An Introduction to Data Ethics
MODULE AUTHOR:[1]
Shannon Vallor, Ph.D.
William J. Rewak, S.J. Professor of Philosophy, Santa Clara University

TABLE OF CONTENTS

---

[1] Thanks to Anna Lauren Hoffman and Irina Raicu for their very helpful comments on an early draft of this module.

An Introduction to Data Ethics

MODULE AUTHOR:

Shannon Vallor, Ph.D.
William J. Rewak, S.J. Professor of Philosophy, Santa Clara University

1. What do we mean when we talk about 'ethics'?

Ethics in the broadest sense refers to the concern that humans have always had for figuring out *how best to live.* The philosopher Socrates is quoted as saying in 399 B.C. that "the most important thing is not life, but the good life."[2] We would all like to avoid a bad life, one that is shameful and sad, fundamentally lacking in worthy achievements, unredeemed by love, kindness, beauty, friendship, courage, honor, joy, or grace. Yet what is the best way to obtain the opposite of this – a life that is not only *acceptable*, but even excellent and worthy of admiration? How do we identify a *good* life, one worth choosing from among all the different ways of living that lay open to us? This is the question that the study of ethics attempts to answer.

Today, the study of ethics can be found in many different places. As an academic field of study, it belongs primarily to the discipline of philosophy, where it is studied either on a **theoretical** level ('what is the best theory of the good life?') or on a **practical, applied level** as will be our focus ('how should we act in this or that situation, based upon our best theories of ethics?'). In community life, ethics is pursued through diverse cultural, religious, or regional/local ideals and practices, through which particular groups give their members guidance about how best to live. This *political* aspect of ethics introduces questions about power, justice, and responsibility. On a *personal* level, ethics can be found in an individual's moral reflection and continual strivings to become a better person. In *work* life, ethics is often formulated in formal codes or standards to which all members of a profession are held, such as those of medical or legal ethics. Professional ethics is also taught in dedicated courses, such as business ethics. It is important to recognize that the political, personal, and professional dimensions of ethics are not separate—they are interwoven and mutually influencing ways of seeking a good life with others.

2. What does ethics have to do with *technology*?

There is a growing international consensus that ethics is of increasing importance to education in technical fields, and that it must become part of the language that technologists are comfortable using. Today, the world's largest technical professional organization, IEEE (the Institute for Electrical and Electronics Engineers), has an entire division devoted just to **technology ethics**.[3] In 2014 IEEE began holding its own international conferences on ethics in engineering, science, and technology practice. To supplement its overarching professional code of ethics, IEEE is also working on **new ethical standards** in emerging areas such as AI, robotics, and data management.

What is *driving* this growing focus on technology ethics? What is the reasoning behind it? The basic rationale is really quite simple. Technology **increasingly shapes *how* human beings seek the good life**, and with what degree of success. Well-designed and well-used technologies can

---

[2] Plato, *Crito* 48b.

[3] https://techethics.ieee.org

make it easier for people to live well (for example, by allowing more efficient use and distribution of essential resources for a good life, such as food, water, energy, or medical care). Poorly designed or misused technologies can make it harder to live well (for example, by toxifying our environment, or by reinforcing unsafe, unhealthy or antisocial habits). **Technologies are not ethically 'neutral'**, for they reflect the values that we 'bake in' to them with our design choices, as well as the values which guide our distribution and use of them. Technologies both reveal and shape what humans value, what we think is 'good' in life and worth seeking.

Of course, this always been true; technology has never been separate from our ideas about the good life. We don't build or invest in a technology hoping it will make no one's life better, or hoping that it makes all our lives worse. **So what is new, then?** Why is ethics now such an important topic in technical contexts, more so than ever?

The answer has partly to do with the unprecedented **speeds, scales** and **pervasiveness with** which technical advances are transforming the social fabric of our lives, and the inability of regulators and lawmakers to keep up with these changes. Laws and regulations have historically been important instruments of preserving the good life within a society, but today they are being outpaced by the speed, scale, and complexity of new technological developments and their increasingly pervasive and hard-to-predict social impacts.

Additionally, many lawmakers lack the **technical expertise** needed to guide effective technology policy. This means that technical experts are increasingly called upon to help anticipate those social impacts and to think proactively about how their technical choices are likely to impact human lives. This means making ethical design and implementation choices in a dynamic, complex environment where the few legal 'handrails' that exist to guide those choices are often outdated and inadequate to safeguard public well-being.

For example: face- and voice-recognition algorithms can now be used to track and create a lasting digital record of your movements and actions in public, even in places where previously you would have felt more or less anonymous. There is no consistent legal framework governing this kind of data collection, even though such data could potentially be used to expose a person's medical history (by recording which medical and mental health facilities they visit), their religiosity (by recording how frequently they attend services and where), their status as a victim of violence (by recording visits to a victims services agency) or other sensitive information, up to and including the content of their personal conversations in the street.
What does a person given access to all that data, or tasked with analyzing it, need to understand about its ethical significance and power to affect a person's life?

Another factor driving the recent explosion of interest in technology ethics is the way in which 21st century **technologies are reshaping the global distribution of power, justice, and responsibility**. Companies such as Facebook, Google, Amazon, Apple, and Microsoft are now seen as having levels of global political influence comparable to, or in some cases greater than, that of states and nations. In the wake of revelations about the unexpected impact of social media and private data analytics on 2017 elections around the globe, the idea that technology companies can safely focus on profits alone, leaving the job of protecting the public interest wholly to government, is increasingly seen as naïve and potentially destructive to social flourishing.

Not only does technology greatly impact our opportunities for living a good life, but its **positive and negative impacts are often distributed unevenly** among individuals and groups. Technologies can create widely disparate impacts, creating 'winners' and 'losers' in the social lottery or magnifying existing inequalities, as when the life-enhancing benefits of a new technology are enjoyed only by citizens of wealthy nations while the life-degrading burdens of environmental contamination produced by its manufacture fall upon citizens of poorer nations. In other cases, technologies can help to create fairer and more just social arrangements, or create new access to means of living well, as when cheap, portable solar power is used to allow children in rural villages without electric power to learn to read and study after dark.

How do we ensure that access to the enormous benefits promised by new technologies, and exposure to their risks, are distributed in the right way? This is a question about technology *justice*. Justice is not only a matter of law, it is also even more fundamentally a matter of *ethics*.

3. What does ethics have to do with *data*?

'Data' refers to any form of recorded information, but today most of the data we use is recorded, stored, and accessed in digital form, whether as text, audio, video, still images, or other media. Networked societies generate an unending torrent of such data, through our interactions with our digital devices and a physical environment increasingly configured to read and record data about us. *Big Data* is a widely used label for the many new computing practices that depend upon this century's rapid expansion in the volume and scope of digitally recorded data that can be collected, stored, and analyzed. Thus **'big data' refers to more than just the existence and explosive growth of large digital datasets; it also refers to the new techniques, organizations, and processes that are necessary to transform large datasets into valuable human knowledge.** The big data phenomenon has been enabled by a wide range of computing innovations in data generation, mining, scraping, and sampling; artificial intelligence and machine learning; natural language and image processing; computer modeling and simulation; cloud computing and storage, and many others. Thanks to our increasingly sophisticated tools for turning large datasets into useful insights, new industries have sprung up around the production of various forms of **data analytics**, including predictive analytics and user analytics.

Ethical issues are everywhere in the world of data, because data's collection, analysis, transmission and use can and often does profoundly impact the ability of individuals and groups to live well.

For example, which of these life-impacting events, both positive and negative, might be the direct result of data practices?

A. Rosalina, a promising and hard-working law intern with a mountain of student debt and a young child to feed, is denied a promotion at work that would have given her a livable salary and a stable career path, even though her work record made her the objectively best candidate for the promotion.

B. John, a middle-aged father of four, is diagnosed with an inoperable, aggressive, and advanced brain tumor. Though a few decades ago his tumor would probably have been judged untreatable

and he would have been sent home to die, today he receives a customized treatment that in people with his very rare tumor gene variant, has a 75% chance of leading to full remission.

C. The Patels, a family of five living in an urban floodplain in India, receive several days advance warning of an imminent, epic storm that is almost certain to bring life-threatening floodwaters to their neighborhood. They and their neighbors now have sufficient time to gather their belongings and safely evacuate to higher ground.

D. By purchasing personal information from multiple data brokers operating in a largely unregulated commercial environment, **Peter**, a violent convict who was just paroled, is able to obtain a large volume of data about the movements of his ex-wife and stepchildren, who he was jailed for physically assaulting, and which a restraining order prevents him from contacting. Although his ex-wife and her children have changed their names, have no public social media accounts, and have made every effort to conceal their location from him, he is able to infer from his data purchases their new names, their likely home address, and the names of the schools his ex-wife's children now attend. They are never notified that he has purchased this information.

Which of these hypothetical cases raise ethical issues concerning data? The answer, as you probably have guessed, is **'All of them.'**

Rosalina's deserved promotion might have been denied because her law firm ranks employees using a poorly-designed predictive HR software package trained on data that reflects previous industry hiring and promotion biases against even the best-qualified women and minorities, thus **perpetuating the unjust bias**. As a result, especially if other employers in her field use similarly trained software, Rosalina might never achieve the economic security she needs to give her child the best chance for a good life, and her employer and its clients lose out on the promise of the company's best intern.

John's promising treatment plan might be the result of his doctors' use of an AI-driven diagnostic support system that can **identify rare, hard-to-find patterns** in a massive sea of cancer patient treatment data gathered from around the world, data that no human being could process or analyze in this way even if given an entire lifetime. As a result, instead of dying in his 40's, John has a great chance of living long enough to walk his daughters down the aisle at their weddings, enjoying retirement with his wife, and even surviving to see the birth of his grandchildren.

The Patels might owe their family's survival to advanced meterological data analytics software that allows for **much more accurate and precise disaster forecasting** than was ever possible before; local governments in their state are now able to predict with much greater confidence which cities and villages a storm is likely to hit and which neighborhoods are most likely to flood, and to what degree. Because it is often logistically impossible or dangerous to evacuate an entire city or region in advance of a flood, a decade ago the Patels and their neighbors would have had to watch and wait to see where the flooding will hit, and perhaps learn too late of their need to evacuate. But now, because these new data analytics allow officials to identify and evacuate only those neighborhoods that will be most severely affected, the Patels lives are saved from destruction.

Peter's ex-wife and her children might have their lives endangered by the absence of regulations on who can purchase and analyze personal data about them that they have not consented to make

public. Because the data brokers Peter sought out had no internal policy against the **sale of personal information** to violent felons, and because no law prevented them from making such a sale, Peter was able to get around every effort of his victims to evade his detection. And because there is no system in place allowing his ex-wife to be notified when someone purchases personal information about her or her children, or even a way for her to learn what data about her is available for sale and by whom, she and her children get no warning of the imminent threat that Peter now poses to their lives, and no chance to escape.

The combination of increasingly powerful but also potentially misleading or misused data analytics, a data-saturated and poorly regulated commercial environment, and the absence of widespread, well-designed standards for data practice in industry, university, non-profit, and government sectors has created a **'perfect storm' of ethical risks.** Managing those risks wisely requires understanding the vast potential for data to generate ethical benefits as well.

But this doesn't mean that we can just 'call it a wash' and go home, hoping that everything will somehow magically 'balance out.' Often, ethical choices do require accepting difficult trade-offs. But some risks are too great to ignore, and in any event, we don't want the result of our data practices to be a 'wash.' **We don't actually want the good and bad effects to *balance*!** Remember, the whole point of scientific and technical innovation is to make lives *better*, to maximize the human family's chances of living well and minimize the harms that can obstruct our access to good lives.

Developing a broader and better understanding of data ethics, especially among those who design and implement data tools and practices, is increasingly recognized as essential to meeting this **goal of beneficial data innovation and practice**.

This free module, developed at the Markkula Center for Applied Ethics at Santa Clara University in Silicon Valley, is one contribution to meeting this growing need. It provides an introduction to some key issues in data ethics, with working examples and questions for students that prompt active ethical reflection on the issues. Instructors and students using the module do not need to have any prior exposure to data ethics or ethical theory to use the module. However, this is only an introduction; **thinking about data ethics can begin here, but it should not *stop* here**. One big challenge for teaching data ethics is the immense territory the subject covers, given the ever-expanding variety of contexts in which data practices are used. Thus **no single set of ethical rules or guidelines will fit all data circumstances**; **ethical insights in data practice must be adapted to the needs of many kinds of data practitioners operating in different contexts.**

This is why many companies, universities, non-profit agencies, and professional societies whose members develop or rely upon data practices are funding an increasing number of their own data ethics-related programs and training tools. Links to many of these resources can be found in **Appendix A** to this module. **These resources can be used to build upon this introductory module and provide more detailed and targeted ethical insights for specific kinds of data practitioners.**

In the remaining sections of this module, you will have the opportunity to learn more about:

Part 1: The potential ethical harms and benefits presented by data

Part 2: Common ethical challenges faced by data professionals and users

Part 3: The nature and source of data professionals' ethical obligations to the public

Part 4: General frameworks for ethical thinking and reasoning

Part 5: Ethical 'best practices' for data practitioners

In each section of the module, you will be asked to fill in answers to specific questions and/or examine and respond to case studies that pertain to the section's key ideas. This will allow you to practice using all the tools for ethical analysis and decision-making that you will have acquired from the module.

PART ONE

What ethically significant harms and benefits can data present?

1. What makes a harm or benefit 'ethically significant'?

In the Introduction we saw that the 'good life' is what ethical action seeks to protect and promote. We'll say more later about the 'good life' and why we are ethically obligated to care about the lives of others beyond ourselves.

But for now, we can **define a harm or a benefit as 'ethically significant'** when it has a substantial possibility of making a difference to certain individuals' chances of having a good life, or the chances of a group to live well: that is, to flourish in society together. Some harms and benefits are not ethically significant. Say I prefer Coke to Pepsi. If I ask for a Coke and you hand me a Pepsi, even if I am disappointed, you haven't impacted my life in any ethically significant way. Some harms and benefits are too trivial to make a meaningful difference to how our life goes. Also, **ethics implies human choice**; a harm that is done to me by a wild tiger or a bolt of lightning might be very significant, but won't be ethically significant, for it's unreasonable to expect a tiger or a bolt of lightning to take my life or welfare into account. Ethics also requires **more than 'good intentions'**: many unethical choices have been made by persons who meant no harm, but caused great harm anyway, by acting with recklessness, negligence, bias, or blameworthy ignorance of relevant facts.[4]

In many technical contexts, such as the engineering, manufacture, and use of aeronautics, nuclear power containment structures, surgical devices, buildings, and bridges, it is very easy to see the ethically significant harms that can come from poor technical choices, and very easy to see the ethically significant benefits of choosing to follow the best technical practices known to us. All of these contexts present obvious issues of 'life or death' in practice; innocent people will die if

---

[4] Even acts performed without any direct intent, such as driving through a busy crosswalk while drunk, or unwittingly exposing sensitive user data to hackers, can involve ethical choice (e.g., the reckless choice to drink and get behind the wheel, or the negligent choice to use subpar data security tools)

we disregard public welfare and act negligently or irresponsibly, and people will generally enjoy better lives if we do things right.

Because 'doing things right' in these contexts preserves or even enhances the opportunities that other people have to enjoy a good life, **good technical practice in such contexts is also *ethical* practice.** A civil engineer who willfully or recklessly ignores a bridge design specification, resulting in the later collapse of said bridge and the deaths of a dozen people, is not just bad at his or her job. Such an engineer is also guilty of an *ethical failure*—and this would be true even if they just so happened to be shielded from legal, professional, or community punishment for the collapse.

In the context of data practice, the potential harms and benefits are no less real or ethically significant, up to and including matters of life and death. But due to the more complex, abstract, and often widely distributed nature of data practices, as well as the interplay of technical, social, and individual forces in data contexts, the harms and benefits of data can be **harder to see and anticipate**. This part of the module will help make them more recognizable, and hopefully, easier to anticipate as they relate to our choices.

2. What significant ethical benefits and harms are linked to data?

One way of thinking about benefits and harms is to understand what our **_life interests_** are; like all animals, humans have significant vital interests in food, water, air, shelter, and bodily integrity. But we also have strong life interests in our health, happiness, family, friendship, social reputation, liberty, autonomy, knowledge, privacy, economic security, respectful and fair treatment by others, education, meaningful work, and opportunities for leisure, play, entertainment, and creative and political expression, among other things.[5]

What is so powerful about data practice is that it has the potential to significantly impact all of these fundamental interests of human beings. In this respect, then, **data has a broader ethical sweep** than some of the stark examples of technical practice given earlier, such as the engineering of bridges and airplanes. Unethical design choices in building bridges and airplanes can destroy bodily integrity and health, and through such damage make it harder for people to flourish, but unethical choices in the use of data can cause many more different kinds of harm. While selling my personal data to the wrong person could in certain scenarios cost me my life, as we noted in the Introduction, mishandling my data could also leave my body physically intact but my reputation, savings, or liberty destroyed. Ethical uses of data can also generate a vast range of benefits for society, from better educational outcomes and improved health to expanded economic security and fairer institutional decisions.

Because of the massive scope of social systems that data touches, and the difficulty of anticipating what might be done *by* or *to* others with the data we handle, **data practitioners must confront a far more complex ethical landscape** than many other kinds of technical professionals, such as civil and mechanical engineers, who might limit their attention to a narrow range of goods such as public safety and efficiency.

---

[5] See Robeyns (2016) https://plato.stanford.edu/entries/capability-approach/) for a helpful overview of the highly influential capabilities approach to identifying these fundamental interests in human life.

ETHICALLY SIGNIFICANT BENEFITS OF DATA PRACTICES

The most common benefits of data are typically easier to understand and anticipate than the potential harms, so we will go through these fairly quickly:

1. HUMAN UNDERSTANDING: Because data and its associated practices can uncover previously unrecognized correlations and patterns in the world, data can greatly enrich our understanding of ethically significant relationships—in nature, society, and our personal lives. Understanding the world is good in itself, but also, **the more we understand about the world and how it works, the more intelligently we can act in it.** Data can help us to better understand how complex systems interact at a variety of scales: from large systems such as weather, climate, markets, transportation, and communication networks, to smaller systems such as those of the human body, a particular ecological niche, or a specific political community, down to the systems that govern matter and energy at subatomic levels. Data practice can also shed new light on previously unseen or unattended harms, needs, and risks. For example, big data practices can reveal that a minority or marginalized group is being harmed by a drug or an educational technique that was originally designed for and tested only on a majority/dominant group, allowing us to innovate in safer and more effective ways that bring more benefit to a wider range of people.

2. SOCIAL, INSTITUTIONAL, AND ECONOMIC EFFICIENCY: Once we have a more accurate picture of how the world works, we can design or intervene in its systems to improve their functioning. This **reduces wasted effort and resources and improves the alignment between a social system or institution's policies/processes and our goals**. For example, big data can help us create better models of systems such as regional traffic flows, and with such models we can more easily identify the specific changes that are most likely to ease traffic congestion and reduce pollution and fuel use—ethically significant gains that can improve our happiness and the environment. Data used to better model voting behavior in a given community could allow us to identify the distribution of polling station locations and hours that would best encourage voter turnout, promoting ethically significant values such as citizen engagement. Data analytics can search for complex patterns indicating fraud or abuse of social systems. The potential efficiencies of big data go well beyond these examples, enabling social action that streamlines access to a wide range of ethically significant goods such as health, happiness, safety, security, education, and justice.

3. PREDICTIVE ACCURACY AND PERSONALIZATION: Not only can good data practices help to make social systems work more efficiently, as we saw above, but they can also used to **more precisely tailor actions to be effective in achieving good outcomes for *specific individuals, groups, and circumstances,*** and to be **more responsive to user input in (approximately)** *real time*. Of course, perhaps the most well-known examples of this advantage of data involves personalized search and serving of advertisements. Designers of search engines, online advertising platforms, and related tools want the content they deliver to you to be the most relevant to you, *now*. Data analytics allow them to predict *your* interests and needs with greater accuracy. But it is important to recognize that the predictive potential of data goes well beyond this familiar use, enabling personalized and targeted interactions that can deliver many kinds of ethically significant goods. From targeted disease therapies in medicine that are tailored specifically to a patient's genetic fingerprint, to customized homework assignments that build upon an individual student's existing skills and focus on practice in areas of weakness, to

predictive policing strategies that send officers to the specific locations where crimes are most likely to occur, to timely predictions of mechanical failure or natural disaster, a key goal of data practice is to more accurately fit our actions to specific needs and circumstances, rather than relying on more sweeping and less reliable generalizations. In this way the choices we make in seeking the good life for ourselves and others can be more effective more often, and for more people.

ETHICALLY SIGNIFICANT HARMS OF DATA PRACTICES

Alongside the ethically significant benefits of data are ways in which data practice can be harmful to our chances of living well. Here are some key ones:

1. HARMS TO PRIVACY & SECURITY: Thanks to the ocean of personal data that humans are generating today (or, to use a better metaphor, the many different lakes, springs, and rivers of personal data that are pooling and flowing across the digital landscape), most of us do not realize how exposed our lives are, or can be, by common data practices.

Even *anonymized* datasets can, when linked or merged with other datasets, reveal intimate facts (or in many cases, *falsehoods*) about us. As a result of your multitude of data-generating activities (and of those you interact with), your sexual history and preferences, medical and mental health history, private conversations at work and at home, genetic makeup and predispositions, reading and Internet search habits, political and religious views, may all be part of data profiles that have been constructed and stored somewhere unknown to you, often **without your knowledge or informed consent.** Such profiles exist within a **chaotic data ecosystem** that gives individuals little to no ability to personally curate, delete, correct, or control the release of that information. Only thin, regionally inconsistent, and weakly enforced sets of data regulations and policies protect us from the **reputational, economic, and emotional harms** that release of such intimate data into the wrong hands could cause. In some cases, as with data identifying victims of domestic violence, or political protestors or sexual minorities living under oppressive regimes, the potential **harms can even be fatal**.

And of course, this level of exposure does not just affect *you* but virtually everyone in a networked society. Even those who choose to live 'off the digital grid' cannot prevent intimate data about them from being generated and shared by their friends, family, employers, clients, and service providers. Moreover, **much of this data does not stay confined to the digital context in which it was originally shared**. For example, information about an online purchase you made in college of a politically controversial novel might, without your knowledge, be sold to third-parties (and then sold again), or hacked from an insecure cloud storage system, and eventually included in a digital profile of you that years later, a prospective employer or investigative journalist could purchase. Should you, and others, be able to protect your employability or reputation from being irreparably harmed by such data flows? **Data privacy isn't just about our online activities, either.** Facial, gait, and voice-recognition algorithms, as well as geocoded mobile data, can now identify and gather information about us as we move and act in many public and private spaces.

Unethical or ethically negligent data privacy practices, from poor data security and data hygiene, to unjustifiably intrusive data collection and data mining, to reckless selling of user data to third-parties, can expose others to profound and unnecessary harms. **In Part Two of this module,**

we'll discuss the specific challenges that avoiding privacy harms presents for data practitioners, and explore possible tools and solutions.

2. HARMS TO FAIRNESS AND JUSTICE: We all have a significant life interest in being judged and treated fairly, whether it involves how we are treated by law enforcement and the criminal and civil court systems, how we are evaluated by our employers and teachers, the quality of health care and other services we receive, or how financial institutions and insurers treat us.

All of these systems are being radically transformed by new data practices and analytics, and the preliminary evidence suggests that the values of fairness and justice are too often endangered by poor design and use of such practices. The most common causes of such harms are: **arbitrariness**; **avoidable errors and inaccuracies**; and **unjust and often hidden biases** in datasets and data practices.

For example, investigative journalists have found compelling evidence of hidden racial bias in data-driven predictive algorithms used by parole judges to assess convicts' risk of reoffending.[6] Of course, bias is not always harmful, unfair, or unjust. A bias against, for example, convicted bank robbers when reviewing job applications for an armored-car driver is entirely reasonable! But **biases that rest on falsehoods, sampling errors**, **and unjustifiable discriminatory practices** are all too common in data practice.

Typically, such biases are not explicit, but *implicit* in the data or data practice, and thus harder to see. For example, in the case involving racial bias in criminal risk-predictive algorithms cited above, the race of the offender was not in fact a label or coded variable in the system used to assign the risk score. The racial bias in the outcomes was not intentionally placed there, but rather 'absorbed' from the racially-biased data the system was trained on. We use the term 'proxies' to describe how data that are not explicitly labeled by race, gender, location, age, etc. can still function as *indirect but powerful indicators* of those properties, especially when combined with other pieces of data. A very simple example is the function of a zip code as a strong proxy, in many neighborhoods, for race or income. So, a risk-predicting algorithm could generate a racially-biased prediction about you even if it is never 'told' your race. This makes the bias no less harmful or unjust; a criminal risk algorithm that inflates the *actual* risk presented by black defendants relative to otherwise similar white defendants leads to judicial decisions that are *wrong*, both factually and morally, and profoundly harmful to those who are misclassified as high-risk. If anything, implicit data bias is *more* dangerous and harmful than explicit bias, since it can be more challenging to expose and purge from the dataset or data practice.

In other data practices the harms are driven not by bias, but by **poor quality, mislabeled, or error-riddled data (i.e., 'garbage in, garbage out')**; **inadequate design and testing of data analytics**; or **a lack of careful training and auditing** to ensure the correct implementation and use of the data system. For example, such flawed data practices by a state Medicaid agency in Idaho led it to make large, arbitrary, and very possibly unconstitutional cuts in disability benefit payments to over 4,000 of its most vulnerable citizens.[7] In Michigan, flawed data practices led

---

[6] See the *ProPublica* series on 'Machine Bias' published by Angwin et. al. (2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[7] See Stanley (2017) https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case

another agency to levy false fraud accusations and heavy fines against at least 44,000 of its innocent, unemployed citizens for two years. It was later learned that its data-driven decision-support system had been operating at a shockingly high false-positive error rate of *93 percent*.[8]

While not all such cases will involve datasets on the scale typically associated with 'big data', they all involve ethically negligent failures to adequately design, implement and audit data practices to promote fair and just results. Such failures of ethical data practice, whether in the use of small datasets or the power of 'big data' analytics, **can and do result in economic devastation, psychological, reputational, and health damage, and for some victims, even the loss of their physical freedom.**

3. HARMS TO TRANSPARENCY AND AUTONOMY: In this context, *transparency* is the **ability to see how a given social system or institution works**, and to be able to inquire about the basis of life-affecting decisions made within that system or institution. So, for example, if your bank denies your application for a home loan, transparency will be served by you having access to information about exactly *why* you were denied the loan, and by whom.

Autonomy is a distinct but related concept; *autonomy* refers to one's **ability to govern or steer the course of one's own life**. If you lack autonomy altogether, then you have no ability to control the outcome of your life and are reliant on sheer luck. The more autonomy you have, the more your chances for a good life depend on your own choices.

The **two concepts are related** in this way; to be effective at steering the course of my own life (to be autonomous), I must have a certain amount of accurate information about the other forces acting upon me in my social environment (that is, I need some transparency in the workings of my society). Consider the example given above: if I know why I was denied the loan (for example, a high debt-to-asset ratio), I can figure out what I need to change to be successful in a new application, or in an application to another bank. The fate of my aspiration to home ownership remains at least somewhat in my control. But if I have no information to go on, then I am blind to the social forces blocking my aspiration, and have no clear way to navigate around them. Data practices have the potential to create or diminish social transparency, but **diminished transparency is currently the greater risk** because of two factors.

The **first risk factor** has to do with the sheer volume and complexity of today's data, and of the algorithmic techniques driving big data practices. For example, machine learning algorithms trained on large datasets can be used to make new assessments based on fresh data; that is why they are so useful. The problem is that especially with 'deep learning' algorithms, it can be difficult or impossible to reconstruct the machine's 'reasoning' behind any particular judgment.[9] This means that if my loan was denied on the basis of this algorithm, the loan officer and even the system's programmers might be unable to tell me why—even if they wanted to. And it is

---

[8] See Egan (2017) http://www.freep.com/story/news/local/michigan/2017/07/30/fraud-charges-unemployment-jobless-claimants/516332001/ and Levin (2016) https://levin.house.gov/press-release/state%E2%80%99s-automated-fraud-system-wrong-93-reviewed-unemployment-cases-2013-2105 For discussion of the broader issues presented by these cases of bias in institutional data practice see Cassel (2017) https://thenewstack.io/when-ai-is-biased/

[9] See Knight (2017) https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ for a discussion of this problem and its social and ethical implications.

unclear how I would appeal such an opaque machine judgment, since I lack the information needed to challenge its basis. In this way my autonomy is restricted. Because of the lack of transparency, my choices in responding to a life-affecting social judgment about me have been severely limited.

The **second risk factor** is that often, data practices are cloaked behind trade secrets and proprietary technology, including proprietary software. While laws protecting intellectual property are necessary, they can also impede social transparency when the protected property (the technique or invention) is a key part of the mechanisms of social functioning. These competing interests in intellectual property rights and social transparency need to be appropriately balanced. In some cases the courts will decide, as they did in the aforementioned Idaho case. In that case, K.W. v. Armstrong, a federal court ruled that citizens' due process was violated when, upon requesting the reason for the cuts to their disability benefits, the citizens were told that trade secrets prevented releasing that information.[10] Among the remedies ordered by the court was a testing regime to ensure the reliability and accuracy of the automated decision-support systems used by the state.

However, not every obstacle to data transparency can or should be litigated in the courts. **Securing an ethically appropriate measure of social transparency in data practices will require considerable public discussion and negotiation, as well as good faith efforts by data practitioners to respect the ethically significant interest in transparency.**

You now have an overview of many common and significant ethical issues raised by data practices. But the scope of these issues is by no means limited to those in Part One. **Data practitioners need to be attentive to the many ways in which data practices can significantly impact the quality of people's lives**, and must learn to better anticipate their potential harms and benefits so that they can be effectively addressed.

Now, you will get some practice in doing this yourself.

Case Study 1

> **Fred and Tamara**, a married couple in their 30's, are applying for a business loan to help them realize their long-held dream of owning and operating their own restaurant. Fred is a highly promising graduate of a prestigious culinary school, and Tamara is an accomplished accountant. They share a strong entrepreneurial desire to be 'their own bosses' and to bring something new and wonderful to their local culinary scene; outside consultants have reviewed their business plan and assured them that they have a very promising and creative restaurant concept and the skills needed to implement it successfully. The consultants tell them they should have no problem getting a loan to get the business off the ground.
>
> For evaluating loan applications, Fred and Tamara's local bank loan officer relies on an off-the-shelf software package that synthesizes a wide range of data profiles purchased from hundreds of private data brokers. As a result, it has access to information about Fred and Tamara's lives that goes well beyond what they were asked to disclose on their loan application. Some of this information is clearly relevant to the application, such as their on-time bill payment history. But

---

[10] See Morales (2016) https://www.acluidaho.org/en/news/federal-court-rules-against-idaho-department-health-and-welfare-medicaid-class-action

a lot of the data used by the system's algorithms is of the sort that no human loan officer would normally think to look at, or have access to—including inferences from their drugstore purchases about their likely medical histories, information from online genetic registries about health risk factors in their extended families, data about the books they read and the movies they watch, and inferences about their racial background. Much of the information is accurate, but some of it is not.

A few days after they apply, Fred and Tamara get a call from the loan officer saying their loan was not approved. When they ask why, they are told simply that the loan system rated them as 'moderate-to-high risk.' When they ask for more information, the loan officer says he doesn't have any, and that the software company that built their loan system will not reveal any specifics about the proprietary algorithm or the data sources it draws from, or whether that data was even validated. In fact, they are told, not even the system's designers know how what data led it to reach any particular result; all they can say is that statistically speaking, the system is 'generally' reliable. Fred and Tamara ask if they can appeal the decision, but they are told that there is no means of appeal, since the system will simply process their application again using the same algorithm and data, and will reach the same result.

Question 1.1:

What ethically significant harms, as defined in Part One, might Fred and Tamara have suffered as a result of their loan denial? (Make your answers as full as possible; identify as many kinds of possible harm done to their significant life interests as you can think of).

Question 1.2:
What sort of ethically significant benefits, as defined in Part One, could come from banks using a big-data driven system to evaluate loan applications?

Question 1.3:
Beyond the impacts on Fred and Tamara's lives, what broader harms to society could result from the widespread use of this particular loan evaluation process?

Question 1.4:
Could the harms you listed in 1.1 and 1.3 have been anticipated by the loan officer, the bank's managers, and/or the software system's designers and marketers? Should they have been anticipated, and why or why not?

Question 1.5:
What measures could the loan officer, the bank's managers, or the employees of the software company have taken to lessen or prevent those harms?

PART TWO

Common ethical challenges for data practitioners and users

We saw in Part One that a broad range of ethically significant harms and benefits to individuals, and to society, are associated with data practices. Here in Part Two, we will see how those harms and benefits relate to eight types of common practical challenges encountered by data practitioners and users. **Even when a data practice is legal, it may not be *ethical*, and unethical data practices can result in significant harm and reputational damage to users, companies, and data practitioners alike.** These are the just some of the common challenges that we must prepared to address through the ethical 'best practices' we will summarize in Part Five. They have been framed as **questions**, since these are the questions that data practitioners and users will frequently need to ask themselves in real-world data contexts, in order to promote ethical data practice.

These questions may apply to **data practitioners in a variety of roles and contexts**, for example: an individual researcher in academia, government, non-profit sector, or commercial industry; members of research teams; app, website, or Internet platform designers or team members; organizational data managers or team leaders; chief privacy officers, and so on. Likewise, **data subjects** (the sharers, owners, or generators of the data) may be found in a similar range of roles and contexts.

1. ETHICAL CHALLENGES IN APPROPRIATE DATA COLLECTION AND USE:

How can we properly acknowledge and respect the *purpose* for, and *context* within which, certain data was shared with us or generated for us? (For example, if the original owner or source of a body of personal data shared it with me for the explicit purpose of aiding my medical research program, may I then sell that data to a data broker who may sell it for any number of non-medical commercial purposes?)[11]

How can we avoid *unwarranted* or *indiscriminate* data collection—that is, collecting *more* data than is justified in a particular context? When is it ethical to *scrape* websites for public data, and does it depend on the purpose for which the data is to be used?

Have we adequately considered the ethical implications of selling or sharing subjects' data with *third-parties*? Do we have a clear and consistent policy outlining the circumstances in which data will leave our control, and do we honor that policy? Have we thought about *who* those third-parties are, and the very different risks and advantages of making data open to the public vs. putting it in the hands of a private data broker or other commercial entity?

Have we given data subjects appropriate forms of *choice* in data sharing? For example, have we favored opt-in or opt-out privacy settings, and have we determined whether those settings are reasonable and ethically justified?

---

[11] Helen Nissenbaum's 2009 book *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Palo Alto, Stanford University Press) is especially relevant to this challenge.

Are data subjects 'boxed in' by the circumstances in which they are asked to share data, or do they have clear and acceptable alternatives? Are unreasonable or punitive costs (in inconvenience, loss of time, or loss of functionality) imposed on subjects who decline to share their data?

Are the terms of our data policy laid out in a *clear, direct, and understandable way*, and made *accessible* to all data subjects? Or are they full of unnecessarily legalistic or technical jargon, obfuscating generalizations and evasions, or ambiguous, vague, misleading and disingenuous claims? Does the design of our interface encourage careful reading of the data policy, or a 'click-through' response?

Are data subjects given clear paths to obtaining *more* information or context for a data practice? (For example, buttons such as: 'Why am I seeing this ad?'; 'Why am I being asked for this information?'; 'How will my data be secured?'; 'How do I disable sharing'?)

Are data subjects being *appropriately compensated* for the benefits/value of their data? If the data subjects are not being compensated monetarily, then what service or value does the data subject get in return? Would our data subjects agree to this data collection and use if they understood as much about the context of the interaction as we do, or would they likely feel exploited or taken advantage of?

Have we considered what *control or rights* our data subjects should *retain* over their data? Should they be able to withdraw, correct, or update the data later if they choose? Will it be technically feasible for the data to be deleted, corrected, or updated later, and if not, is the data subject fully aware of this and the associated risks? Who should *own* the shared data and hold rights over its transfer and commercial and noncommercial use?

2. DATA STORAGE, SECURITY AND RESPONSIBLE DATA STEWARDSHIP:

How can we *responsibly and safely store* personally identifying information? Are data subjects given clear and accurate information about our terms of storage? Is it clear which members of our organization are responsible for which aspects of our data stewardship?

Have we *reflected on the ethical harms* that may be done by a data breach, both in the short-term and long-term, and to whom? Are we taking into account the significant interests of *all* stakeholders who may be affected, or have we overlooked some of these?

What are our *concrete action plans for the worst-case-scenarios*, including mitigation strategies to limit or remedy harms to others if our data stewardship plan goes wrong?

Have we made *appropriate investments in our data security/storage infrastructure* (relative to our context and the potential risks and harms)? Or have we endangered data subjects or other parties by allocating insufficient resources to these needs, or contracting with unreliable/low-quality data storage and security vendors?

What *privacy-preserving techniques* such as data anonymization, obfuscation, and differential privacy do we rely upon, and what are their various advantages and limitations? Have we invested appropriate resources in maintaining the most appropriate and effective privacy-preserving techniques for our data context? Are we keeping up-to-date on the evolving vulnerabilities of existing privacy-preserving techniques, and updating our practices accordingly?

What are the *ethical risks of long-term data storage*? How long we are justified in keeping sensitive data, and when/how often should it be purged? (Either at a data subject's request, or for security purposes). Do we have a data deletion/destruction plan in place?

Do we have an end-to-end *plan for the lifecycle of the data* we collect or use, and do we regularly examine that plan to see if it needs to be improved or updated?

What measures should we have in place to allow data to be deleted, corrected, or updated by affected/interested parties? How can we best ensure that those measures are communicated to or easily accessible by affected/interested parties?

3. DATA HYGIENE AND DATA RELEVANCE

How '*dirty*' (inaccurate, inconsistent, incomplete, or unreliable) is our data, and how do we know? Is our data clean 'enough' to be effective and beneficial for our purposes? Have we established **what *significant harms* 'dirty' data** in our practice could do to others?

What are our practices and procedures for *validation* and *auditing* of data in our context, to ensure that the data conform to the necessary constraints of our data practice?

How do we establish proper *parsing and consistency* of data field labels, especially when integrating data from different sources/systems/platforms? How do we ensure the ***integrity*** of our data across transfer/conversion/transformation operations?

What are our established tools and practices for *scrubbing* dirty data, and what are the risks and limitations of those scrubbing techniques?

Have we considered the *diversity* of the data sources and/or training datasets we use, ensuring that they are appropriately reflective of the population we are using it to produce insights about? (For example, does our health care analytics software rely upon training data sourced from medical studies in which white males were vastly overrepresented?)

Is our data appropriately *relevant* to the problem it will be used to solve, or the nature of the judgments it will be used to support?

How long is this data likely to *remain* accurate, useful or relevant? What is our plan for replacing/refreshing datasets that have become out-of-date?

4. IDENTIFYING AND ADDRESSING ETHICALLY HARMFUL DATA BIAS

What inaccurate, unjustified, or otherwise harmful human biases are reflected in our data? Are these data explicit in our data or implicit? What is our plan for identifying, auditing, eliminating, offsetting or otherwise effectively responding to harmful data bias?

Have we distinguished carefully between the forms of bias we should *want* to be reflected in our data or application, and those that are harmful or otherwise unwarranted? What practices will serve us well in anticipating and addressing the latter?

Have we sufficiently understood how this bias could do harm, and to whom? Or have we perhaps ignored or minimized the harms, or failed to see them at all due to a lack of moral imagination and perspective, or due to a desire not to think about the risks of our practice?

How might harmful or unwarranted bias in our data get *magnified, transmitted, obscured, or perpetuated* **by our use of it?** What methods do we have in place to prevent such effects of our practice?

5. VALIDATION AND TESTING OF DATA MODELS & ANALYTICS

How can we ensure that we have adequately tested our analytics/data models to validate their performance, especially 'in the wild' (against 'real-world' data)?

Have we fully considered the ethical harms that may be caused by inadequate validation and testing, or have we allowed a rush to production or customer pressures to affect our judgment of these risks?

What distinctive ethical challenges might arise as a result of the lack of transparency in 'deep-learning' or any other opaque, 'black-box' techniques driving our analytics?

How can we test our data analytics and models to **ensure their reliability across new, unexpected contexts?** Have we anticipated circumstances in which our analytics might **get used in contexts or to solve problems for which they were not designed**, and the ethical harms that might result for such 'off-label' uses or abuses? Have we identified measures to limit the harmful effects of such uses?

In what cases might we be ethically obligated to ensure that the **results, applications, or other consequences of our analytics are audited for disparate and unjust outcomes**? How will we respond if our systems or practices are accused by others of leading to such outcomes, or other social harms?

6. HUMAN ACCOUNTABILITY IN DATA PRACTICES AND SYSTEMS

Who will be designated as responsible for each aspect of ethical data practice, if I am involved in a group or team of data practitioners? How will we avoid a scenario where ethical data practice is a high-level goal of the team or organization, but no specific individuals are made responsible for taking action to help the group achieve that goal?

Who should and will be held accountable for various harms that might be caused by our data or data practice? How will we avoid the 'problem of many hands,' where no one is held accountable for the harmful outcomes of a practice to which many contributed?

Have we **established effective organizational or team practices and policies** for safeguarding/promoting ethical benefits, and anticipating, preventing and remedying possible ethical harms, of our data practice? (For example: 'premortem' and 'postmortem' exercises as a form of 'data disaster planning' and learning from mistakes).

Do we have a clear and effective process for any harmful outcomes of our data practice to be surfaced and investigated? Or do our procedures, norms, incentives, and group/team culture make it likely that such harms will be ignored or swept under the rug?

What processes should we have in place to allow an affected party to appeal the result or challenge the use of a data practice? Is there an established process for correction, repair, and iterative improvement of a data practice?

To what extent should our data systems and practices be open for public inspection and comment? Beyond ourselves, to whom are we *responsible* for what we do? How do our responsibilities to a broad 'public' differ from our responsibilities to the specific populations most impacted by our data practices?

7. EFFECTIVE CUSTOMER/USER TRAINING IN USE OF DATA AND ANALYTICS

Have we placed data tools in appropriately skilled and responsible hands, with appropriate levels of instruction and training? Or do we sell data or analytics 'off the shelf' with no follow-up, support, or guidance? What harms can result from inadequate instruction and training (of data users, clients, customers, etc.)

Are our data customers/users given an accurate view of the *limits* and *proper use* of the data, data practice or system we offer, not just its potential power? Or are we taking advantage of or perpetuating 'big data hype' to sell inappropriate technology?

8. UNDERSTANDING PERSONAL, SOCIAL, AND BUSINESS IMPACTS OF DATA PRACTICE

Overall, have we fully considered how our data/data practice or system will be used, and how it might impact data subjects or other parties later on? Are the relevant decision-making teams developing or using this data/data practice **sufficiently diverse to understand and anticipate its effects?** Or might we be ignoring or minimizing the effects on people or groups unlike ourselves?

Has sufficient input been gathered from other stakeholders who might represent very different interests/values/experiences from ours?

Has the testing of the practice taken into account how its impact might vary across a variety of individuals, identities, cultures and interest groups?

Does the collection or use of this data **violate anyone's legal or *moral* rights**, **limit their fundamental human *capabilities***, or otherwise **damage their fundamental *life interests*?** Does the data practice in any way **impinge on the *autonomy* or *dignity*** of other moral agents? Is the data practice likely to **damage or interfere with the *moral and intellectual habits, values,* or *character development*** of any affected parties or users?

Would information about this data practice be morally or socially controversial or damaging to professional reputation of those involved if widely known and understood? Is it consistent with the organization's image and professed values? Or is it a PR disaster waiting to happen, and if so, why is it being done?

CASE STUDY *2*

> In 2014 it was learned that Facebook had been experimenting on its own users' emotional manipulability, by altering the news feeds of almost 700,000 users to see whether Facebook engineers placing more positive or negative content in those feeds could create effects of positive or negative 'emotional contagion' that would spread between users. Facebook's published study,

which concluded that such emotional contagion could be induced via social networks on a "massive scale," was highly controversial, since the affected users were unaware that they were the subjects of a scientific experiment, or that their news feed was being used to manipulate their emotions and moods.[12]

Facebook's Data Use Policy, which users must agree to before creating an account, did not include the phrase "constituting informed consent for research" until four months after the study concluded. However, the company argued that their activities were still covered by the earlier data policy wording, even without the explicit reference to 'research.'[13] Facebook also argued that the purpose of the study was consistent with the user agreement, namely, to give Facebook knowledge it needs to provide users with a positive experience on the platform.

Critics objected on several grounds, claiming that:

A) Facebook violated long-held standards for ethical scientific research in the U.S. and Europe, which require *specific* and *explicit* informed consent from human research subjects involved in medical or psychological studies;

B) That such informed consent should not in any case be implied by agreements to a generic Data Use Policy that few users are known to carefully read or understand;

C) That Facebook abused users' trust by using their online data-sharing activities for an *undisclosed and unexpected purpose*;

D) That the researchers seemingly ignored the *specific harms* to people that can come from emotional manipulation. For example, thousands of the 689,000 study subjects almost certainly suffer from clinical depression, anxiety, or bipolar disorder, but were not excluded from the study by those higher risk factors. The study lacked key mechanisms of research ethics that are commonly used to minimize the potential emotional harms of such a study, for example, a mechanism for debriefing unwitting subjects after the study concludes, or a mechanism to exclude participants under the age of 18 (another population especially vulnerable to emotional volatility).

On the next page, you'll answer some questions about this case study. Your answers should highlight connections between the case and the content of Part Two.

---

[12] Kramer, Guillory, and Hancock (2014); see http://www.pnas.org/content/111/24/8788.full

[13] https://www.forbes.com/sites/kashmirhill/2014/06/30/facebook-only-got-permission-to-do-research-on-users-after-emotion-manipulation-study/#f0b433a7a62d

Question 2.1: Of the eight types of ethical challenges for data practitioners that we listed in Part Two, which two types are most relevant to the Facebook emotional contagion study? Briefly explain your answer.

Question 2.2: Were Facebook's users justified and reasonable in reacting negatively to the news of the study? Was the study *ethical*? Why or why not?

Question 2.3: To what extent should those involved in the Facebook study have *anticipated* that the study might be ethically controversial, causing a flood of damaging media coverage and angry public commentary? If the negative reaction *should* have been anticipated by Facebook researchers and management, why do you think it *wasn't?*

Question 2.4: Describe *2* or *3* things Facebook could have done differently, to acquire the benefits of the study in a less harmful, less reputationally damaging, and more ethical way.

Question 2.5: Who is morally *accountable* for any harms caused by the study? Within a large organization like Facebook, how should *responsibility* for preventing unethical data conduct be distributed, and why might that be a challenge to figure out?

CASE STUDY 3

In a widely cited 2016 study, computer scientists from Princeton University and the University of Bath demonstrated that significant harmful racial and gender biases are consistently reflected in the performance of learning algorithms commonly used in natural language processing tasks to represent the relationships between meanings of words.[14]

For example, one of the tools they studied, GloVe (Global Vectors for Word Representation), is a learning algorithm for creating word embeddings—visual maps that represent similarities and associations among word meanings in terms of distance between vectors.[15] Thus the vectors for the words 'water' and 'rain' would appear much closer together than will the vectors for the terms 'water' and 'red.' As with other similar data models for natural language processing, when GloVe is trained on a body of text from the Web, it learns to reflect in its own outputs "accurate imprints of [human] historic biases" (Caliskan-Islam, Bryson, and Naryanan, 2016). Some of these biases are based in objective reality (like our 'water' and 'rain example above). Others reflect subjective values that are (for the most part) morally neutral—for example, names for flowers (rose, lilac, tulip) are much more strongly associated with pleasant words (such as freedom, honest, miracle, and lucky), whereas names for insects (ant, beetle, hornet) are much more strongly associated (have nearer vectors) with unpleasant words (such as filth, poison, and rotten.)

---

[14] Caliskan-Islam, Bryson, & Narayanan (2016); see https://motherboard.vice.com/en_us/article/z43qka/its-our-fault-that-ai-thinks-white-names-are-more-pleasant-than-black-names

[15] See Pennington (2014) https://nlp.stanford.edu/projects/glove/

However, other biases in the data models, especially those concerning race and gender, are neither objective nor harmless. As it turns out, for example, common *European American* names such as Ryan, Jack, Amanda, and Sarah were far more closely associated in the model with the pleasant terms (such as joy, peace, wonderful, and friend), while common *African American* names such as Tyrone, Darnell, and Keisha were far more likely to be associated with the unpleasant terms (such as terrible, nasty, and failure).

Common names for men were also much more closely associated with career related words such as 'salary' and 'management' than for women, whose names were more closely associated with domestic words such as 'home' and 'relatives.' Career and educational stereotypes by gender were also strongly reflected in the model's output. The study's authors note that this is not a deficit of a particular tool, such as GloVe, but a pervasive problem across many data models and tools trained on a corpus of human language use. Because people are (and have long been) biased in harmful and unjust ways, data models that learn from human output will carry those harmful biases forward. Often the human biases are actually concentrated or *amplified* by the data model.

Does it raise ethical concerns that biased tools are used to drive many tasks in big data analytics, from sentiment analysis (e.g., determining whether an interaction with a customer is pleasant), to hiring solutions (e.g., ranking resumes), to ad service and search (e.g., showing you customized content), to social robotics (understanding and responding appropriately to humans in a social setting) and many other applications? **Yes**.

On this page, you'll answer some questions about this case. Your answers should make connections between case study 3 and the content of Part Two.

Question 2.6: Of the eight types of ethical challenges for data practitioners that we listed in Part Two, which types are *most* relevant to the word embedding study? Briefly explain your answer.

Question 2.7: What ethical concerns should data practitioners have when relying on word embedding tools in natural language processing tasks and other big data applications? To say it in another way, what ethical questions should such practitioners ask themselves when using such tools?

Question 2.8: Some researchers have designed 'debiasing techniques' to address the solution to the problem of biased word embeddings. (Bolukbasi 2016) Such techniques quantify the harmful biases, and then use algorithms to reduce or cancel out the harmful biases that would otherwise appear and be amplified by the word embeddings. Can you think of any significant tradeoffs or risks of this solution? Can you suggest any other possible solutions or ways to reduce the ethical harms of such biases?

Question 2.9: Identify four different uses/applications of data in which racial or gender biases in word embeddings might cause significant ethical harms, then briefly describe the specific harms that might be caused in each of the four applications, and who they might affect.

Question 2.10: Bias appears not only in language datasets but in image data. In 2016, a site called beauty.ai, supported by Microsoft, Nvidia and other sponsors, launched an online 'beauty contest' which solicited approximately 6000 selfies from 100 countries around the world. Of the entrants, 75% were of white and of European descent. Contestants were judged on factors such as facial symmetry, lack of blemishes and wrinkles, and how young the subjects looked for their age group. But of the 44 winners picked by a 'robot jury' (i.e., by beauty-detecting algorithms trained by data scientists), only 2% (1 winner) had dark skin, leading to media stories about the 'racist' algorithms driving the contest.[16] *How might the bias have got into the algorithms built to judge the contest, if we assume that the data scientists did not <u>intend</u> a racist outcome?*

---

[16] Levin (2016), https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people

PART THREE

What are data practitioners' obligations to the public?

To what extent are data practitioners across the spectrum—from data scientists, system designers, data security professionals, database engineers, users of third-party data analytics and other big data techniques—obligated by **ethical duties to the public**? **Where do those obligations come from**? And **who is 'the public'** that deserves a data practitioner's ethical concern?

1. WHY DO DATA PRACTITIONERS HAVE OBLIGATIONS TO THE PUBLIC?

One *simple* answer is, 'because data practitioners are human beings, and all human beings have ethical obligations to one another.' The vast majority of people, upon noticing a small toddler crawling toward the opening to a deep mineshaft, will feel obligated to redirect the toddler's path or otherwise stop to intervene, even if the toddler is unknown and no one else is around. If you are like most people, you just accept that you have some basic ethical obligations toward other human beings.

But of course, our **ethical obligations to an overarching 'public' always co-exist with ethical obligations to one's family, friends, employer, local community, and even oneself**. In this part of the module we highlight the public obligations because **too often, important obligations to the public are ignored in favor of more familiar ethical obligations** we have to specific known others in our social circle—even in cases when the ethical obligation we have to the public is objectively much **stronger** than the more local one.

If you're tempted to say 'well of course, I *always* owe my family/friends/employer/myself more than I owe to a bunch of strangers,' consider that this is **not how we judge things when we stand as an objective observer.** If the owner of a school construction company knowingly buys subpar/defective building materials to save on costs and boost his kids' college fund, resulting in a school cafeteria collapse that kills fifty children and teachers, we don't cut him any slack because he did it to benefit *his* family. We don't excuse his employees either, if they were knowingly involved and could anticipate the risk to others, *even* if they were told they'd be fired if they didn't cooperate. We'd tell them that keeping a job isn't worth sacrificing fifty strangers' lives. If we're thinking straight, we'd tell them that keeping a job doesn't give them permission to sacrifice even *one* strangers' life.

As we noted in Part One, **some data contexts *do* involve life and death risks** to the public. If my recklessly negligent cost-cutting on data hygiene, validation or testing results in a medical diagnostics error that causes one, or fifty, or a hundred strangers' deaths, it's really no different, morally speaking, than the reckless negligence of the school construction. Notice, however, that it may take us longer at first to make the connection, since the cause-and-effect relationship in the data case can be harder to visualize.

Other **risks of harm to the public that we must guard against** include those we described in Part One, from reputational harm, economic damage, and psychological injury, to reinforcement of unfair or unjust social arrangements.

However, it remains true that **the nature and details of our obligations to the public as data practitioners can be unclear.** How *far* do such obligations go, and *when do they take precedence* over other obligations? To what extent and in what cases do I *share* those obligations with others on my team or in my company? **These are not easy questions**, and often, the answers depend considerably on the details of the specific situation confronting us. But there are some ways of thinking about our obligations to the public that can help dispel some of the fog; Part Three outlines several of these.

2. DATA PROFESSIONALS AND THE PUBLIC GOOD

Remember that if the good life requires making a positive contribution to the world in which others live, then it would be perverse if we accomplished none of that in our professional lives, where we spend many or most of our waking hours, and to which we devote a large proportion of our intellectual and creative energies. Excellent doctors contribute health and vitality to the public. Excellent professors contribute knowledge, skill and creative insights to the public domain of education. Excellent lawyers contribute balance, fairness and intellectual vigor to the public system of justice. **Data professionals of various sorts contribute goods to the public sphere as well.**

What is a data *professional*? You may not have considered that the word 'professional' is etymologically connected with the English verb 'to profess.' What is it to profess something? It is to stand publicly for something, to express a belief, conviction, value or promise to a general audience that you expect that audience to hold you accountable for, and to identify you with. When I *profess* something, I say to others that this is something about which I am serious and sincere; and which I want them to know about me. So when we identify someone as a professional *X* (**whether 'X' is a lawyer, physician, soldier, data scientist, data analyst, or data engineer**), we are saying that being an 'X' is not just a job, but a *vocation*—a form of work to which the individual is committed and with which they would like to be identified. If I describe myself as just having a 'job,' I don't identify myself with it. But if I talk about '*my work*' or '*my profession*,' I am saying something more. This is part of why most professionals are expected to undertake **continuing education and training in their field**; not only because they need the expertise (though that too), but also because this is an important sign of their investment in and commitment to the field. Even if I leave a profession or retire, I am likely to continue to identify with it—an ex-lawyer will refer to herself as a 'former lawyer,' an ex-soldier calls himself a 'veteran.'

So how does being a professional create special ethical obligations for the data practitioner? Consider that members of most professions enjoy an elevated status in their communities; doctors, professors, scientists and lawyers generally get more respect from the public (rightly or wrongly) than retail clerks, toll booth operators, and car salespeople. But why? It can't just be the difference in skill; after all, car salespeople have to have very specialized skills in order to thrive in their job. The distinction lies in the perception that **professionals secure a vital public good**, not something of merely private and conditional value. For example, without doctors, public health would certainly suffer – and a good life is virtually impossible without some measure of health. Without lawyers and judges, the public would have no formal access to justice – and without recourse for injustice done to you or others, how can the good life be secure?

So each of these professions is supported and respected by the public precisely because they deliver something vital to the good life, and something needed not just by a few, but by us all.

Although data practices are employed by a range of professionals in many fields, from medical research to law and social science, many data practices are turning into new professions *of their own*, and these will continue to gain more and more public recognition and respect. **What do data scientists, data analysts, data engineers, and other data professionals do to earn that respect?** How must they act in order to continue to earn it? After all, special public respect and support are not given for free or given unconditionally—they are given in recognition of some service or value. That support and respect is also something that translates into real power; the power of public funding and consumer loyalty, the power of influence over how people live and what systems they use to organize their lives; in short, the power to guide the course of other human beings' technological future. And as we are told in the popular Spiderman saga, "With great power comes great responsibility." This is a further reason, even above their general ethical obligations as human beings, that data professionals have special ethical obligations to the public they serve.

Question 3.1: What sort of goods can data professionals contribute to the public sphere? (Answer as fully/in as many ways as you are able):

Question 3.2: What kinds of character traits, qualities, behaviors and/or habits do you think mark the kinds of data professionals who will contribute most to the public good? (Answer as fully/in as many ways as you are able):

3. JUST WHO *IS* THE 'PUBLIC'?

Of course, one can respond simply with, 'the public is everyone.' But the public is not an undifferentiated mass; the public is composed of our families, our friends and co-workers, our employers, our neighbors, our church or other local community members, our countrymen and women, and people living in every other part of the world. To say that we have ethical obligations to 'everyone' is to tell us very little about how to actually work responsibly as in the public interest, since **each of these groups and individuals that make up the public are in a unique relationship to us and our work, and are potentially impacted by it in very different ways**. And as we have noted, we also have special obligations to some members of the public (our children, our employer, our friends, our fellow citizens) that exist alongside the broader, more general obligations we have to all.

One concept that ethicists use to clarify our public obligations is that of a ***stakeholder***. A stakeholder is anyone who is potentially impacted by my actions. Clearly, certain persons have ***more*** at stake than other stakeholders in any given action I might take. When I consider, for example, how much effort to put into cleaning up a dirty dataset that will be used to train a 'smart' pacemaker, it is obvious that the patients in whom the pacemakers with this programming will be implanted are the primary stakeholders in my action; their very lives are potentially at risk in my choice. And this stake is so ethically significant that it is hard to see how any other stakeholder's interest could weigh as heavily.

4. DISTINGUISHING AND RANKING COMPETING STAKEHOLDER INTERESTS

Still, in most data contexts there are a **variety of stakeholders** potentially impacted by my action, and **their interests do not always align** with each other. For example, my employer's interests in cost-cutting and an on-time product delivery schedule may be in tension with the interest of other stakeholders in having the highest quality and most reliable data product on the market. Yet even such **stakeholder conflicts** are rarely so stark as they might first appear. In our example, the consumer also has an interest in an *affordable* and *timely* data product, and my employer also has an interest in earning a reputation for product excellence in its sector, and maintaining the profile of a responsible corporate citizen. Thinking about the public in terms of *stakeholders*, and **distinguishing them by the different 'stakes' they hold in what we do as data practitioners**, can help to sort out the tangled web of our varied ethical obligations to one amorphous 'public.'

Of course, **I *too* am a stakeholder**, since my actions impact my own life and well-being. Still, my *trivial* or *non-vital* interests (say, in shirking a necessary but tedious data obfuscation task, or concealing rather than reporting and patching an embarrassing security hole in my app) *will never trump a critical moral interest* of another stakeholder (say, their interest in not being unjustly arrested, injured, or economically damaged due to my professional laziness). **Ignoring the health, safety, or other vital interests of those who *rely* upon my data practice is simply not justified by my own stakeholder standing**. Typically, doing so would imperil my reputation and long-term interests anyway.

Ethical decision-making thus requires cultivating the habit of reflecting carefully upon the range of stakeholders who together make up the 'public' to whom I am obligated, and **weighing what is at stake for each of us** in my choice, or the choice facing my team or group. **On the next few pages is a case study you can use to help you think about what this reflection process can entail.**

CASE STUDY 4

In 2016, two Danish social science researchers used data scraping software developed by a third collaborator to amass and analyze a trove of public user data from approximately 68,000 user profiles on the online dating website OkCupid. The purported aim of the study was to analyze "the relationship of cognitive ability to religious beliefs and political interest/participation" among the users of the site.

However, when the researchers published their study in the open access online journal *Open Differential* Psychology, they included *their entire dataset*, without use of any deanonymizing or other privacy-preserving techniques to obscure the sensitive data. Even though the real names and photographs of the site's users were not included in the dataset, the publication of usernames, bios, age, gender, sexual orientation, religion, personality traits, interests, and answers to popular dating survey questions was immediately recognized by other researchers as an acute privacy threat, since this sort of data is easily re-identifiable when combined with other publically available datasets.

That is, the real-world identities of many of the users, even when not reflected in their chosen usernames, could easily be uncovered and relinked to the highly sensitive data in their profiles, using commonly available re-identification techniques. The responses to the survey questions were especially sensitive, since they often included information about users' sexual habits and desires, history of relationship fidelity and drug use, political views, and other extremely personal information. Notably, this information was public only to others logged onto the site as a user who had answered the same survey questions; that is, users expected that the only people who could see their answers would be other users of OkCupid seeking a relationship. The researchers, of course, had logged on to the site and answered the survey questions for an entirely different purpose—to gain access to the answers that thousands of others had given.

When immediately challenged upon release of the data and asked via social media if they had made any efforts to anonymize the dataset prior to publication, the lead study author Emil Kirkegaard responded on Twitter as follows: "No. Data is already public." In follow-up media interviews later, he said: "We thought this was an obvious case of public data scraping so that it would not be a legal problem."[17] When asked if the site had given permission, Kirkegaard replied by tweeting "Don't know, don't ask. :)"[18] A spokesperson for OkCupid, which the researchers had not asked for permission to scrape the site using automated software, later stated that the researchers had violated their Terms of Service and had been sent a take-down notice instructing them to remove the public dataset. The researchers eventually complied, but not before the dataset had already been accessible for two days.

Critics of the researchers argued that even *if* the information *had* been legally obtained, it was also a flagrant *ethical* violation of many professional norms of research ethics (including informed consent from data subjects, who never gave permission for their profiles to be used or published by the researchers). Aarhus University, where the lead researcher was a student, distanced itself from the study saying that it was an independent activity of the student and not funded by Aarhus, and that "We are sure that [Kirkegaard] has not learned his methods and ethical standards of research at our university, and he is clearly not representative of the about 38,000 students at AU."

The authors did appear to anticipate that their actions might be ethically controversial. In the draft paper, which was later removed from publication, the authors wrote that "Some may object to the ethics of gathering and releasing this data…However, all the data found in the dataset are or were already publicly available, so releasing this dataset merely presents it in a more useful form."[19]

---

[17] Hackett (2016): http://fortune.com/2016/05/18/okcupid-data-research/

[18] Resnick (2016): https://www.vox.com/2016/5/12/11666116/70000-okcupid-users-data-release

[19] Hackett (2016) http://fortune.com/2016/05/18/okcupid-data-research/

Question 3.3: What specific, significant harms to members of the public did the researchers' actions risk? List as many types of harm as you can think of.

Question 3.4: How should those potential harms have been evaluated alongside the prospective *benefits* of the research claimed by the study's authors? Could the benefits hoped for by the authors have been significant enough to justify the risks of harm you identified above in 3.3?

Question 3.5: List the various *stakeholders* involved in the OkCupid case, and for each type of stakeholder you listed, identify what was *at stake* for them in this episode. Be sure your list is as complete as you can make it, including all possible affected stakeholders.

Question 3.6: The researchers' actions potentially affected tens of thousands of people. Would the members of the public whose data were exposed by the researchers be justified in feeling *abused*, *violated*, or otherwise unethically treated by the study's authors, even though they have never had a personal interaction with the authors? If those feelings are justified, does this show that the study's authors had an ethical *obligation* to those members of the public that they failed to respect?

Question 3.7: The lead author repeatedly defended the study on the grounds that the data was technically public (since it was made accessible by the data subjects to other OkCupid users). The author's implication here is that no individual OkCupid user could have reasonably objected to their data being viewed by any other individual OkCupid user, so, the authors might argue, how could they reasonably object to what the authors did with it? How would you evaluate that argument? Does it make an ethical difference that the authors accessed the data in a very different way, to a far greater extent, with highly specialized tools, and for a very different purpose than an 'ordinary' OkCupid user?

Question 3.8: The authors clearly did anticipate some criticism of their conduct as unethical, and indeed they received an overwhelming amount of public criticism, quickly and widely. How meaningful is that public criticism? To what extent are big data practitioners *answerable* to the public for their conduct, or can data practitioners justifiably ignore the public's critical response to what they do? Explain your answer.

Question 3.9: As a follow up to Question 3.7, how meaningful is it that much of the criticism of the researchers' conduct came from a range of well-established data professionals and researchers, including members of professional societies for social science research, the profession to which the study's authors presumably aspired? How should a data practitioner want to be judged by his or her peers or prospective professional colleagues? Should the evaluation of our conduct by our professional peers and colleagues hold special sway over us, and if so, why?

Question 3.10: A Danish programmer, Oliver Nordbjerg, specifically designed the data scraping software for the study, though he was not a co-author of the study himself. What ethical obligations did he have in the case? Should he have agreed to design a tool for this study? To what extent, if any, does he share in the ethical responsibility for any harms to the public that resulted?

Question 3.11 How do you think the OkCupid study likely impacted the reputations and professional prospects of the researchers, and of the designer of the scraping software?

PART FOUR

What general ethical frameworks might guide data practice?

We noted above that data practitioners, in addition to their special professional obligations to the public, also have the same ethical obligations to their fellow human beings that we all share. What might those obligations be, and how should they be evaluated alongside our professional obligations? There are a number of familiar concepts that we already use to talk about how, in general, we ought to treat others. Among them are the concepts of rights, justice and the common good. But how do we define the concrete meaning of these important ideals? Here are three common frameworks for understanding our general ethical duties to others:

1. VIRTUE ETHICS

Virtue approaches to ethics are found in the ancient Greek and Roman traditions, in Confucian, Buddhist and Christian moral philosophies, and in modern secular thinkers like Hume and Nietzsche. Virtue ethics focuses not on rules for good or bad actions, but on the qualities of morally excellent persons (e.g., virtues). Such theories are said to be character based, insofar as they tell us what a person of virtuous character is like, and how that moral character develops. Such theories also focus on the habits of action of virtuous persons, such as the habit of moderation (finding the 'golden mean' between extremes), as well as the virtue of prudence or

practical wisdom (the ability to see what is morally required even in new or unusual situations to which conventional moral rules do not apply).

How can virtue ethics help us to understand what our moral obligations are? It can do so in three ways. The first is by helping to see that we have a basic moral obligation to make a consistent and conscious effort to develop our moral character for the better; as the philosopher Confucius said, the real ethical failing is not having faults, 'but rather failing to amend them.' The second thing virtue theories can tell us is where to look for standards of conduct to follow; virtue theories tell us to look for them in our own societies, in those special persons who are exemplary human beings with qualities of character (virtues) to which we should aspire. The third thing that virtue ethics does is direct us toward the lifelong cultivation of practical wisdom or good moral judgment: the ability to discern which of our obligations are most important in a given situation and which actions are most likely to succeed in helping us to meet those obligations. Virtuous persons with this ability flourish in their own lives by acting justly with others, and contribute to the common good by providing a moral example for others to admire and follow.

Question 4.1: How would a conscious habit of thinking about how to be a better human being contribute to a person's character, especially over time?

Question 4:2: Do you know what specific aspects of your character you would need to work on/improve in order to become a better person? (Yes or No)

Question 4:3: Do you think most people make enough of a regular effort to work on their character or amend their shortcomings? Do you think we are morally obligated to make the effort to become better people? Why or why not?

Question 4:4: Who do you consider a model of moral excellence that you see as an example of how to live, and whose qualities of character you would like to cultivate? Who would you want your children (or future children) to see as examples of such human (and especially moral) excellence?

Question 4:5: What are three strengths of moral character (*virtues*) that you think are particularly important for data practitioners to practice and cultivate in order to be excellent models of data practice in their profession? Explain your answers.

2. CONSEQUENTIALIST/UTILITARIAN ETHICS

Consequentialist theories of ethics derive principles to guide moral action from the likely consequences of those actions. The most famous form of consequentialism is utilitarian ethics, which uses the principle of the 'greatest good' to determine what our moral obligations are in any given situation. The 'good' in utilitarian ethics is measured in terms of happiness or pleasure (where this means not just physical pleasure but also emotional and intellectual pleasures). The absence of pain (whether physical, emotional, etc.) is also considered good, unless the pain somehow leads to a net benefit in pleasure, or prevents greater pains (so the pain of exercise would be good because it also promotes great pleasure as well as health, which in turn prevents more suffering). When I ask what action would promote the 'greater good,' then, I am asking which action would produce, in the long run, the greatest net sum of good (pleasure and absence of pain), taking into account the consequences for all those affected by my action (not just myself). This is known as the hedonic calculus, where I try to maximize the overall happiness produced in the world by my action.

Utilitarian thinkers believe that at any given time, whichever action among those available to me is most likely to boost the overall sum of happiness in the world is the right action to take, and my moral obligation. This is yet another way of thinking about the 'common good.' But utilitarians are sometimes charged with ignoring the requirements of individual rights and justice; after all, wouldn't a good utilitarian willingly commit a great injustice against one innocent person as long as it brought a greater overall benefit to others? Many utilitarians, however, believe that a society in which individual rights and justice are given the highest importance just is the kind of society most likely to maximize overall happiness in the long run.

After all, how many societies that deny individual rights, and freely sacrifice individuals/minorities for the good of the many, would we call happy?

Question 4:6: What would be the hardest part of living by the utilitarian principle of the 'greatest good'? What would be the most rewarding part?

Question 4:7: What different kinds of pleasure/happiness are there? Are some pleasures more or less valuable or of higher or lower quality than others? Why or why not? Explain your intuitions about this:

Question 4:8: Utilitarians think that pleasure and the absence of pain are the highest goods that we can seek in life, and that we should always be seeking to produce these goods for others (and for ourselves). They claim that every other good thing in life is valued simply because it produces pleasure or reduces pain. Do you agree? Why or why not?

Question 4:9: A utilitarian might say that to measure a 'good life,' you should ask: 'how much overall happiness did this life bring into the world?' Do you agree that this is the correct measure of a good life, or not? Briefly explain.

Question 4:10: In what ways do you think data practitioners can promote the 'greater good' through their work, that is, increase human happiness?

3. DEONTOLOGICAL ETHICS

Deontological ethics are rule or principle-based systems of ethics, in which one or more rules/principles are claimed to tell us what our moral obligations are in life. In Judeo-Christian thought, the Ten Commandments can be thought of as a deontological system. Among modern, secular forms of ethics, many deontological systems focus on lists of 'rights' (for example, the rights not to be unjustly killed, enslaved, or deprived of your property). Consider also the modern idea of 'universal human rights' that all countries must agree to respect. In the West, moral rights are often taken as a basis for law, and are often invoked to justify the making of new laws, or the revision or abolition of existing ones. In many cultures of East Asia, deontological systems may focus not on rights but on duties; these are fixed obligations to others (parents, siblings, rulers, fellow citizens etc.) that must be fulfilled according to established rules of conduct that govern various types of human relationships.

Another well-known deontological system is that of the 18th century philosopher Immanuel Kant, who identified a single moral rule called the categorical imperative. This principle tells us to only act in ways that we would be willing to have all other persons follow, all of the time. He related this to another principle that tells us never to treat a human being as a 'mere means to an end,' that is, as an object to be manipulated for our own purposes. For example, I might want to tell a lie to get myself out of trouble in a particular case. But I certainly would not want everyone in the world to lie every time they felt like it would help them avoid trouble. And if someone lies to me to get me to do something that benefits them, I am rightly upset about being treated as a mere object to be manipulated for gain. So, I cannot logically give myself permission to lie, since there is nothing about me that exempts me from my own general moral standards for human

behavior. For if I am willing to give myself permission to act in this way for this reason, how could I logically justify withholding the same permission from others?

According to this principle, human lives are the ultimate sources of all moral value. I thus have a universal moral obligation to treat other human lives in ways that acknowledge and respect their unconditional value, and to not treat them merely as tools to manipulate for lesser purposes. And since I myself am human, I cannot morally allow even my own existence to be used as a mere tool for some lesser purpose (for example, to knowingly sell out my personal integrity for money, fame or approval). This principle highlights my duty to always respect the dignity of all human lives. This theory is also linked with a particular idea of *justice*, as treatment that recognizes the basic equality and irreplaceable dignity of every human being, no matter who they are or where they live. Such thinking is often considered to be at the heart of the modern doctrine of inalienable human rights.

Question 4:11: How often, when making decisions, do you think about whether you would willingly allow or support others acting in the same way that you are choosing to act? Does it seem like something you *should* think about?

Question 4:12: What are two cases you can think of in data practice in which a person or persons were treated as a 'mere means to an end', that is, treated as nothing more than a useful tool to achieve someone else's goal? (Feel free to draw from any of the working examples in previous parts of the module).

Question 4:13: Do you agree that human lives are of the highest possible value and beyond any fixed 'price'? In your opinion, how well does our society today reflect this view on morality and justice? *Should* it reflect this view?

Question 4:14: While each of the 3 distinct types of ethical frameworks/theories reviewed in this section is subject to certain limitations or criticisms, what aspects of the good life/ethics do you think each one captures best?

PART FIVE

What are ethical best practices for data practitioners?

The phrase 'best practices' refers to known techniques for doing something that tend to work well, better than the alternative ways of doing something. It's not a phrase unique to ethics, in fact it's used in a range of corporate and government settings; but it's often used in contexts where it is very important that the thing be done well, and where there are significant costs or risks to doing it in a less than optimal way.

For data practitioners, we describe two types of best practices. The first set focuses on **best practices for functioning ethically *in data practice***; they are adapted specifically to the ethical challenges that we studied in Part Two of this module. The second set identifies **best practices for living and acting ethically *in general***; these practices can be adopted by anyone, regardless of their career or professional interests. Data practitioners can benefit from drawing upon *both* sets of practices in creative ways to manage ethical challenges wisely and well.

1. BEST PRACTICES FOR DATA ETHICS

As noted in the Introduction, no single, detailed code of data ethics can be fitted to all data contexts and practitioners; organizations and data-related professions should therefore be encouraged to develop explicit internal policies, procedures, guidelines and best practices for data ethics that are specifically adapted to their own activities (e.g., data science, machine learning, data security and storage, data privacy protection, medical and scientific research, etc.) However, those specific codes of practice can be well shaped by reflecting on these 14 general norms and guidelines for ethical data practice:

I. Keep Data Ethics in the Spotlight—and Out of the Compliance Box: As earlier modules and examples have shown, data ethics is a *pervasive* aspect of data practice. Because of the immense social power of data, ethical issues are virtually *always* actively in play when we handle data. Even when our work is highly technical and not directly client-facing, ethical issues are never simply *absent* from the context of our work. However, the 'compliance mindset' found many organizations, especially concerning *legal* matters, can, when applied to data ethics, encourage a dangerous tendency to 'sideline' ethics as an external constraint rather than see it as an integral part of our daily work. If we fall victim to that mindset, we are more likely to view our ethical obligations as a box to 'check off' and then happily forget about, once we feel we have done the minimum needed to 'comply' with our ethical obligations. Unfortunately, this often leads to disastrous consequences, for individuals and organizations alike. Because data practice involves ethical considerations that are *ubiquitous* and *central*, not intermittent and marginal, our individual and organizational efforts need to strive to keep ethics in the spotlight.

II. Consider the Human Lives and Interests Behind the Data: Especially in technical contexts, it's easy to lose sight of what most of the data we work with *are*: namely, reflections of human lives and interests. Even when the data we handle are generated by non-human entities (for example, recordings of ocean temperatures), these data are being collected for important *human* purposes and interests. And much of the data under the 'big data' umbrella concern the most *sensitive* aspects of human lives: the condition of people's bodies, their finances, their social likes

and dislikes, or their emotional and mental states. A decent human would never handle another person's body, money, or mental condition without due care; but it can be easy to forget that this is often what we are doing when we handle data.

III. Focus on Downstream Risks and Uses of Data: As noted above, often we focus too narrowly on whether *we* have complied with ethical guidelines and we forget that ethical issues concerning data don't just 'go away' once we have performed a particular task diligently. Thus it is essential to think about what happens to or with the data later on, even after it leaves our hands. Even if, for example, we obtained explicit and informed consent to collect certain data from a subject, we cannot ignore how that data might impact the subject, or others, down the road. If the data poses clear risks of harm if inappropriately used or disclosed, then I should be asking myself where that data might be five or ten years from now, in whose hands, for what purposes, and with what safeguards. I should also consider how long that data will remain accurate and relevant, or how its sensitivity and vulnerability to abuse might increase in time. If I can't answer any of those questions—or have not even asked them—then I have not fully appreciated the ethical stakes of my *current* data practice.

IV. Don't Miss the Forest for the Trees: Envision the Data Ecosystem: This is related to the former item; but broader in scope. Not only is it important to keep in view where the data I handle today is going tomorrow, and for what purpose, I also need to keep in mind the full context in which it exists *now*. For example, if I am a university genetics researcher handling a large dataset of medical records, I might be inclined to focus narrowly on how *I* will collect and use the genetic data responsibly. But I also have to think about who else might have an interest in obtaining such data, and for different purposes than mine (for example, employers and insurance companies). I may have to think about the cultural and media context in which I'm collecting the data, which might embody expectations, values, and priorities concerning the collection and use of personal genetic data that conflict with those of my academic research community. I may need to think about where the server or cloud storage company I'm currently using to store the data is located, and what laws and standards for data security exist there. The point here is that my data practices are never isolated from a broader data ecosystem that includes powerful social forces and instabilities not under my control; it is essential that I consider my ethical practices and obligations in light of that bigger social picture.

V. Mind the Gap Between Expectations and Reality: When collecting or handling personal or otherwise sensitive data, it's essential that I keep in mind how the expectations of data subjects or other stakeholders may vary from reality. For example, do my data subjects know as much about the risks of data disclosure (from hacking, phishing, etc.) as I do? Might my data disclosure and use policy lead to inflated expectations about how safe users' data are from such threats? Do I intend to use this data for additional purposes beyond what the consenting subjects would know about or reasonably anticipate? Can I keep all the promises I have made to my data subjects, or do I know that there is a good chance that their expectations will not be met? For example, might I one day sell my product and/or its associated data to a third-party who may not honor those promises? Often we make the mistake of regarding parties we contract with as information equals, when we may in fact operate from a position of epistemic advantage—we know a lot more than they do. Agreements with data subjects who are 'in the dark' or subject to illusions about the nature of the data agreement are not, in general, ethically legitimate.

VI. Treat Data as a Conditional Good: Some of the most dangerous data practices involve treating data as unconditionally good. One such practice is to follow the policy of 'collect and store it all now, and figure out what we actually need later.' Data (at least *good* data) is incredibly *useful*, but its power also makes it capable of doing damage. Think about personal data like guns: only some of us should be licensed to handle guns, and even those of us who are licensed should keep only as many guns as we can reasonably think we actually need, since they are so often stolen or misused in harmful ways. The same is often true for sensitive data. We should collect only as much of it as we need, when we need it, store it carefully for only as long as we need it, and purge it when we no longer need it. The second dangerous practice that treats data as an unconditional good is the flawed policy that more data is always better, regardless of data quality or the reliability of the source. The motto 'garbage in, garbage out' is of critical importance to remember, and just because our algorithms and systems are incredibly thirsty for data, doesn't mean that we should open the firehose and send them all the data we can get our hands on— especially if that data is dirty, incomplete, or unreliably sourced. Data are a *conditional* good— only as beneficial and useful as we take the care to make them.

VII. Avoid Dangerous Hype and Myths around 'Big Data': Data is powerful, but it isn't *magic*, and it isn't a silver bullet for complex social problems. There are, however, significant industry and media incentives to portray 'big data' as exactly that. This can lead to many harms, including unrealized hopes and expectations that can easily lead to consumer, client, and media backlash. The saying 'to a man with a hammer, everything looks like a nail' is also instructive here. Not all problems have a big data solution, and we may overlook more economical and practical solutions if we believe otherwise. We should also remember the joke about the drunk man who, when asked why he's looking for his lost car keys under the street lamp, says 'because that's where the light is.' For some problems we have abundant sources of high-quality, relevant data and powerful analytics that can use them to produce new insights and solutions. For others, we don't. But we shouldn't ignore problems that might require other kinds of solutions, or employ inappropriate solutions, just because we are in the thrall of 'big data' hype.

VIII. Establish Chains of Ethical Responsibility and Accountability: In organizational settings, the 'problem of many hands' is a constant challenge to responsible practice and accountability. To avoid a diffusion of responsibility in which no one on a team may feel empowered or obligated to take the steps necessary to ensure ethical data practice, it is important that clear *chains of responsibility* are established and made explicit to everyone involved in the work, at the earliest possible stages of a project. It should be clear who is responsible for each aspect of ethical risk management and prevention of harm, in each of the relevant areas of risk-laden activity (data collection, use, security, analysis, disclosure, etc.) It should also be clear who is ultimately *accountable* for ensuring an ethically executed project or practice. Who will be expected to provide answers, explanations, and remedies if there is a failure of ethics or significant harm caused by the team's work? The essential function of chains of responsibility and accountability is to assure that members of a data-driven project or organization *take explicit ownership of the work's ethical significance.*

IX. Practice Data Disaster Planning and Crisis Response: Most people don't want to anticipate failure, disaster, or crisis; they want to focus on the positive potential of a project. While this is understandable, the dangers of this attitude are well known, and have often *caused* failure, disaster, or crisis that could easily have been avoided. This attitude also often prevents effective crisis response since there is no planning for a worst-case-scenario. This is why

engineering fields whose designs can impact public safety have long had a culture of *encouraging* thinking about failure. Understanding how a product will function in non-ideal conditions, at the boundaries of intended use, or even outside those boundaries, is essential to building in appropriate margins of safety and developing a plan for product failures or other unwelcome scenarios. Thinking about failure makes engineers' work *better*, not worse. Data practitioners must begin to develop the same cultural habit in their work. Known failures should be carefully analyzed and discussed ('post-mortems') and results projected into the future. 'Pre-mortems' (imagining together how a current project could fail or produce a crisis, so that we can design to prevent that outcome) can be a great data practice. It's also essential to develop crisis plans that go beyond deflecting blame or denying harm (often the first mistake of a PR team when the harm is evident). Crisis plans should be intelligent, responsive to public input, and most of all, able to effectively mitigate or remedy harm being done. This is much easier to plan *before* a crisis has actually happened.

X. Promote Values of Transparency, Autonomy, and Trustworthiness: The most important thing to preserve a healthy relationship between data practitioners and the public is for data practitioners to understand the importance of transparency, autonomy, and trustworthiness to that relationship. Hiding a risk or a problem behind legal language, disempowering users or data subjects, and betraying public trust are almost never good strategies in the long run. Clear and understandable data collection, use, and privacy policies, when those policies give users and data subjects actionable information and encourage them to use it, help to promote these values. Favoring 'opt-in' rather than 'opt-out' options and offering other clear avenues of choice for data participants can enhance autonomy and transparency, and promote greater trust. Of course, we can't always be completely transparent about everything we do with data: company interests, intellectual property rights, and privacy concerns of other parties often require that we balance transparency with other legitimate goods and interests. Likewise, sometimes the autonomy of users will be in tension with our obligations to prevent harmful misuse of data. But *balancing* transparency and autonomy with other important rights and ethical values is not the same as *sacrificing* these values or ignoring their critical role in sustaining public trust in data-driven practices and organizations.

XI. Consider Disparate Interests, Resources, and Impacts: It is important to understand the profound risk in many data practices of producing or magnifying disparate impacts; that is, of *making some people better off and others worse off*, whether this is in terms of their social share of economic well-being, political power, health, justice, or other important goods. *Not all disparate impacts are unjustifiable or wrong.* For example, an app that flags businesses with a high number of consumer complaints and lawsuits will make those businesses worse off relative to others in the same area—but if the app and its data are sufficiently reliable, then there's an argument that this disparate impact is a *good* thing. But imagine *another* app, created for the same purpose, that sources its data from consumer complaints in a way that reflects and magnifies existing biases in a given region against women business owners, business owners of color, and business owners from certain religious backgrounds. The fact that more complaints per capita are registered against *those* businesses might be an artifact of those harmful biases in the region, which my app then just blindly replicates and reinforces. This is why there ought to be a *presumption in data practice of ethical risk from disparate impacts; they must be anticipated, actively audited for, and carefully examined for their ethical acceptability.* Likewise, we must investigate the extent to which different populations affected by our practice have different interests and resources, that give them a differential ability to benefit from our product or project. If a data-driven product produces

immense health benefits but is inaccessible to people who are blind, deaf, or non-native English speakers, or to people who cannot afford the latest high-end mobile devices, then there are disparate impacts of this work that at a minimum must be reflected upon and evaluated.

XII. Invite Diverse Stakeholder Input: One way to avoid 'groupthink' in ethical risk assessment and design is to invite input from diverse stakeholders outside of the team and organization. It is important that stakeholder input not simply reflect the same perspectives one already has within the organization. Often, data practitioners work in fields with unusually high levels of educational achievement and economic status, and in many technical fields, there may be skewed representation of the population in terms of gender, ethnicity, age, disability, and other characteristics. Also, the nature of the work may attract people who have common interests and values, for example, a shared optimism about the potential of science and technology to promote social good, and comparatively less faith in other social mechanisms. All of these factors can lead to organizational *monocultures*, which magnify the dangers of groupthink, blind spots, and insularity of interests. For example, many of the best practices above can't be carried out successfully if members of a team struggle to imagine how a data practice would be perceived by, or how it might affect, people unlike themselves. Actively recognizing the limitations of a team perspective is essential. Fostering more diverse data organizations and teams is one obvious way to mitigate those limitations, but soliciting external input from a more truly representative body of those likely to be impacted by our data practice is another.

XIII. Design for Privacy and Security: This might seem like an obvious one, but nevertheless its importance can't be overemphasized. 'Design' here means not only technical design (of databases, algorithms, or apps), but also social and organizational design (of groups, policies, procedures, incentives, resource allocations, and techniques) that promote data privacy and data security objectives. How this is best done in each context will vary, but the essential thing is that along with other project goals, the values of data privacy and security remain at the forefront of project design, planning, execution, and oversight, and are never treated as marginal, external, or 'after-the-fact' concerns.

XIV. Make Ethical Reflection & Practice Standard, Pervasive, Iterative, and Rewarding: Ethical reflection and practice, as we have already said, is an essential and central part of professional excellence in data-driven applications and fields. Yet it is still in the process of being fully integrated into every data environment. The work of making ethical reflection and practice *standard* and *pervasive*, that is, accepted as a necessary, constant, and central component of every data practice, must continue to be carried out through active measures taken by individual data practitioners and organizations alike. Ethical reflection and practice in data environments must also, to be effective, be instituted in *iterative* ways. That is, because data practice is so increasingly complex in its interactions with society, we must treat data ethics as an active and unending learning cycle in which we continually observe the outcomes of our data practice, learn from our mistakes, gather more information, acquire further ethical expertise, and then update and improve our ethical practice accordingly. Most of all, ethical practice in data environments must be made *rewarding*: team, project, and institutional/company incentives must be well aligned with the ethical best practices described above, so that those practices are reinforced and so that data practitioners are empowered and given the necessary resources to carry them out.

Question 5:1: Of these fourteen best practices for data ethics, which two do you think are the most challenging to carry out? What do you think could be done (by an individual, a team, or an organization) to make those practices easier?

Question 5:2: What benefits do you think might come from successfully instituting these practices in data environments—for society overall, *and* for big data professionals?

2. GENERAL BEST PRACTICES FOR LIVING WELL

There are a number of unfortunate habits and practices that create *obstacles* to living well in the moral sense; fortunately, there are also a number of common habits and practices that are highly *conducive* to living well. Here are five ethically beneficial habits of mind and action:

I. Practice Self- Reflection/Examination: This involves spending time on a regular basis (even daily) thinking about the person you want to become, in relation to the person you are today. It involves identifying character traits and habits that you would like to change or improve in your private and professional life; reflecting on whether you would be happy if those whom you admire and respect most knew all that you know about your actions, choices and character; and asking yourself how fully you are living up to the values you profess to yourself and others.

II. Look for Moral Exemplars: Many of us spend a great deal of our time, often more than we realize, judging the shortcomings of others. We wallow in irritation or anger at what we perceive as unfair, unkind or incompetent behavior of others, we comfort ourselves by noting the even greater professional or private failings of others, and we justify ignoring the need for our own ethical improvement by noting that many others seem to be in no hurry to become better people either. What we miss when we focus on the shared faults of humanity are those exemplary actions we witness, and the exemplary persons in our communities, that offer us a path forward in our own self-development. Exemplary acts of forgiveness, compassion, grace, courage, creativity and justice have the power to draw our aspirations upward; especially when we consider that there is no reason why we would be incapable of these actions ourselves. But this cannot happen unless we are in the habit of looking for, and taking notice of, moral exemplars in the world around us. We can also look specifically to moral exemplars in our chosen profession.

III. Exercise Moral Imagination: It can be hard to notice our ethical obligations, or their importance, because we have difficulty imagining how what we do might affect others. In some sense we all know that our personal and professional choices almost always have consequences for the lives of others, whether good or bad. But rarely do we try to really imagine what it will be like to suffer the pain that our action is likely going to cause someone – or what it will be like to experience the joy, or relief of pain or worry that another choice of ours might bring. This becomes even harder as we consider stakeholders who live outside of our personal circles and beyond our daily view. The pain of your best friend who you have betrayed is easy to see, and not difficult to imagine before you act – but it is easy not to see, and not to imagine, the pain of a person on another continent, unknown to you, whose life has been ruined by identity theft or political persecution because you recklessly allowed their sensitive data to be exposed. The suffering of that person, and your responsibility for it, would be no less great simply because you had difficulty imagining it. Fortunately, our powers of imagination can be increased. Seeking out news, books, films and other sources of stories about the human condition can help us to better envision the lives of others, even those in very different circumstances from our own. This capacity for imaginative empathy, when habitually exercised, enlarges our ability to envision the likely impact of our actions on other stakeholders. Over time, this can help us to fulfill our ethical obligations and to live as better people.

IV. Acknowledge Our Own Moral Strength: For the most part, living well in the ethical sense makes life easier, not harder. Acting like a person of courage, compassion and integrity is, in most

circumstances, also the sort of action that garners respect, trust and friendship in both private and professional circles, and these are actions that we ourselves can enjoy and look back upon with satisfaction rather than guilt, disappointment or shame. But it is inevitable that sometimes the thing that is right will not be the easy thing, at least not in the short term. And all too often our moral will to live well gives out at exactly this point – under pressure, we take the easy (and wrong) way out, and try as best we can to put our moral failure and the harm we may have done or allowed out of our minds.

One of the most common reasons why we fail to act as we know we should is that we think we are too weak to do so, that we lack the strength to make difficult choices and face the consequences of doing what is right. But this is often more of a self-justifying and self-fulfilling fantasy than a reality; just as a healthy person may tell herself that she simply can't run five miles, thus sparing her the effort of trying what millions of others just like her have accomplished, a person may tell herself that she simply can't tell the truth when it will greatly inconvenience or embarrass her, or that she simply can't help someone in need when it will cost her something she wants for herself. But of course people do these things every day; they tell the morally important truth and take the heat, they sell their boat so that their disabled friend's family does not become homeless, they report frauds from which they might otherwise have benefited financially. These people are not a different species from the rest of us; they just have not forgotten or discounted their own moral strength. And in turn, they live very nearly as they should, and as we at any time can, if we simply have the will.

V. Seek the Company of Other Moral Persons: Many have noted the importance of friendship in moral development; in the 4th century B.C. the Greek philosopher Aristotle argued that a virtuous friend can be a 'second self,' one who represents the very qualities of character that we value and aspire to preserve in ourselves. He notes also that living well in the ethical sense requires ethical actions, and that activity is generally easier and more pleasurable in the company of others. Thus seeking the company of other moral persons can keep us from feeling isolated and alone in our moral commitments; friends of moral character can increase our pleasure and self-esteem when we do well alongside them, they can call us out when we act inconsistently with our own professed ideals and values, they can help us reason through difficult moral choices, and they can take on the inevitable challenges of ethical life with us, allowing us to weather them together.

Aside from this, and as compared with persons who are ethically compromised, persons of moral character are direct sources of pleasure and comfort – we benefit daily from their kindness, honesty, mercy, wisdom and courage, just as they find comfort and happiness in ours. On top of all of this, Aristotle said, it is only in partnership with other good and noble people that we can produce good and noble things, since very little of consequence can be accomplished in life without the support and help of at least *some* others.

Question 5:3: Of these five moral habits and practices, which do you think you are best at presently? Which of these habits, if any, would you like to do more to cultivate?

Question 5.4: In what specific ways, small or large, do you think adopting some or all of these habits could make a person a better data practitioner?

CASE STUDY 5

In the summer of 2017, a published study by Stanford University researchers prompted alarm and criticism from LBGTQ groups and others who questioned the ethics of the study. The study sampled tens of thousands of dating website photos to create a deep learning algorithm for detecting sexual orientation, which the study's authors claim was able to perform this task with an accuracy between 74% and 81% -- notably better than human judges.[20] It was noted, however, that the algorithm in more realistic test conditions would likely yield a significant number of false positives; that is, ranking some straight persons as more likely to be gay or lesbian than others who actually are.[21]

Critics asserted that the study was methodologically flawed and biased. For example, it did not include any images of faces of people of color, a significant exclusion. There was also no consideration in the study design of transgender or bisexual persons.[22] Critics also asserted that the study was highly dangerous, insofar as such a tool could potentially be used by oppressive governments or other hostile parties to 'detect' and 'out' gay and lesbian persons and target them for social exclusion or punishment, even death. Such a tool might also be used by parents to try to predict homosexual behavior in children, or by spouses to 'test' their mate's sexuality, or by teenagers to 'test' the sexuality of their peers.

The study's authors defended their research by asserting that such technology is already available to create and abuse (although they did not make *their* algorithm public), and that their research helpfully brings this potential to light. They claimed that it had a legitimate scientific purpose, namely, to provide further evidence that sexual orientation has a biological basis, as opposed to being entirely a personal choice. The study's author also noted that similar techniques might be used with other datasets to detect IQ or political orientation. In an interview with *The Economist*, the study's author characterized the use of such data-driven algorithms to erode personal privacy as "inevitable."

Question 5.5: Identify the 5 most significant ethical issues/questions raised by this study.

---

[20] (Levin 2017a) https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph

[21] https://www.economist.com/news/science-and-technology/21728614-machines-read-faces-are-coming-advances-ai-are-used-spot-signs

[22] (Levin 2017b) https://www.theguardian.com/world/2017/sep/08/ai-gay-gaydar-algorithm-facial-recognition-criticism-stanford

Question 5.6: Identify 3 ethical best practices listed in Part Five that seem to you to be closely related to the issues you identified in Q5.5, and to their potential remedies.

CASE STUDY 6

In this concluding exercise, you (or, if your instructor chooses, a team) will design your own case study involving a hypothetical data project. (Alternatively, your instructor may provide you or your team with an existing case study for you to analyze.)

After coming up with your case outline, **you or your group must *identify*:**

1. The purpose/intended function of the data practice or practices involved in the hypothetical project. This will the outline of your case study, which might be built around a hypothetical big data-driven application, a data collection context, or a machine learning or analytics context.

2. The various types of stakeholders that might be involved in such a practice, and the different stakes/interests they have in the outcome.

3. The potential benefits and risks of harm that could be created by such a project, including 'downstream' impacts.

4. The ethical challenges most relevant to this project (be sure to draw your answers from the list of challenges outlined in Part Two of this module, although feel free to note any other ethical challenges not included in that section).

5. The ethical obligations to the public that such a project might entail for the data professionals working on it.

6. Any potential for disparate impacts of the project that should be anticipated, and how those might differently affect various stakeholders.

7. The ethical best-case scenario (the maximum social benefit the data practitioners would hope to come out of the project) *and* a worst-case scenario (how the project could lead to an ethical disaster or at least substantial harm to the significant interests of others).

8. One way that the risk of the worst-case-scenario could be reduced in advance, and one way that the harm could be mitigated after-the-fact by an effective crisis response.

9. At least *three* brief proposals or ideas for carrying out the project in the most ethical way possible. Or, if the project as outlined could *never* be carried out in an ethical way, identify a redesign or alternative project that would be more ethically sound. Use the module content, especially Parts Two and Five, to help you come up with your ideas.

APPENDIX A. RELEVANT PROFESSIONAL ETHICS CODES & GUIDELINES

As noted in the Introduction to this module, the sheer variety of professional and personal contexts in which data are involved is such that no single code of professional ethics or list of professional guidelines will be relevant for all data practitioners. However, below are some available resources that will be relevant to many readers:

"Building Digital Trust: The Role of Ethics in the Digital Age" from Accenture
https://www.accenture.com/t20160613T024441Z__w__/us-en/_acnmedia/PDF-22/Accenture-Data-Ethics-POV-WEB.pdf#zoom=50

"Universal Principles of Data Ethics: 12 Guidelines for Developing Ethics Codes" from Accenture
https://www.accenture.com/t20160629T012639Z__w__/us-en/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf#zoom=50

Ethics Guidelines from AOIR (Association of Internet Researchers)
https://aoir.org/ethics

Code of Ethics and Professional Conduct of ACM (Association for Computing Machinery)
https://www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct

Software Engineering Code of Ethics and Professional Practice of ACM (Association for Computing Machinery) and IEEE-Computer Society
http://www.acm.org/about/se-code

Code of Conduct of Data Science Association
http://www.datascienceassn.org/code-of-conduct.html

The Web Analyst's Code of Ethics of the Digital Analytics Association
https://www.digitalanalyticsassociation.org/codeofethics

Report on "Ethics Codes: History, Context, and Challenges" from The Council for Big Data, Ethics, and Society
http://bdes.datasociety.net/council-output/ethics-codes-history-context-and-challenges/

Code of Ethics and Professional Conduct of The Association of Clinical Research Practitioners
https://www.acrpnet.org/about/code-of-ethics/

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
(includes several IEEE P7000 Working Groups on Standards for Ethics in Data/AI Practice)
http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Open Data Institute, 'The Data Ethics Canvas'
https://theodi.org/the-data-ethics-canvas

APPENDIX B. BIBLIOGRAPHY/ADDITIONAL READING

Online Resources (see also Appendix A)
ABET (Accreditation Board for Engineering and Technology). http://www.abet.org/

ACM/IEEE-Computer Society. Software Engineering Code of Ethics and Professional Practice. Version 5.2. http://www.acm.org/about/se-code

Council for Big Data, Ethics & Society. http://bdes.datasociety.net/

Data & Society. https://datasociety.net/

National Academy of Engineering's Center for Engineering, Ethics and Society (CEES). http://www.nae.edu/26187.aspx

NSPE (National Society of Professional Engineers). Engineering Ethics. http://www.nspe.org/Ethics/index.html

Online Ethics Center for Engineering and Research. http://www.onlineethics.org/

Selected Books and Edited Collections (in reverse chronological order)
Bunnik, Anno et. al., Eds., (2016) *Big Data Challenges: Society, Security, Innovation and Ethics*, Palgrave Macmillan, 140 pages.

Collmann, Jeff and Matai, Sorin Adam, Eds., (2016) *Ethical Reasoning in Big Data: A Exploratory Analysis*, Springer, 192 pages.

Mittelstadt, Brent and Floridi, Luciano, Eds. (2016) *The Ethics of Biomedical Big Data*, Springer, 480 pages.

Lane, Julia, et al., Eds., (2014) *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press, 339 pages.

Spinello, Richard (2014) *Cyberethics: Morality and Law in Cyberspace*, 5th ed., Jones & Bartlett; 246 pages.

Tavani, Herman T. (2013) *Ethics and Technology: Controversies, Questions, and Strategies in Ethical Computing*, 4th Ed., John Wiley & Sons; 454 pages.

Davis, Kord (with Doug Patterson) (2012) *Ethics of Big Data: Balancing Risk and Innovation*, O'Reilly Media; 82 pages.

Solove, Daniel (2011) *Nothing to Hide: The False Tradeoff Between Privacy and Security*. Yale University Press; 256 pages.

Floridi, Luciano, ed. (2010) *The Cambridge Handbook of Information and Computer Ethics*, Cambridge University Press; 342 pages.

Johnson, Deborah G., ed. (2009) *Computer Ethics*, 4th ed., Pearson; 216 pages.

Nissenbaum, Helen (2009) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Stanford University Press; 304 pages.

Himma, Kenneth E. and Tavani, Herman T., eds., (2008) *The Handbook of Information and Computer Ethics*, John Wiley & Sons; 702 pages.

Weckert, John, ed. (2007) *Computer Ethics*, Ashgate; 516 pages.

Spinello, Richard and Tavani, Herman T. eds. (2004) *Readings in Cyberethics*, Jones and Bartlett; 697 pages.

Bynum, Terrell Ward and Rogerson, Simon, eds. (2004) *Computer Ethics and Professional Responsibility*, Blackwell; 378 pages.

Johnson, Deborah G. and Nissenbaum, Helen, eds. (1995) *Computers, Ethics & Social Values*, Prentice Hall; 656 pages.

Selected Articles and Encyclopedia Entries (in reverse chronological order)

Herschel, Richard and Miori, Virginia (2017) "Ethics & Big Data," *Technology in Society* 49, 31-36.

Buchanan, Elizabeth and Zimmer, Michael (2016) "Internet Research Ethics," The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), https://plato.stanford.edu/entries/ethics-internet-research/

Floridi, Luciano, and Taddeo, Mariarosaria (2016) "What is Data Ethics?" *Philosophical Transactions of the Royal Society A*, 374:2083, DOI: 10.1098/rsta.2016.0360. In special issue with the theme *The Ethical Impact of Data Science*, Taddeo and Floridi eds.

Metcalf, Jason and Crawford, Kate (2016) "Where are Human Subjects in Big Data Research? The Emerging Ethics Divide," *Big Data & Society* 3:1, DOI: 10.1177/2053951716650211

O'Leary, Daniel E. (2016) "Ethics for Big Data and Analytics," *IEEE Intelligent Systems*, 31:4, 81-84.

Crawford, Kate, et al. (2014) "Critiquing Big Data: Politics, Ethics, Epistemology." *International Journal of Communication*, 8:1663-1672.

Richards, Neil M. and King, Jonathan H. (2014) "Big Data Ethics," *Wake Forest Law Review*. Available at SSRN: https://ssrn.com/abstract=2384174

Zwitter, Andrej (2014) "Big Data Ethics," *Big Data & Society*, Jul-Dec, 1-6.

Moreno, M.A., et al. (2013) "Ethics of Social Media Research: Common Concerns and Practical Considerations." *Cyberpsychol Behav Soc Netw.* 16(9):708-13. doi: 10.1089/cyber.2012.0334.

Grodzinsky, Frances S., Miller, Keith W. and Wolf, Marty J. (2012) "Moral responsibility for computing artifacts: "the rules" and issues of trust." *ACM SIGCAS Computers and Society*, 42:2, 15-25.

Bynum, Terrell (2011) "Computer and Information Ethics", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2011/entries/ethics-computer/

Berenbach, Brian and Broy, Manfred (2009). "Professional and Ethical Dilemmas in Software Engineering." *IEEE Computer* 42:1, 74-80.

Erdogmus, Hakan (2009). "The Seven Traits of Superprofessionals." *IEEE Software* 26:4, 4-6.

Hall, Duncan (2009). "The Ethical Software Engineer." *IEEE Software* 26:4, 9-10.

Rashid, Awais, Weckert, John and Lucas, Richard (2009). "Software Engineering Ethics in a Digital World." *IEEE Computer* 42:6, p. 34-41.

Gotterbarn, Donald and Miller, Keith W. (2009) "The public is the priority: making decisions using the Software Engineering Code of Ethics." *IEEE Computer*, 42:6, 66-73.

Gotterbarn, Donald. (2008) "Once more unto the breach: Professional responsibility and computer ethics." *Science and Engineering Ethics* 14:1, 235-239.

Johnson, Deborah G. and Miller, Keith W. (2004) "Ethical issues for computer scientists." *The Computer Science and Engineering Handbook* 2nd Ed,, A. Tucker, ed. Springer-Verlag, 2.1-2.12.

Gotterbarn, Donald (2002) "Software Engineering Ethics," *Encyclopedia of Software Engineering*, 2nd ed., John Marciniak ed., John Wiley & Sons.

On General Philosophical Ethics
Aristotle (2011). *Nicomachean Ethics.* Translated by R.C. Bartlett and S.D. Collins. Chicago: University of Chicago Press.

Cahn, Steven M. (2010). *Exploring Ethics: An Introductory Anthology*, 2nd Edition. Oxford: Oxford University Press.

Shafer-Landau, Russ (2007). *Ethical Theory: An Anthology.* Oxford: Blackwell Publishing.