Doing Good Data Science

The hard thing about being an ethical data scientist isn't understanding ethics. It's the junction between ethical ideas and practice. It's doing good data science.

There has been a lot of healthy discussion about data ethics lately. We want to be clear: that discussion is good, and necessary. But it's also not the biggest problem we face. We already have good standards for data ethics. The ACM's code of ethics, which dates back to 1993, and is currently being updated, is clear, concise, and surprisingly forward-thinking; 25 years later, it's a great start for anyone thinking about ethics. The American Statistical Association has a good set of ethical guidelines for working with data. So, we're not working in a vacuum.

And we believe that most people want to be fair. Data scientists and software developers don't want to harm the people using their products. There are exceptions, of course; we call them criminals and con artists. Defining "fairness" is difficult, and perhaps impossible, given the many crosscutting layers of "fairness" that we might be concerned with. But we don't have to solve that problem in advance, and it's not going to be solved in a simple statement of ethical principles, anyway.

The problem we face is different: how do we put ethical principles into practice? We're not talking about an abstract commitment to being fair. Ethical principles are worse than useless if we don't allow them to change our practice, if they don't have any effect on what we do day-to-day. For data scientists, whether you're doing classical data analysis or leading-edge AI, that's a big challenge. We need to understand how to build the software systems that implement fairness. That's what we mean by doing good data science.

Any code of data ethics will tell you that you shouldn't collect data from experimental subjects without informed consent. But that code won't tell you how to implement "informed consent." Informed consent is easy when you're interviewing a few dozen people in person for a psychology experiment. Informed consent means something different when someone clicks an item in an online catalog (hello, Amazon), and ads for that item start following them around *ad infinitum*. Do you use a pop-up to ask for permission to use their choice in targeted advertising? How many customers would you lose if you did so? Informed consent means something yet again when you're asking someone to fill out a profile for a social site, and you might (or might not) use that data for any number of experimental purposes. Do you pop up a consent form in impenetrable legalese that basically says "we will use your data, but we don't know for what"? Do you phrase this agreement as an opt-out, and hide it somewhere on the site where nobody will find it?

That's the sort of question we need to answer. And we need to find ways to share best practices. After the ethical principle, we have to think about the implementation of the ethical principle. That isn't easy; it encompasses everything from user experience design to data management. How do we design the user experience so that our concern for fairness and ethics doesn't make an application unuseable? Bad as it might be to show users a pop-up with thousands of words of legalese, laboriously guiding users through careful and lengthy explanations isn't likely to meet with approval, either. How do we manage any sensitive data that we acquire? It's easy to say that applications shouldn't collect data about race, gender, disabilities, or other protected classes. But if you don't gather that data, you will have trouble testing whether your applications are fair to minorities. Machine learning has proven to be very good at figuring its own proxies for race and other classes. Your application wouldn't be the first system that was unfair despite the best intentions of its developers. Do you keep the data you need to test for fairness in a separate database, with separate access controls?

To put ethical principles into practice, we need space to be ethical. We need the ability to have conversations about what ethics means, what it will cost, and what solutions to implement. As technologists, we frequently share best practices at conferences, write blog posts, and develop open source technologies---but we rarely discuss problems such as how to obtain informed consent.

There are several facets to this space that we need to think about.

Foremost, we need corporate cultures in which discussions about fairness, about the proper use of data, and about the harm that can be done by inappropriate use of data can be considered. In turn, this means that we can't rush products out the door without thinking about how they're used. We can't allow "internet time" to mean ignoring the consequences. Computer security has shown us the consequences of ignoring the consequences: many companies that have never taken the time to implement good security practices and safeguards are now paying with damage to their reputations and their finances. We need to do the same when thinking about issues like fairness, accountability, and unintended consequences.

We particularly need to think about the unintended consequences of our use of data. It will never be possible to predict all the unintended consequences; we're only human, and our ability to foresee the future is limited. But plenty of unintended consequences could easily have been foreseen: for example, Facebook's "Year in Review" that reminded people of deaths and other painful events. Moving fast and breaking things is unacceptable if we don't think about the things we are likely to break. And we need the space to do that thinking: space in project schedules, and space to tell management that a product needs to be rethought.

We also need space to stop the production line when something goes wrong. This idea goes back to Toyota's Kanban: any assembly line worker can stop the line if they see something going wrong. The line doesn't restart until the problem is fixed. Workers don't have have to fear consequences from management for stopping the line; they are trusted, and expected to behave responsibly. What would it mean if we could do this with product features? If anyone at Facebook could have said "wait, we're getting complaints about Year in Review" and pulled it out of production until someone could investigate what was happening?

It's easy to imagine the screams from management. But it's not hard to imagine a Toyota-style "stop button" working. After all, Facebook is the poster child for continuous deployment, and they've often talked about how new employees push changes to production on their first day. Why not let employees pull features out of production? Where are the tools for instantaneous undeployment? They certainly exist; continuous deployment doesn't make sense if you can't roll back changes that didn't work. Yes, Facebook is a big, complicated company, with a big complicated product. So is Toyota. It worked for them.

The issue lurking behind all of these concerns is, of course, corporate culture. Corporate environments can be hostile to anything other than short-term profitability. That's a consequence of poor court decisions and economic doctrine, particularly in the US. But that inevitably leads us to the biggest issue: how to move the needle on corporate culture. Susan Etlinger has suggested that, in a time when public distrust and disenchantment is running high, ethics is a good investment. Upper-level management is only starting to see this; changes to corporate culture won't happen quickly.

Users want to engage with companies and organizations they can trust not to take unfair advantage of them. Users want to deal with companies that will treat them and their data responsibly, not just as potential profit or engagement to be maximized. Those companies will be the ones that create space for ethics within their organizations. We, the data scientists, data engineers, AI and ML developers, and other data professionals, have to demand change. We can't leave it to people that "do" ethics. We can't expect management to hire trained ethicists and assign them to our teams. We need to live ethical values, not just talk about them. We need to think carefully about the consequences of our work. We must create space for ethics within our organizations. Cultural change may take time, but it will happen---if we are that change. That's what it means to do good data science.

Of Oaths and Checklists

"Oaths? We don't need no stinkin' oaths." (With apologies to Humphrey Bogart in *Treasure of the Sierra Madre*.)

Over the past year, there has been a great discussion of data ethics, motivated in part by discomfort over "fake news," targeted advertising, algorithmic bias, and the effect that data products have on individuals and on society. Concern about data ethics is hardly new; the ACM, IEEE, and the American Statistical Association all have ethical codes that address data. But the intensity with which we've discussed ethics shows that something significant is happening: data science is coming of age and realizing its responsibilities. A better world won't come about simply because we use data; data has its dark underside.

The recent discussion frequently veers into a discussion of data oaths, looking back to the ancient Hippocratic Oath for doctors. Much as we appreciate the work and the thought that goes into oaths, we are skeptical about their value. Oaths have several problems:

- They're one-shots. You take the oath once (if at all), and that's it. There's no reason to keep it in the front of your consciousness. You don't recite it each morning. Or evaluate regularly whether you're living up to the ideals.
- Oaths are a set of very general and broad principles. Discussions of the Hippocratic Oath begin with the phrase "First, do no harm," words that don't actually appear in the oath. But what does "do no harm" mean? For centuries doctors did very little but harm (many people died because doctors didn't believe they needed to wash their hands). The doctors just didn't know they were doing harm. Nice idea, but short on the execution. And data science (like medicine) is all about execution.
- Oaths can actually give cover to people and organizations who are doing unethical work. It's easy to think "we can't be unethical, because we endorsed this oath." It's not enough to say "don't be evil." You have to not be evil.
- Oaths do very little to connect theories and principles to practice. It is one thing to say "researchers must obtain informed consent"; it's an entirely different thing to get informed consent at internet scale. Or to teach users what "informed consent" means.

We are not suggesting that the principles embodied in oaths aren't important, just that they don't get us to the endpoint we want. They don't connect our ideas about what's good or just to the practices that create goodness and justice. We can talk a lot about the importance of being fair and unbiased without knowing about how to be fair and unbiased. At this point, the oath actually becomes dangerous: it becomes a tool to convince yourself that you're one of the good guys, that you're doing the right thing, when you really don't know.

Oaths are good at creating discussion---and, in the past year, they have created quite a lot of discussion. The discussion has been tremendously helpful in making people aware of issues like algorithmic fairness. The discussion has helped software developers and data scientists to understand that their work isn't value-neutral, that their work has real impact, both good and bad, on real people. And there has been a vigorous debate about what self-government means for data scientists, and what guiding principles would last longer than a few years. But we need to take the next step, and connect these ideas to practice. How will we do that?

In 2009, Atul Gawande wrote *The Checklist Manifesto* (Macmillan), a short book on how not to make big mistakes. He writes a lot about his practice as a surgeon. In a hospital, everyone knows what to do. Everyone knows that you're supposed to scrub down before the surgery. Everyone knows that you're not supposed to amputate the wrong leg. Everyone knows that you're not supposed to leave sponges and other equipment in patients when you close the incision.

But mistakes are made, particularly when people are in stressful environments. The surgeon operates on the wrong leg; the sponge is left behind; and so on. Gawande found that, simply by creating checklists for basic things you shouldn't forget, these mistakes could be eliminated almost completely. Yes, there were some doctors who found the idea of checklists insultingly simple; they were the ones who continued making mistakes.

Unlike oaths, checklists connect principle to practice. Everyone knows to scrub down before the operation. That's the principle. But if you have to check a box on a form after you've done it, you're not likely to forget. That's the practice. And checklists aren't one-shots. A checklist isn't something you read once at some initiation ceremony; a checklist is something you work through with every procedure.

What would a checklist for data science and machine learning look like? The UK Government's Data Ethics Framework and Data Ethics Workbook is one approach. They isolate seven principles, and link to detailed discussions of each. The workbook asks a number of open-ended questions to probe your compliance with these principles. Our criticism is that their process imposes a lot of overhead. While anyone going through their entire process will certainly have thought carefully about ethical issues, in practice, asking developers to fill out a workbook with substantive answers to 46 questions is an effective way to ensure that ethical thought doesn't happen.

We believe that checklists are built around simple, "have we done this?" questions---and they are effective because they are simple. They don't leave much room to wiggle. Either you've analyzed how a project can be abused, or you haven't. You've built a mechanism for gathering consent, or you haven't. Granted, it's still possible to take shortcuts: your analysis might be inadequate and your consent mechanism might be flawed, but you've at least gone on record for saying that you've done it.

Feel free to use and modify this checklist in your projects. It covers most of the bases that we've seen discussed in various data oaths. Go over the checklist when starting a project so the developers know what's needed and aren't surprised by a new set of requirements at the last minute. Then work through it whenever you release software. Go through it, and actually check off all the boxes before your product hits the public.

Here's a checklist for people who are working on data projects:

❏ Have we listed how this technology can be attacked or abused?

❏ Have we tested our training data to ensure it is fair and representative?

❏ Have we studied and understood possible sources of bias in our data?

❏ Does our team reflect diversity of opinions, backgrounds, and kinds of thought?

❏ What kind of user consent do we need to collect to use the data?

❏ Do we have a mechanism for gathering consent from users?

❏ Have we explained clearly what users are consenting to?

❏ Do we have a mechanism for redress if people are harmed by the results?

❏ Can we shut down this software in production if it is behaving badly?

❏ Have we tested for fairness with respect to different user groups?

❏ Have we tested for disparate error rates among different user groups?

❏ Do we test and monitor for model drift to ensure our software remains fair over time?

❏ Do we have a plan to protect and secure user data?

Oaths and codes of conduct have their value. The value of an oath isn't the pledge itself, but the process you go through in developing the oath. People who work with data are now having discussions that would never have taken place a decade ago. But discussions don't get the hard work done, and we need to get down to the hard work. We don't want to talk about how to use data ethically; we want to use data ethically. It's hypocritical to talk about ethics, but never do anything about it. We want to put our principles into practice. And that's what checklists will help us do.

The Five Cs

What does it take to build a good data product or service? Not just a product or service that's useful, or one that's commercially viable, but one that uses data ethically and responsibly.

We often talk about a product's technology or its user experience, but we rarely talk about how to build a data product in a responsible way that puts the user in the center of the conversation. Those products are badly needed. News that people "don't trust" the data products they use---or that use them---is common. While Facebook has received the most coverage, lack of trust isn't limited to a single platform. Lack of trust extends to nearly every consumer internet company, to large traditional retailers, and to data collectors and brokers in industry and government.

Users lose trust because they feel abused by malicious ads; they feel abused by fake and misleading content, and they feel abused by "act first, and apologize profusely later" cultures at many of the major online companies. And users ought to feel abused by many abuses they don't even know about. Why was their insurance claim denied? Why weren't they approved for that loan? Were those decisions made by a system that was trained on biased data? The slogan goes, "Move fast and break things." But what if society is broken?

Data collection is a big business. Data is valuable: "the new oil," as the Economist proclaimed. We've known that for some time. But the public provides the data under the assumption that we, the public, benefit from it. We also assume that data is collected and stored responsibly, and those who supply the data won't be harmed. Essentially it's a model of trust. But how do you restore trust once it's been broken? It's no use pretending that you're trustworthy when your actions have proven that you aren't. The only way to get trust back is to be trustworthy, and regaining that trust once you've lost it takes time.

There's no simple way to regain users' trust, but we'd like to suggest a "golden rule" for data as a starting point: "treat others' data as you would have others treat your own data." However, implementing a golden rule in the actual research and development process is challenging---just as it's hard to get from short, pithy oaths and pledges to actual practice.

What does it mean to treat others' data as you would treat your own? How many data scientists have actually thought about how their own data might be used and abused? And once you know how you'd like to see your data (and others' data) respected, how do you implement those ideas? The golden rule isn't enough by itself. We need guidelines to force discussions with the application development teams, application users, and those who might be harmed by the collection and use of data.

Five framing guidelines help us think about building data products. We call them the five Cs: consent, clarity, consistency, control (and transparency), and consequences (and harm). They're a framework for implementing the golden rule for data. Let's look at them one at a time.

Consent

You can't establish trust between the people who are providing data and the people who are using it without agreement about what data is being collected and how that data will be used. Agreement starts with obtaining consent to collect and use data. Unfortunately, the agreements between a service's users (people whose data is collected) and the service itself (which uses the data in many ways) are binary (meaning that you either accept or decline) and lack clarity. In business, when contracts are being negotiated between two parties, there are multiple iterations (redlines) before the contract is settled. But when a user is agreeing to a contract with a data service, they either accept the terms or they don't get access. It's nonnegotiable.

For example, when you check into a hospital you are required to sign a form that gives them the right to use your data. Generally, there's no way to say that your data can be used for some purposes but not others. When you sign up for a loyalty card at your local pharmacy, you're agreeing that they can use your data in unspecified ways. Those ways certainly include targeted advertising (often phrased as "special offers"), but may also include selling your data (with or without anonymization) to other parties. And what happens to your data when one company buys another and uses data in ways that you didn't expect?

Data is frequently collected, used, and sold without consent. This includes organizations like Acxiom, Equifax, Experian, and Transunion, that collect data to assess financial risk, but many common brands also collect data without consent. In Europe, Google collected data from cameras mounted on cars to develop new mapping products. AT&T and Comcast both used cable set top boxes to collect data about their users, and Samsung collected voice recordings from TVs that respond to voice commands. There are many, many more examples of nonconsensual data collection. At every step of building a data product, it is essential to ask whether appropriate and necessary consent has been provided.

Clarity

Clarity is closely related to consent. You can't really consent to anything unless you're told clearly what you're consenting to. Users must have clarity about what data they are providing, what is going to be done with the data, and any downstream consequences of how their data is used. All too often, explanations of what data is collected or being sold are buried in lengthy legal documents that are rarely read carefully, if at all. Observant readers of Eventbrite's user agreement recently discovered that listing an event gave the company the right to send a video team, and exclusive copyright to the recordings. And the only way to opt out was by writing to the company. The backlash was swift once people realized the potential impact, and Eventbrite removed the language.

Facebook users who played Cambridge Analytica's "This Is Your Digital Life" game may have understood that they were giving up their data; after all, they were answering questions, and those answers certainly went somewhere. But did they understand how that data might be used? Or that they were giving access to their friends' data behind the scenes? That's buried deep in Facebook's privacy settings.

Even when it seems obvious that their data is in a public forum, users frequently don't understand how that data could be used. Most Twitter users know that their public tweets are, in fact, public; but many don't understand that their tweets can be collected and used for research, or even that they are for sale. This isn't to say that such usage is unethical; but as Casey Fiesler points out, the need isn't just to get consent, but to inform users what they're consenting to. That's clarity.

It really doesn't matter which service you use; you rarely get a simple explanation of what the service is doing with your data, and what consequences their actions might have. Unfortunately, the process of consent is often used to obfuscate the details and implications of what users may be agreeing to. And once data has escaped, there is no recourse. You can't take it back. Even if an organization is willing to delete the data, it's very difficult to prove that it has been deleted.

There are some notable exceptions: people like John Wilbanks are working to develop models that help users to understand the implications of their choices. Wilbanks' work helps people understand what happens when they provide sensitive medical and health data to a service.

Consistency and Trust

Trust requires consistency over time. You can't trust someone who is unpredictable. They may have the best intentions, but they may not honor those intentions when you need them to. Or they may interpret their intentions in a strange and unpredictable way. And once broken, rebuilding trust may take a long time. Restoring trust requires a prolonged period of consistent behavior.

Consistency, and therefore trust, can be broken either explicitly or implicitly. An organization that exposes user data can do so intentionally or unintentionally. In the past years, we've seen many security incidents in which customer data was stolen: Yahoo!, Target, Anthem, local hospitals, government data, data brokers like Experian, and the list grows longer each day. Failing to safeguard customer data breaks trust---and safeguarding data means nothing if not consistency over time.

We've also seen frustration, anger, and surprise when users don't realize what they've agreed to. When Cambridge Analytica used Facebook's data to target vulnerable customers with highly specific advertisements, Facebook initially claimed that this was not a data breach. And while Facebook was technically correct, in that data was not stolen by an intruder, the public's perception was clearly different. This was a breach of trust, if not a breach of Facebook's perimeter. Facebook didn't consistently enforce its agreement with its customers. When the news broke, Facebook became unpredictable because most of its users had no idea what it would or wouldn't do. They didn't understand their user agreements, they didn't understand their complex privacy settings, and they didn't understand how Facebook would interpret those settings.

Control and Transparency

Once you have given your data to a service, you must be able to understand what is happening to your data. Can you control how the service uses your data? For example, Facebook asks for (but doesn't require) your political views, religious views, and gender preference. What happens if you change your mind about the data you've provided? If you decide you're rather keep your political affiliation quiet, do you know whether Facebook actually deletes that information? Do you know whether Facebook continues to use that information in ad placement?

All too often, users have no effective control over how their data is used. They are given all-or-nothing choices, or a convoluted set of options that make controlling access overwhelming and confusing. It's often impossible to reduce the amount of data collected, or to have data deleted later.

A major part of the shift in data privacy rights is moving to give users greater control of their data. For example, Europe's General Data Protection Regulation (GDPR) requires users' data to be provided to them at their request and removed from the system if they so desire.

Consequences

Data products are designed to add value for a particular user or system. As these products increase in sophistication, and have broader societal implications, it is essential to ask whether the data that is being collected could cause harm to an individual or a group. We continue to hear about unforeseen consequences and the "unknown unknowns" about using data and combining data sets. Risks can never be eliminated completely. However, many unforeseen consequences and unknown unknowns could be foreseen and known, if only people had tried. All too often, unknown unknowns are unknown because we don't want to know.

Due to potential issues around the use of data, laws and policies have been put in place to protect specific groups: for example, the Children's Online Privacy Protection Act (COPPA) protects children and their data. Likewise, there are laws to protect specific sensitive data sets: for example, the Genetic Information Nondiscrimination Act (GINA) was established in 2008 in response to rising fears that genetic testing could be used against a person or their family. Unfortunately, policy doesn't keep up with technology advances; neither of these laws have been updated. Given how rapidly technology is being adopted by society, the Obama administration realized that the pace of the regulatory process couldn't keep up. As a result, it created the roles of the US chief technology officer and chief data scientist. The Obama administration also established more than 40 chief data officers and scientists across the federal government. The result has been to make sure the regulatory process fosters innovation while ensuring the question of potential of harm is asked regularly and often.

Even philanthropic approaches can have unintended and harmful consequences. When, in 2006, AOL released anonymized search data to researchers, it proved possible to "deanonymize" the data and identify specific users. In 2018, Strava opened up their data to allow users to discover new places to run or bike. Strava didn't realize that members of the US military were using GPS-enabled wearables, and their activity exposed the locations of bases and patrol routes in Iraq and Afghanistan. Exposure became apparent after the product was released to the public, and people exploring the data started talking about their concerns.

While Strava and AOL triggered a chain of unforeseen consequences by releasing their data, it's important to understand that their data had the potential to be dangerous even if it wasn't released publicly. Collecting data that may seem innocuous and combining it with other data sets has real-world implications. Combining data sets frequently gives results that are much more powerful and dangerous than anything you might get from either data set on its own. For example, data about running routes could be combined with data from smart locks, telling thieves when a house or apartment was unoccupied, and for how long. The data could be stolen by an attacker, and the company wouldn't even recognize the damage.

It's easy to argue that Strava shouldn't have produced this product, or that AOL shouldn't have released their search data, but that ignores the data's potential for good. In both cases, well-intentioned data scientists were looking to help others. The problem is that they didn't think through the consequences and the potential risks.

It is possible to provide data for research without unintended side-effects. For example, the US Internal Revenue Service (IRS), in collaboration with researchers, opened a similar data set in a tightly controlled manner to help understand economic inequality. There were no negative repercussions or major policy implications. Similarly the Department of Transportation releases data about traffic fatalities. The UK Biobank (one of the largest collections of genomic data) has a sophisticated approach to opening up different levels of data. Other companies have successfully opened up data for the public benefit, including LinkedIn's Economic Graph project and Google Books' ngram viewer.

Many data sets that could provide tremendous benefits remain locked up on servers. Medical data that is fragmented across multiple institutions limits the pace of research. And the data held on traffic from ride-sharing and GPS/mapping companies could transform approaches for traffic safety and congestion. But opening up that data to researchers requires careful planning.

Implementing the Five Cs

Data can improve our lives in many ways, from the mundane to the amazing. Good movie recommendations aren't a bad thing; if we could consolidate medical data from patients around the world, we could make some significant progress on treating diseases like cancer. But we won't get either better movie recommendations or better cancer treatments if we can't ensure that the five Cs are implemented effectively. We won't get either if we can't treat others' data as carefully as we'd treat our own.

Over the past decade, the software industry has put significant effort into improving user experience (UX). Much of this investment has been in user-centric approaches to building products and services that depend on the data the collective user base provides. All this work has produced results: using software is, on the whole, easier and more enjoyable. Unfortunately, these teams have either intentionally or unintentionally limited their efforts to providing users with immediate gratification or the ability to accomplish near-term goals. "Growth hacking" focuses on getting people to sign up for services through viral mechanisms. We've seen few product teams that try to develop a user experience that balances immediate experience with long-term values.

In short, product teams haven't considered the impacts of the five Cs. For example, how should an application inform users about how their data will be used, and get their consent? That part of user experience can't be swept under the rug. And it can't mean making it easy for users to give consent, and difficult to say "no." It's all part of the total user experience. Users need to understand what they are consenting to and what effects that consent might have; if they don't, the designer's job isn't done.

Responsibility for the five Cs can't be limited to the designers. It's the responsibility of the entire team. The data scientists need to approach the problem asking "what if" scenarios that get to all of the five C's. The same is true for the product managers, business leaders, sales, marketing, and also executives.

The five Cs need to be part of every organization's culture. Product and design reviews should go over the five Cs regularly. They should consider developing a checklist before releasing a product to the public. All too often, we think of data products as minimal viable products (MVPs: prototypes to test whether the product has value to users). While that's a constructive approach for developing and testing new ideas, even MVPs must address the five Cs. The same is true for well-established products. New techniques may have been developed that could result in harm in unforeseen ways. In short, it's about taking responsibility for the products that are built. The five Cs are a mechanism to foster dialogue to ensure the products "do no harm."

Data's Day of Reckoning

Our lives are bathed in data: from recommendations about whom to "follow" or "friend" to data-driven autonomous vehicles. But in the past few years, it has become clear that the products and technologies we have created have been weaponized and used against us. Although we've benefited from the use of data in countless ways, it has also created a tension between individual privacy, public good, and corporate profits. Cathy O'Neil's *Weapons of Math*

Destruction (Broadway Books) and Virginia Eubanks' *Automating Inequality* (Macmillan) document the many ways that data has been used to harm the broader population.

Data science, machine learning, artificial intelligence, and related technologies are now facing a day of reckoning. It is time for us to take responsibility for our creations. What does it mean to take responsibility for building, maintaining, and managing data, technologies, and services? Responsibility is inevitably tangled with the complex incentives that surround the creation of any product. These incentives have been front and center in the conversations around the roles that social networks have played in the 2016 US elections, recruitment of terrorists, and online harassment. It has become very clear that the incentives of the organizations that build and own data products haven't aligned with the good of the people using those products.

These issues aren't new to the consumer internet. Other fields have had their days of reckoning. In medicine, widely publicized abuses include the Tuskegee syphilis experiment, the case of Henrietta Lacks (whose cells were used for cancer research without her permission and without compensation), and human experiments performed during World War II by Nazis. The physics community had to grapple with the implications of the atomic bomb. Chemists and biologists have had to address the use of their research for chemical and biological weapons. Other engineering disciplines have realized that shoddy work has an impact on people's lives; it's hard to ignore bridge collapses. As a result, professional societies were formed to maintain and enforce codes of conduct; government regulatory processes have established standards and penalties for work that is detrimental to society.

Ethics and Security Training

In many fields, ethics is an essential part of professional education. This isn't true in computer science, data science, artificial intelligence, or any related field. While courses on ethics exist at many schools, the ideas taught in ethics classes often aren't connected to existing projects or course work. Students may study ethical principles, but they don't learn how to implement those principles in their projects. As a result, they are ill-prepared for the challenges of the real world. They're not trained to think about ethical issues and how they affect design choices. They don't know how to have discussions about projects or technologies that may cause real-world harm.

Software security and ethics frequently go hand in hand, and our current practices for teaching security provide an example of what not to do. Security is usually taught as an elective, isolated from other classes about software development. For example, a class on databases may never discuss SQL injection attacks. SQL injection would be addressed in classes on security, but not in the required database course. When a student submits a project in a database course, its vulnerability to hostile attack doesn't affect the grade; an automated grading system won't even test it for vulnerabilities. Furthermore, a database course might not discuss architectural decisions that limit damage if an attacker gains access---for example, storing data elements such as names and Social Security numbers in different databases.

Teaching security in an elective is better than not teaching it at all; but the the best way to produce programmers who really understand security is to incorporate it into assignments and grading within the core curriculum, in addition to teaching it in electives. The core curriculum ensures that everyone can recognize and deal with basic security problems; the electives can go into greater depth, and tie together issues from different disciplines ranging from physical security to cryptography. Security as an afterthought doesn't work in product development; why do we expect it to work in education? There is no industry in which security lapses haven't led to stolen data, affecting millions of individuals. Poor security practices have led to serious vulnerabilities in many consumer devices, from smart locks to smart light bulbs.

Ethics faces the same problem. Data ethics is taught at many colleges and universities, but it's isolated from the rest of the curriculum. Courses in ethics help students think seriously about issues, but can't address questions like getting informed consent in the context of a real-world application. The White House report "Preparing for the Future of Artificial Intelligence" highlights the need for training in both ethics and security:

Ethical training for AI practitioners and students is a necessary part of the solution. Ideally, every student learning AI, computer science, or data science would be exposed to curriculum and discussion on related ethics and security topics. However, ethics alone is not sufficient. Ethics can help practitioners understand their responsibilities to all stakeholders, but ethical training should be augmented with technical tools and methods for putting good intentions into practice by doing the technical work needed to prevent unacceptable outcomes.

Ethics and security must be at the heart of the curriculum, not only as electives, or even isolated requirements. They must be integrated into every course at colleges, universities, online courses, and programming boot camps. They can't remain abstract, but need to be coupled with "technical tools and methods for putting good intentions into practice." And training can't stop upon graduation. Employers need to host regular forums and offer refresher courses to keep people up-to-date on the latest challenges and perspectives.

Developing Guiding Principles

The problem with ethical principles is that it's easy to forget about them when you're rushing: when you're trying to get a project finished on a tight, perhaps unrealistic, schedule. When the clock is ticking away toward a deadline, it's all too easy to forget everything you learned in class---even if that class connected ethics with solutions to real-world problems.

Checklists are a proven way to solve this problem. A checklist, as described by Atul Gawande in *The Checklist Manifesto* (Metropolitan Books), becomes part of the ritual. It's a short set of questions that you ask at the start of the project, and at every stage as you move toward release. You don't go to the next stage until you've answered all the questions affirmatively. Checklists have been shown to reduce mistakes in surgery; they're used very heavily by airline pilots, especially in emergencies; and they can help data professionals to not forget ethical issues, even when they are under pressure to deliver.

In Chapter 2, we proposed a checklist for developers working on data-driven applications. Feel free to use and to modify this checklist to fit your situation and use it in your projects. Our checklist doesn't reflect all the issues that you should be considering, and certainly doesn't reflect all the applications that people are currently developing, let alone in the future. If you add to it, though, try to keep the additions short; that's why checklists work.

The Fairness, Accountability, and Transparency in Machine Learning group (FAT/ML) advocates a similar approach. Their Principles for Accountable Algorithms and a Social Impact Statement for Algorithms suggests assessing the social impact statement of a project at least three times during its life: during design, pre-launch, and post-launch. Working through a social impact statement requires developers to think about the ethical consequences of their projects and address any problems that turn up. In a similar vein, the Community Principles on Ethical Data Practices, which arose out of the Data for Good Exchange (D4GX), provides a set of values and principles that have been gathered through community discussion. They're a great start for any group that wants to create its own checklist. And Cathy O'Neil has proposed auditing machine learning algorithms for fairness.

Building Ethics into a Data-Driven Culture

Individual responsibility isn't sufficient. Ethics needs to be part of an organization's culture. We've seen many organizations recognize the value of developing a data-driven culture; we need to ensure ethics and security become part of that culture, too.

Security is gradually becoming a part of corporate culture: the professional, financial, legal, and reputational consequences of being a victim are too large to ignore. Organizations are experimenting with bug-bounty programs, sharing threats with each other, and collaborating with government agencies. Security teams are no longer simply corporate naysayers; they're charged with preventing serious damage to an organization's reputation and to finances.

Integrating ethics into corporate culture has been more challenging. A single team member may object to an approach, but it's easy for an individual to be overruled, and if there's no support for ethical thinking within the organization, that's likely to be where it ends. Ethical thinking is important with or without corporate support, but it's more likely to make a difference when ethical action is a corporate value. Here are some ideas for building ethics into culture:

An individual needs to be empowered to stop the process before damage is done.

Toyota and W. Edwards Deming pioneered the use of the andon cord to improve quality and efficiency. Anyone who saw a problem could pull the cord, which would halt the production line. Senior managers as well as production line operators would then discuss the issue, make improvements, and restart the process.

Any member of a data team should be able to pull a virtual "andon cord," stopping production, whenever they see an issue. The product or feature stays offline until the team has a resolution. This way, an iterative process can be developed that avoids glossing over issues.

Anyone should be able to escalate issues for remediation without fear of retaliation.

There needs to be an escalation process for team members who don't feel their voice has been heard. The US Department of State has a dissent channel where any diplomat can make sure the Secretary of State hears their concerns. In health care, a path to escalate legal and ethical issues is required by law. For health care plans in the US, there is a compliance officer who reports directly to the board of directors.

Data-driven organizations need a similar model that allows people to escalate issues without the fear of reprisal. An escalation process could be implemented in several forms. For example, companies could work with an organization such as the Electronic Frontier Foundation (EFF) to develop a program that accepts and investigates whistleblower reports. The problem would be kept from public scrutiny unless specific criteria are violated. A similar approach could be implemented under an existing or new agency (e.g., a Consumer Data Protection Agency).

An ethical challenge should be part of the hiring process.

When hiring, companies frequently assess whether a candidate will be a "cultural fit." Interviewers ask questions that help them understand whether a candidate will work well with other team members. However, interviewers rarely ask questions about the candidate's ethical values.

Rather than asking a question with a right/wrong answer, we've found that it's best to pose a problem that lets us see how the candidate thinks about ethical and security choices. Here's a question we have used:

Assume we have a large set of demographic data. We're trying to evaluate individuals and we're not supposed to use race as an input. However, you discover a proxy for race with the other variables. What would you do?

This kind of question can start a dialogue about how to use the proxy variable. What effects does it have on people using the product? Are we making recommendations, or deciding whether to provide services? Are we implementing a legal requirement, or providing guidance about compliance? Discussing the question and possible answers will reveal the candidate's values.

Product reviews must ask questions about the product's impact.

Environmental impact statements predict the impact of construction projects on the public. We've already mentioned FAT/ML's proposed Social Impact Statements as an example of what might be done for data. In the social sciences and the biomedical industry, Institutional Review Boards (IRBs) assess the possible consequences of experiments before they're performed.

While both environmental impact statements and IRBs present problems for data products, data teams need to evaluate the impact of choices they make. Teams need to think about the consequences of their actions before releasing products. We believe that using a checklist is the best approach for ensuring good outcomes.

Teams must reflect diversity of thought, experiences, race, and background.

All too often, we hear about products that are culturally insensitive or overtly racist. One notorious example is an automated passport control system that doesn't let an individual proceed until a good digital image is captured. People of Asian ancestry reported that the system kept asking them to open their eyes, even though their eyes were open. Many cringe-worthy examples are well documented; they can often be traced to a lack of data or a lack of insight into the diversity of the population that will be impacted.

While there's no general solution to these problems of cultural sensitivity, diversity and inclusion are a tremendous help. Team members should be from the populations that will be impacted. They'll see issues well before anyone else. External peer reviews can help to reveal ethical issues that your team can't see. When you're deeply involved with a project, it can be hard to recognize problems that are obvious to outsiders.

Corporations must make their own principles clear.

Google's "Don't be evil" has always been a cute, but vague, maxim. Their recent statement, Artificial Intelligence at Google: Our Principles, is more specific. In a similar vein, the face recognition startup Kairos has said that they won't do business with law enforcement companies. Kairos' CEO writes that "the use of commercial face recognition in law enforcement or government surveillance of any kind is wrong."

However, it's important to realize that advocating for corporate ethical principles has consequences. Significant internal protest, and the resignation of several developers in protest over Google's defense contracts, were needed to get their AI principles in place. Kairos is probably leaving a lot of money on the table. It's also important to realize that organizations frequently point to their ethical principles to divert attention from unethical projects.

Over the past few years, we've heard a lot about software startups that begin with a "minimal viable product," and adhere to Facebook's slogan, "move fast and break things." Is that incompatible with the approach we've just described? This is a false choice. Going fast doesn't mean breaking things. It is possible to build quickly and responsibly.

The lean/agile methodology used in many startups is a good way to expose ethical issues before they become problems. Developers start with a very simple product ("the simplest thing that could possibly work," according to Ward Cunningham's seminal phrase), demo it to users, get feedback, develop the next version, and repeat. The process continues for as many iterations as needed to get a product that's satisfactory. If a diverse group of users tests the product, the product development loop is likely to flush out systematic problems with bias and cultural insensitivity. The key is testing the product on a truly diverse group of users, not just a group that mirrors the expected customer base or the developers' backgrounds.

Regulation

In some industries, ethical standards have been imposed by law and regulation. The Nuremberg Code was developed in response to Nazi atrocities. It focuses on individual consent to participation in an experiment. After the Tuskegee syphilis experiments became public knowledge, the code was put into law in the 1974 National Research Act and the 1975 Declaration of Helsinki. This push to codify ethical guidelines established the role of the institutional review board (IRB), and was adopted widely in the US via the Common Rule.

In other industries, other regulatory bodies enforce ethical standards. These include the US Federal Trade Commission (FTC), which oversees commerce; the Nuclear Regulatory Commission (NRC), which oversees nuclear power plants; the Federal Food and Drug Administration (FDA), which oversees the safety of pharmaceuticals; and, most recently, the Consumer Finance Protection Bureau (CFPB), which oversees bankers and lenders on behalf of consumers.

The European Union's General Data Protection Regulation (GDPR) takes an aggressive approach to regulating data use and establishing a uniform data policy. In June 2018, California passed a digital privacy law similar to GDPR, despite the reservations of many online companies. One challenge of developing a policy framework is that the policy development process nearly always lags the pace of innovation, and isn't agile enough to keep policy iterative. By the time a policy has been formulated and approved, it almost always lags behind technology; but it's impossible for policy makers to iterate quickly enough to catch up with the newest technology.

Another problem is that the committees that make policy often lack experts with the necessary technical background. That can be good; technologists are too easily influenced by "what technology wants." But policies created by people who are technologically uninformed are frequently out of touch with reality: look at the debate over back doors to encryption protocols.

Some have argued that organizations using data should adopt the Institutional Review Board (IRB) model from the biomedical industry. Unfortunately, while there are many positive aspects of the IRB, this isn't a viable approach. IRBs are complex, and they can't be agile; it's very difficult for IRBs to adapt to new ideas and technologies. It's why the Obama administration pushed for nearly eight years to update the Common Rule's models for consent to be consistent with digital technologies and to enable data mining.

Building Our Future

For some time, we've been aware of the ethical problems that arise from the use and abuse of data. Public outcry over Facebook will die down eventually, but the problems won't. We're looking at a future in which most vehicles are autonomous; we will be talking to robots with voices and speech patterns that are indistinguishable from humans; and where devices are listening to all our conversations, ready to make helpful suggestions about everything from restaurants and recipes to medical procedures. The results could be wonderful---or they could be a nightmarish dystopia.

It's data's day of reckoning. The shape of the future will depend a lot on what we do in the next few years. We need to incorporate ethics into all aspects of technical education and corporate culture; we need to give people the freedom to stop production if necessary, and to escalate concerns if they're not addressed; we need to incorporate diversity and ethics into hiring decisions; and we may need to consider regulation to protect the interests of individual users, and society as a whole.

Above all, talk about ethics! In "It's time for data ethics conversations at the dinner table," Natalie Evans Harris and others write that "we need to be having difficult conversations about our individual and collective responsibility to handle data ethically." This is the best single thing you can do to further data ethics: talk about it in meetings, at lunch, and even at dinner. Signing a data oath, or agreeing to a code of conduct, does little if you don't live and breathe ethics. Once you are living and breathing it, you will start to think differently about the code you write, the models you build, and the applications you create. The only way to create an ethical culture is to live it. The change won't take place magically, nor will it be easy---but it's necessary.

We can build a future we want to live in, or we can build a nightmare. The choice is up to us.

Case Studies

To help us think seriously about data ethics, we need case studies that we can discuss, argue about, and come to terms with as we engage with the real world. Good case studies give us the opportunity to think through problems before facing them in real life. And case studies show us that ethical problems aren't simple. They are multifaceted, and frequently there's no single right answer. And they help us to recognize there are few situations that don't raise ethical questions.

Princeton's Center for Information Technology Policy and Center for Human Values have created four anonymized case studies to promote the discussion of ethics. (More are in the pipeline, and may be available by the time you read this.) The first of these studies, Automated Healthcare App, discusses a smartphone app designed to help adult onset diabetes patients. It raises issues like paternalism, consent, and even language choices. Is it OK to "nudge" patients toward more healthy behaviors? What about automatically moderating the users' discussion groups to emphasize scientifically accurate information? And how do you deal with minorities who don't respond to treatment as well? Could the problem be the language itself that is used to discuss treatment?

The next case study, Dynamic Sound Identification, covers an application that can identify voices, raising issues about privacy, language, and even gender. How far should developers go in identifying potential harm that can be caused by an application? What are acceptable error rates for an application that can potentially do harm? How can a voice application handle people with different accents or dialects? And what responsibility do developers have when a small experimental tool is bought by a large corporation that wants to commercialize it?

The Optimizing Schools case study deals with the problem of finding at-risk children in school systems. Privacy and language are again an issue; it also raises the issue of how decisions to use data are made. Who makes those decisions, and who needs to be informed about them? What are the consequences when people find out how their data has been used? And how do you interpret the results of an experiment? Under what conditions can you say that a data experiment has really yielded improved educational results?

The final case study, Law Enforcement Chatbots, raises issues about the trade-off between liberty and security, entrapment, openness and accountability, and compliance with international law.

None of these issues are simple, and there are few (if any) "right answers." For example, it's easy to react against perceived paternalism in a medical application, but the purpose of such an application is to encourage patients to comply with their treatment program. It's easy to object to monitoring students in a public school, but students are minors, and schools by nature handle a lot of private personal data. Where is the boundary between what is, and isn't, acceptable? What's important isn't getting to the correct answer on any issue, but to make sure the issue is discussed and understood, and that we know what trade-offs we are making. What is important is that we get practice in discussing ethical issues and put that practice to work in our jobs. That's what these case studies give us.