu/news/news-2020-boulicault-and-phillips-brown-ethics-technical-curriculum.

In order to overcome these limitations, we proposed to ground teaching strategies within a virtue ethics framework, and to think about ethical training as a way to help students and practitioners to cultivate two main virtues (i.e., moral attention and appropriate extension of moral concerns). However, we have also complemented this approach with ideas from microethics, which emphasizes the ethical relevance of small acts and, unlike traditional virtue theory, is able to provide a framework to understand and grasp the granularity and uniqueness of each situation in which we act. Finally, we have described how this framework works within the curriculum of open and responsible research of SDRDS developed by one of the authors (LB).

We strongly believe that this novel approach, grounded in virtue ethics, offers an important contribution to discussions on digital/data ethics. It demonstrates how the focus on character development and daily routine actions can provide a consistent approach across a wide variety of disciplinary applications.

Acknowledgements We want to thank Thomas Stapleford, Don Howard, Nathaniel Warne, Celia Deane-Drummond, Darcia Narvaez, Dori Beeler, Emily Dumler-Winckler, and Mark Graves for inspiring discussions about virtues. We are also indebted to instructors and students of the CODATA-RDA Schools for Research Data Science for their valuable inputs.

Author contributions LB and ER have contributed equally.

Funding Open access funding provided by Johannes Kepler University Linz.

Compliance with ethical standards

Conflict of interest Not applicable.

References

Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine bias - there's a software used across the country to predict future criminals. And it's biased against blacks. ProPublica

Annas J (2011) Intelligent virtue. Oxford University Press, Oxford

Aristotle (2014) Nicomachean ethics. In: Reeve CD (ed) Hackett Publishing Company, Indianapolis

Beauchamp TL (2015) Theory, method, and practice of principlism. Oxford Handbook of Psychiatric Practice, no. October 1–24. https://doi.org/10.1093/oxfordhb/9780198732365.013.31.

Beauchamp T, Childress J (2009) Principles of biomedical ethics, 6th edn. Oxford University Press, Oxford

Bezuidenhout L, Quick R, Shanahan H (2020) "Ethics when you least expect it": a modular approach to short course data ethics instruction. Sci Eng Ethics. https://doi.org/10.1007/s11948-020-00197-2

Binns R (2018) Fairness in machine learning: lessons from political philosophy. pp 1–11. http://arxiv.org/abs/1712.03586

Bourgeois M (Forthcoming) Virtue ethics and social responsibilities of researchers. In: Ratti E, Stapleford T (Eds) Science, Technology, and the Good Life: Perspectives on Virtues in Modern Science and Technology

Bozdag E (2013) Bias in algorithmic filtering and personalization. Ethics Inf Technol 15(3):209–227. https://doi.org/10.1007/s10676-013-9321-6

Chen J-Y (Forthcoming) Integrating virtue ethics into responsible conduct of research programs: challenges and opportunities. In: Ratti E, Stapleford T (Eds) Science, Technology, and the Good Life: Perspectives on Virtues in Modern Science and Technology

Dressel J, Farid H (2018) The accuracy, fairness, and limits of predicting recidivism. Sci Adv 4(1):1–6. https://doi.org/10.1126/sciadv.aao5580

Floridi L (2018) Soft ethics and the governance of the digital. Philos Technol 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. Harv Data Sci Rev 1:1–13. https://doi.org/10.1162/99608f92.8cd550d1

Grosz BJ, Grant DG, Vredenburgh K, Behrends J, Hu L, Simmons A, Waldo J (2019) Embedded EthiCS: integrating ethics broadly across computer science education. https://cacm.acm.org/magazines/2019/8/238345-embedded-ethics/fulltext#comments

Hagendorff T (2020) The ethics of AI ethics. Minds Mach 30:99–120. https://doi.org/10.1007/s11023-020-09517-8

Kitto K, Knight S (2019) Practical ethics for building learning analytics. Brit J Educ Technol 50(6):2855–2870. https://doi.org/10.1111/bjet.12868

Komesaroff P (1995) From bioethics to microethics: ethical debate and clinical medicine. In: Komesaroff P (ed) Troubled bodies - critical perspectives on postmodernism, medical ethics, and the body. Duke University Press, Durham

Lerman J (2013) Big Data and its exclusions. Stan Law Rev

MacIntyre A (2011) After Virtues. Bloomsbury Academic, New York

Madaio MA, Stark L, Wortman Vaughan J, Wallach H (2020) Co-Designing checklists to understand organizational challenges and opportunities around fairness in AI. In: Conference on Human Factors in Computing Systems - Proceedings, pp 1–14. https://doi.org/10.1145/3313831.3376445

Martin K (2019) Ethical implications and accountability of algorithms. J Bus Ethics 160(4):835–850. https://doi.org/10.1007/s10551-018-3921-3

McLennan S, Fiske A, Celi LA, Müller R, Harder J, Ritt K, Haddadin S, Buyx A (2020) An embedded ethics approach for AI development. Nat Mach Intell 2(9):488–490. https://doi.org/10.1038/s42256-020-0214-1

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, pp 729–733. https://doi.org/10.1145/3236024.3264833

Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. Nat Mach Intell 501–507

Morley J, Floridi L, Kinsey L, Elhalal A (2019) From what to how. An overview of ai ethics tools, methods and research to translate

principles into practices. Sci Eng Ethics. https://doi.org/10.1007/s11948-019-00165-5
O'Neil C (2016) Weapons of math destruction. Crown Publishing Group, New York
Pennock RT (2019) An instinct for truth: curiosity and the moral character of science. MIT Press
Ratti E (2020) Phronesis and automated science: the case of machine learning and biology. In: Sterpetti F, Bertolaso M (eds) A critical reflection on automated science - Will Science Remain Human, Springer
Robbins S (2019) A misdirected principle with a catch: explicability for AI. Minds Mach 29(4):495–514. https://doi.org/10.1007/s11023-019-09509-3
Russell D (2015) Aristotle on cultivating virtue. In: Snow N (ed) Cultivating virtue - perspective from philosophy, theology, and psychology. Oxford University Press, Oxford, pp 17–48
Susser D, Roessler B, Nissenbaum H (2019) Online manipulation: hidden influences in a digital world. Geol Tech Rev 1
Tomsett R, Braines D, Harborne D, Preece A, Chakraborty S (2018) Interpretable to whom? A role-based model for analyzing interpretable machine learning systems. http://arxiv.org/abs/1806.07552
Truog R, Brown S, Browning D, Hundert E, Rider E, Bell S, Meyer E (2015) Microethics: the ethics of everyday clinical practice. The Hastings Center Report
Vallor S (2016) Technology and the virtues - a philosophical guide to a future worth wanting. Oxford University Press, Oxford
Vakkuri V, Kemell KK, Abrahamsson P (2020) ECCOLA - a method for implementing ethically aligned AI systems. https://doi.org/10.1109/seaa51224.2020.00043
Whittlestone J, Nyrup R, Alexandrova A, Cave S (2019) The role and limits of principles in AI ethics. pp 195–200. https://doi.org/10.1145/3306618.3314289
Williams BA, Brooks CF, Shmargad Y (2018) How algorithms discriminate based on data they lack: challenges, solutions, and policy implications. J Inf Pol 8(2018):78. https://doi.org/10.5325/jinfopoli.8.2018.0078
Žliobaitė I (2017) Measuring discrimination in algorithmic decision making. Data Min Knowl Discov 31(4):1060–1089. https://doi.org/10.1007/s10618-017-0506-1
Zuboff S (2015) Big other: surveillance capitalism and the prospects of an information civilization. J Inf Technol 30(1):75–89. https://doi.org/10.1057/jit.2015.5
Zwolinski M, Schmidtz D (2013) Environmental virtue ethics. In: Russell D (ed) The Cambridge companion to virtue ethics. Cambridge University Press, Cambridge, pp 221–239