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).

HARMS TO TRANSPARENCY AND AUTONOMY since tamara and fred will get the same result everytime they apply for a loan because the system will always scan the same data causing autonomous and transparency harm to the couple.

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?

PREDICTIVE ACCURACY AND PERSONALIZATION because, From targeted disease therapies in medicine that are tailoredspecifically to a patients genetic fingerprint, to customized homework assignments that build
upon an individual students existing skills and focus on practice in areas of weakness, topredictive 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.

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?

Harm to fairness in justice because after a while the system will have caused a type of segregation between the people that have gotten their loans accepted and the ones that havent gotten their loans accepted. this could lead to massive controversy between the non accepted and the accepted.

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?

Yes because if someone takes a loan outof the bank the officer would have had to have an idea of why the applier would want to get a loan, wether that be for autonomy for starting a business or if it is a more fair loan system.

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?

invent a kind of verification system that lets the appliers prove they are low risk, sort of like a second chance for a loan if it firstly gets denied by the system. then use manpower for a second verification using previous experiences with loans.