theFuture ofScience andEthics

Rivista scientifica a cura del Comitato Etico
della Fondazione Umberto Veronesi

Volume 6 ■ 2021 ■ ISSN 2421-3039

Fondazione
Umberto Veronesi
– per il progresso
delle scienze

Articoli

Microethics for healthcare data science: attention to capabilities in sociotechnical systems

Per una microetica della scienza dei dati nella sanita pubblica: capacità fondamentali e sistemi socio-tecnici

MARK GRAVES[1]
EMANUELE RATTI[2]
mnl.ratti@gmail.com

AFFILIAZIONE
1. Parexel AI Labs
2. Johannes Kepler University Linz

SOMMARIO

Negli ultimi anni c'è stato un dibattito molto acceso sulla efficacia di certi sistemi etici applicati alla scienza dei dati. Tra le varie limitazioni, è stata messa in dubbio la capacità di implementare questi sistemi nella pratica scientifica. In questo breve articolo, avanziamo una proposta per superare questa limitazione. La nostra proposta analizza come ogni singola scelta tecnica fatta dagli scienziati dei dati possa avere potenzialmente una valenza morale. Usando il contesto medico come esempio, mostriamo che concentrarsi sui fattori socioculturali dei dati sanitari risulta utile per promuovere una consapevolezza etica nel momento stesso in cui vengono prese decisioni tecniche che, di solito, vengono considerate moralmente neutre. Invece di applicare vaghi e generali principii etici, il nostro approccio promuove l'etica come esercizio riflessivo per capire l'impatto delle decisioni tecniche sui fattori socioculturali.

PAROLE CHIAVE

Etica dei dati

Scienza dei dati

Cartella medica elettronica

Attenzione morale

Etica delle virtù

ABSTRACT

It has been argued that ethical frameworks for data science often fail to foster ethical behavior, and they can be difficult to implement due to their vague and ambiguous nature. In order to overcome these limitations of current ethical frameworks, we propose to integrate the analysis of the connections between technical choices and sociocultural factors into the data science process, and show how these connections have consequences for what data subjects can do, accomplish, and be. Using healthcare as an example, attention to sociocultural conversion factors relevant to health can help in navigating technical choices that require broader considerations of the sociotechnical system, such as metric tradeoffs in model validation, resulting in better ethical and technical choices. This approach promotes awareness of the ethical dimension of technical choices by data scientists and others, and that can foster the cultivation of 'ethical skills' as integral to data science.

KEYWORDS

Data ethics

Data science

Electronic health records

Moral attention

Virtue ethics

DOI: DOI: 10.53267/20210106

1. INTRODUCTION

In the past few years, there has been an explosion of ethical frameworks for data science, machine learning, and artificial intelligence (AI) (Jobin et al., 2019). The origin of these ethical frameworks is two-fold. First, they are developed from shared views about the values that democratic societies should promote, such as fairness or sustainability (Floridi et al., 2018; Human Rights Watch, 2018). Second, proponents take inspiration from professional code of ethics of cognate disciplines, such as software engineering or statistics, or from disciplines where ethical concerns have been systematized, such as medicine (Mittelstadt, 2019; Véliz, 2019). These frameworks are usually structured as a list of ethical principles that have to be applied in order to ensure that the data science tools meet some important desiderata.

Recently, Floridi (2018) identified 84 principles for ethical AI that converge around general concepts such as beneficence, non-maleficence, sustainability, transparency, justice, etc. However, two problems have been noticed with this approach. First, there is a problem of *effectiveness*: there is evidence that principles *per se* have little efficacy in fostering ethical behavior (McNamara et al., 2018). Second, there is an *applicability* problem: these principles are difficult to operationalize (Morley et al., 2020). While it is widely acknowledged by data scientists that data science should not do harm, unfairly discriminate, etc., how exactly these insights translate into technical operations is far from clear. Some have proposed to develop mathematical methods that can formalize or quantify ethical principles, such as fairness (Srivastava et al., 2019). However, it has been also found that the many, often incompatible, ways of defining the same principle hinder the feasibility of the approach. In response to these issues, there has been a proliferation of AI ethics checklists (Madaio et al., 2020). These AI ethics checklists seek to increase active participation of practitioners, and they can be structured around principles or phases of, for instance, the data science process. However, as noticed in Madaio (2020), these checklists are often framed in binary ways (i.e. answers to 'yes/no' questions) which misleadingly simplify ethical deliberation. Moreover, the 'check-list' approach embraces technological solutionism (Selbst et al., 2019), while ignoring that data science tools are always embedded in sociotechnical contexts, and that ethical considerations are more often than not contested.

In this paper, we propose an approach that conceptualizes ethics in data science as deliberation about the connection between technical choices and sociocultural factors, which form the foundation of sociotechnical systems. While checklists based on principles can be mindlessly applied, here we advocate an approach that promotes awareness of the ethical dimension of technical choices, and that can promote the cultivation of 'ethical skills'. This can be summarized with a slogan: rather than applying ethics, data scientists should learn how to think about ethical problems.

2. THE APPROACH: MICROETHICS AND CAPABILITIES

The goal of our approach is to stimulate the cultivation of ethical skills (which are usually called 'virtues'). In particular, we emphasize the importance of the virtue of moral attention (Vallor, 2016; Bezuidenhout & Ratti, 2020; Ratti & Graves, 2021), which we define as the ability to grasp the ethical relevance of one's actions. Data scientists have to learn first and foremost to identify and assess the ethical relevance of each of their technical acts in the data science process.

We define ethical relevance by adapting the broadly established capability approach (Nussbaum, 2006) to the context of this work. This approach proposes an original conceptualization of well-being and human dignity based on the central concept of 'capabilities.' Capabilities are defined as what people can choose to do or to be. Well-being and human dignity are characterized in terms of substantial freedoms that individuals have in deciding how to live their own lives. Nussbaum (2006) compiles a list of ten basic capabilities, which are life; bodily health (sometimes discussed as 'health' or 'health agency'); bodily integrity; senses, imagination, and thought; emotion; practical reason; affiliation; concern for other species; play; control over one's environment. These capabilities can be 'actualized' or 'exercised' thanks to 'conversion factors', which are understood as personal, social, institutional, or environmental factors that can facilitate the development of a capability in a direction or another. For instance, in a famous example, bicycles are described as a conversion factor that can contribute substantially to the capability of affiliation, because it allows individuals to move around and meet other individuals. However, the bicycle is not the only relevant conversion factor. Depending on other conversion factors, one will be able to expand the capability of affiliation with bicycles only to a certain extent. Using bicycles to expand affiliation is easier for individuals living in Netherland than it is for Bedouin in the desert, because of the social conversion factor of street infrastructures allowing individuals to use bicycles to go anywhere (Oosterlaken, 2014).

By using the capability approach as a heuristic, we define an action as ethically relevant when it impacts any of the ten basic capabilities or substantial freedoms that Nussbaum describes. Capabilities are impacted when the conversion factors (personal, social, and environmental) that are necessary to actualize them are constrained or blocked, or when factors that impede capabilities are ignored or made invisible. In our approach, a data scientist cultivates the ethical skill of moral attention when, for each of the technical acts she performs in the data science pipeline, she systematically asks questions about the impact for basic capabilities (in connection to conversion factors as discussed above). The importance of conversion factors cannot be stressed enough. Often, capabilities are not visible per se, but their status can be inferred by looking at conversion factors. Usually, data sets that data scientists process tell us a lot about these conversion factors. In the case of the bicycle, data about sales of bicycles themselves, plus data about cycle paths infrastructures can give us important hints about the possibility for expanding a capability such as affiliation. As moral attention to capabilities requires specific contexts for application and elucidation, we focus on sociotechnical systems for healthcare, and especially decisions that might affect health agency.

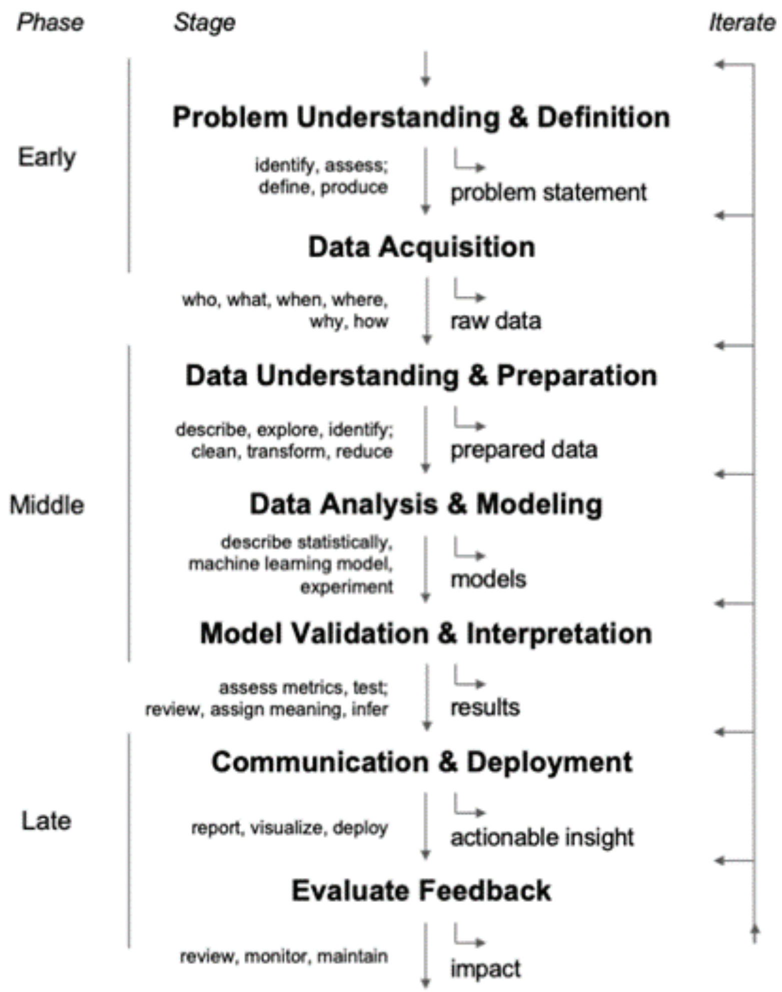
We call our approach 'MicroEThical Approach' to data science (META), because, as any microethics (Komesaroff, 1995), it emphasizes the ethical relevance of micro-, cumulative decisions characterizing the daily activities of data scientists. This means that data scientists can decide to develop the data science tool in one direction rather than another, because of the ethical relevance of a particular technical decision (informed by considerations based on capabilities and conversion factors). This has an important consequence. Even if we imagine the ideal situation where data scientists are omniscient and can identify all the conversion factors at play in each data science phase, this does not mean that individuals will reach the same conclusion. Sometimes impacts on capabilities can stand in a tradeoff relation, and individuals have different ways of handling tradeoffs depending on the fundamental values that they hold. In the next section we operationalize META by describing how questions about capabilities and conversion factors can be embedded in each phase of the data science process.

3. DATA SCIENCE PROCESS

In order to define and visualize data science as a field, we use Drew Conway's classic Venn Diagram of this discipline as the intersection of hacking skill, math and statistics, and substantive expertise (Conway, 2013). One the one hand, computer scientists may see data science as incorporating statistics and domain expertise into machine learning. On the other hand, a statistician may view data science as automating applied statistics for big data. In business, data analysts become data scientists by strengthening their programming, statistics, and machine learning skills. Another perspective will see data scientists as using engineering, statistical, and domain expertise to develop computational models (Weisberg, 2013) for a domain rather than just model-free programs, descriptive statistical models, or business intelligence tools using data from a domain.

To organize the data science process and demonstrate how moral attention can be embedded into it, 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 Figure 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 definition and understanding 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, and thus

Figure 1: Data science process stages, phases, and flow



have more conspicuous ethical demands. We structure the remainder of the article in terms of the stepwise process to clarify specific places a microethical approach adds technical as well as ethical value to data science, and focus particularly on the middle phase and its model validation and interpretation stage.

Early Phase

As a working example, consider the problem of extracting text from Electronic Health Records (EHR). This process can be useful for constructing features on Social Determinants of Health (SDoH; World Health Organization, 2011) to predict, cohort, and possibly intervene on patients with diabetes and/or cardiovascular disease. SDoH are the 'conversion factors' that can affect the capability of health and health agency. For instance, Ruger (2010) describes social, cultural, and personal factors that impact health such as health knowledge, health-seeking skills, self-governance, effective health-decision making, social norms, social networks and capital, group membership, material circumstances, access to health services, etc. Some of these factors can be found as 'features' in EHR data sets, such as the ones in large US healthcare systems (e.g., hospitals, HMO, Medicare Advantage program, etc.). Let us see how our framework can be applied to the different phases of this working example.

The first data science stage is 'problem definition'. This is replete with opportunities to use microethical decisions to impact capabilities. In addition to the programming and statistics expertise, data science also requires expertise in the domain that is sufficient to understand and address the problem. Knowledge of the domain means being able to identify SDoHs or conversion factors affecting capabilities such as health. Conversion factors include broad areas of healthcare access, education, economic stability, and social context. These directly affect health through work, unemployment, access to or lack of transportation, healthy food, stress, social support or exclusion, quality of housing, environmental conditions, and incarceration. Understanding the problem may require additional investigation of these factors sufficient to make the decisions necessary to address the problem. For instance, these factors affect cardiovascular disease and diabetes in multiple ways (Liburd et al., 2005), and although the data scientist would typically not be expected to know all these at the beginning of a project, sufficient background is needed to obtain the necessary expertise while making the microdecisions during the data science process. The requirements of predicting, cohorting, and intervening can focus the understanding on identifying the conversion factors that, by impacting capabilities, have ethical relevance.

The 'data acquisition' stage raises ethical issues that include consent, privacy, HIPAA, security, deidentification, and equitable representation. EHR include EMR (Electronic Medical Records) written by physicians and nurses, lab reports, pharmacy refills, social worker notes, and other records relevant to a person's health care. Ethical principles abound in data acquisition, especially around ownership, despite the stage's simplification in some technical accounts to data ingestion, selection, or discovery. In addition, microethical attention requires, for example, ensuring that extra effort is taken to acquire representative data from underserved communities who may have poorly accessible records, and not just from well-funded health systems whose data may be more accessible and interoperable.

Middle Phase

The middle phase depends most heavily on the technical skills specific to data science, specifically experimentation with computational models. Data science, at its essence, builds upon its computational and statistical foundations to support decisions and other actionable insights within a domain. Although data scientists also use statistical methods, the subtle emphasis on working within the domain of discourse to create interventions and other changes shifts the data scientist's attention from descriptive or predictive aspects of the model, to a more domain-focused support for decision making. The experimental dimension of data science includes not only the validation of the model with respect to acquired or forward-looking data, but also ideally includes the experimentation necessary to evaluate the model within the domain-specific, decision-making context.

The middle phase begins with 'data understanding and preparation' stage, which combines two highly-interrelated processes. 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. Determining whether data is missing or an outlier, for example, can require awareness of social context for the data beyond statistical measures, as those factors may limit truly representative data acquisition. The stage also includes data preparation, which consists of cleaning the data followed by transforming and reducing it to prepare for modeling. Data cleaning involves addressing the data quality problems, and can be an unanticipatedly time-consuming aspect of the process requiring additional effort to ensure representative but less accessible or interoperable data is comparably cleaned. Data transformation changes the data values, format, or structure in a way more amenable to the problem being addressed, and feature engineering can surface distinctions in social determinants. Data reduction modifies the quantity and/or structure of the data by sampling, selecting features, or applying dimensionality reduction methods and requires attention to representative sampling, minimizing bias in feature selection, and retaining transparency in dimension reduction.

In 'data analysis and modeling', we identify three technical levels of modeling and experimentation that have moral significance. First, descriptive statistics with no or minimal modeling can be used to analyze the data, e.g., by business analysts using complex spreadsheets or simple 'dashboard' visualizations, which puts the findings in their sociotechnical context and requires an accurate understanding of the sociotechnical system to avoid obfuscating capabilities and related conversion factors. Second, statistical and machine learning models may suffice for predictions and other actionable insights given the acquired data and well-defined project goals, which affects how people may interact with its encompassing system in the future. Third, experimentation may be required also to effect appropriate change within the sociotechnical system. In data science experimentation, attention to human capabilities is not only valuable, but also morally and technically necessary. Because numerous domain-specific factors affect a variety of technical decisions with complex effects in the sociotechnical system, the data scientist must attend to those effects as part of the experimental process. Because the modeling inputs and outputs are defined internally by the data science process and affected by experimentation, then the experimentation necessarily has a social impact inseparable from its moral implications.

In the stage of 'model validation and interpretation', one compares model outputs to independent observations to judge whether the model is performing as expected. This often combines with tuning the model to maximize some measurement (e.g., accuracy) or balance some tradeoff (e.g., precision-recall). Typical techniques in machine learning include holding out some data from model training to be used for validation and final testing, or iterating over the data set with varying held out sets. Choosing the appropriate metrics can involve ethical considerations: for example, in medical diagnosis one chooses metrics to minimize false positives with non-threatening diseases and costly follow-up, and to minimize false negatives with effective early treatment of serious or contagious disease. Minimizing one or the other is a choice that cannot be dictated by a mindless application of principles, but rather it is the result of ethical reflection and deliberation on the importance of capabilities, such as health agency.

Two tradeoffs between false positives and false negatives that occur frequently in healthcare model validation and tuning are precision-recall and sensitivity-specificity, defined in Table 1. In the precision-recall tradeoff, one either increases precision by reducing false positives *or* increases recall by reducing false negatives. In retrieving EHR of patients with cardiovascular disease, high precision reduces the number of patient records retrieved for patients who lack the disease, and high recall reduces the number of EHRs missed by the retrieval for patients who have the disease. Targeting the proper tradeoff depends upon sufficient medical domain knowledge to understand the relative consequences of the errors, technical skill to make appropriate tradeoff adjustments as desired, and moral judgment to select a proper tradeoff given the possible consequences to health agency. Does too high precision (low recall) mean individuals miss an opportunity for cheaper prescriptions or for a potentially life-saving intervention? Without microethics, data scientists must use their personal intuitions to evaluate tradeoffs with respect

Table 1. Metric Definitions for Model Validation

Term	**Formula**	**Denominator**
Sensitivity = Recall	$\frac{TP}{TP+FN}$	**Actually a positive**
Specificity	$\frac{TN}{TN+FP}$	**Actually a negative**
Precision	$\frac{TP}{TP+FP}$	**Classified as a positive**

to uninformative and general principles, such as beneficence and nonmaleficence. Attention to conversion factors, however, might raise awareness that for a financially impoverished patient pool, lowering drug costs 90% might reduce overall mortality more than a free electrocardiogram-enabled smart watch with high monthly wireless service bills.

The sensitivity-specificity tradeoff is similar to precision-recall, with sensitivity mathematically identical to recall, and specificity also increasing with reduced false positives (like precision). But unlike precision, specificity takes true negatives into account (instead of true positives). For instance, if the target population of a breast cancer screening tool is from an underserved area where receiving appropriate medical treatments may be challenging, then applying a less stringent metric for positives can be a way to make sure that patients have more time to seek medical counsel, and hence it is a way to empower health agency, even if the disease is not present. However, the other way around can be also true. A positive diagnosis may be an unbearable psychological burden, and hence one may choose to minimize false positives, if the SDoHs of the target population seems to suggest this.

Both precision-recall and sensitivity-specificity tradeoffs are important when dealing with low-prevalence disease, where a model to predict a disease occurring in 1:100,000 people that failed to predict any occurrence would be 99.999% accurate but have 0.0 recall and sensitivity. Precision works well in information retrieval to measure irrelevance when retrieving a few documents out of potentially millions, but specificity works better when modeling high-prevalence disease to identify healthy individuals (i.e., true negatives for the disease). Whether a data scientist predominantly considers the precision-recall or sensitivity-specificity tradeoff depends heavily upon training, e.g., in information retrieval or biostatistics, respectively. However, a microethical awareness of the relevant conversion factors in the sociotechnical systems can help the data scientist appropriately choose between two options, namely (i) measuring and reducing the irrelevant EHRs retrieved when searching for patients with diabetes (high precision) and, (ii) measuring and reducing the number of healthy patients identified when searching for patients with diabetes (high specificity). There are not established technical rationales to choose one metric tradeoff over the other, but the choice may affect others in the sociotechnical system through the model and the resulting intervention chosen, because the mathematical results for precision and specificity do differ and that affects their respective tradeoffs with recall/sensitivity.

Late Phase

Finally, the late phase involves 'communicating and deploying the results' and 'evaluating feedback' on the solution. Communicating the results may involve additional visualizations and presentations beyond what is needed for technical interpretation and considering the background of the intended audience and ways in which the results could be misconstrued. Visualizations are persuasive arguments and can affect sociotechnical systems even with minimal additional modeling. Although easy for a technically focused data scientist to consider a project 'done' when validated models and their results are interpreted and communicated, the late phase is essential for project success and causing real change in the sociotechnical system, i.e., being 'done done'. Evaluation includes examining if the data and modeling assumptions are valid *in situ*, and if the impact of the results have inadvertent consequences, including on any data being fed back into the system.

Decisions and insights from the data science process affect capabilities within the sociotechnical system, which can be amplified when the initiated changes feed back into the technical modeling. Data science, as a science, can go beyond building models of existing datasets and can actually experiment with the sociotechnical system, using subsequent data to evaluate the effects of those interventions. This has clear ethical consequences, especially in the healthcare domain, even if not considered a medical intervention. One might set up A/B tests (also called split tests) to compare predicted interventions to see if they have the predicted effect. For example, if a model predicts that transportation conversion factors are insufficient for medication adherence, then one could set up a prescription delivery service for a randomly selected set of patients predicted to have that issue and compare adherence with those remaining in a control group. Although the medical treatments do not differ between groups, a better model could lead to better medical outcomes for the targeted population, which has clear medical and ethical consequences. This effect occurs regardless of whether one explicitly considers transportation as a conversion factor affecting health agency, but including explicit moral awareness of capabilities means the data scientist can more readily ascertain the relationship between the technical modeling and metrics and the sociocultural factors affecting the system and response. Rather than limiting improvements to transportation based upon the data scientist's personal social context, she has the framework to think creatively about improving that conversion factor, such as whether providing bicycles would be an effective sociotechnical intervention.

4. CONCLUSION

In this paper, we argued a microethical framework incorporating moral awareness of capabilities and conversion factors can potentially overcome limitations of current ethical frameworks for data science. By conceptualizing ethics as an analysis of the connections between technical choices and sociocultural factors, we showed how a data scientist could gain insight into how these connections have ethical consequences for data subjects. This stimulates data scientists to understand how their tools can potentially shape the lives of data subjects. This approach promotes awareness of the ethical dimension of technical choices, and that can foster the cultivation of 'ethical skills' or, as known in philosophy, 'virtues'.

Organizing the data science process into discrete stages facilitates the identification of the potential connections between technical choices and sociocultural factors without introducing vague, high-level principles. Within the sociotechnical system of data scientist, data subject, computational tools, and systemic structures, a data scientist's technical choices to measure capabilities and conversion factors of the data subjects affects their health agency. These choices have social and ethical consequences, and they can also lead to better technical results. Technical benefits include more representative data acquisition, more insightful and focused data understanding and preparation, and more accurate and robust data analysis and modeling. In model validation and interpretation, awareness of sociocultural conversion factors and health determinants can help in navigating metric tradeoffs, such as within and between precision-recall and sensitivity-specificity. Without handles to bring social and moral consequences of technical choices into the data science process through explicit identification of capabilities, data scientists must fall back on their personal social context for interpreting features, rather than construct the needed features for modeling the relevant aspects of the sociotechnical system, such as how transportation might affect health agency. Explicit, and trained, moral awareness of capabilities and conversion factors focused on the microethical dimension of technical choices enables more beneficial social interventions and impact, as well as the cultivation of skills and development of good practices that make data scientists 'good' data scientists.

AUTHOR CONTRIBUTION STATEMENT

Authors have contributed equally.

REFERENCES

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 & Society*, *36*, 939–953. https://doi.org/10.1007/s00146-020-01112-w.

Conway, D. (2013, March 26). *The Data Science Venn Diagram*. http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Human Rights Watch, *The Toronto declaration: protecting the right to equality and nondiscrimination in machine learning systems* (2018). https:// www.hrw.org/news/2018/07/03/toronto-declaration-protecting-rightsequality-and-non-discrimination-machine.

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.

Komesaroff, P. A. (1995). From bioethics to microethics: Ethical debate and clinical medicine. In P. A. Komesaroff (Ed.), *Troubled Bodies—Critical Perspectives on Postmodernism, Medical Ethics, and the Body* (pp. 62–86). Duke University Press.

Liburd, L. C., Jack, L., Williams, S., & Tucker, P. (2005). Intervening on the Social Determinants of Cardiovascular Disease and Diabetes. *American Journal of Preventive Medicine*, *29*(5, Supplement 1), 18–24. https://doi.org/10.1016/j.amepre.2005.07.013.

Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020). Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376445.

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? *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*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5.

Nussbaum, M. C. (2006). *Frontiers of justice: Disability, nationality, species membership*. Belknap Press Cambridge, MA.

Oosterlaken, I. (2014). Human Capabilities in Design for Values. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 1–26). Springer Netherlands. https://doi.org/10.1007/978-94-007-6994-6_7-1.

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

Ratti, E., & Graves, M. (2021). Cultivating Moral Attention: A Virtue-Oriented Approach to Responsible Data Science in Healthcare. *Philosophy & Technology*, *34*(4), 1819–1846. https://doi.org/10.1007/s13347-021-00490-3.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598.

Srivastava, M., Heidari, H., & Krause, A. (2019). Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2459–2468. https://doi.org/10.1145/3292500.3330664.

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190498511.003.0001.

Véliz, C. (2019). Three things digital ethics can learn from medical ethics. *Nature Electronics*, *2*, 1–3. https://doi.org/10.1038/s41928-019-0294-2.

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford University Press.

World Health Organization. (2011). *Rio political declaration on social determinants of health*. World Health Organization. https://www.who.int/sdhconference/declaration/en/.