Week 3 transcript

If you require closed captioning for accommodation reasons, please contact the instructor and we will make the necessary accommodations.

3.1.2. Types of Questions

A major theme in this class is to work on how to ask better questions. So how do we do that? Think about the scope of the question. We can think of data science questions that are specific, and that kind of feel like data science questions. For example, how can we optimize a process? How can we identify new customers? How can we identify high quality applicants to recruit to our university, and not just grade point average?

But we could also expand our scope to think about questions that don't immediately appear like data science questions at first glance. For example, how does the university accommodate for interruptions due to power outages, poor air quality, or health concerns? Another way to ask better questions is to listen to the ask and mirror back when your client, manager, or colleague asks to make sure you understand.

Make sure it's clear what the problem is. Listen, think big, and be open to a broad range of approaches to tackle the question. Be careful not to allow known methods and data sources to narrow down your scope.

3.1.3. Ask the Right Question: A National Football League Example

OK. So in this segment, I want to delve a little bit more into the context of asking the right questions, using the right data for the right reasons. It seems obvious on the face of it, but let's play with it a little bit for a moment in an example that is very close to my heart that involves the 2017 Super Bowl.

In the 2017 Super Bowl, if you were following the NFL over the course of that season, one of the things that was really different about it was that in that game, every player had a sensor in his shoulder pad that tracked movement, speed, and direction. And that data is available, or at least parts of it are available, if you go to Next Gen Stats on nfl.com.

Now the NFL has been behind the Major League Baseball in this for quite a while. For those of you familiar, we'll talk about it a bit later in the course. Major League Baseball has been collecting variants of this kind of data for almost a decade. There's a product called PITCHf/x which is aimed at pitchers to understand their velocity, their throwing motion, and all sorts of other things.

The question that I want to pose here is, how useful has that data actually been? Well, interestingly, in the football world, you would think it would really be valuable to know, say, for

example, who's running the fastest, who can cover the most ground on defense. But not so much if you're actually in the football world itself.

In fact, what the teams do depends on what are reflected by themselves acknowledged to be very subjective decisions. And as you know, they're made by coaches, and those coaches are people who tend to watch videos in the Film Room. And they call it the Film Room. They don't call it the data room.

In fact, Matt Patricia, who is the Defensive Coordinator for the New England Patriots-- many of you maybe would recognize the face of the bearded guy, he stands on the sideline next to Bill Belichick-- has explicitly said and repeated after the New England Patriots' victory in the 2017 Super Bowl that-- well, I'll just leave it alone, that terrible moment. Patricia said he never actually uses the data. He doesn't even want to see it.

And it's a fair question to ask, why not? I mean, every single Monday in 2015 through 2016, he was given a report on what would seem to be really useful information. He got the speed of his players, the consistency of their running routes. He could even get the separation between a wide receiver and the defender. And again, you can see some of those stats.

But interestingly, he didn't seem to care. What he wanted was data he didn't get. He no longer was allowed to get data on the brute force of hits on the helmet because of the controversy over concussions in the NFL. And so Patricia was not getting what he wanted, and he was getting what he didn't want.

He said he was interested in other things which are much, much harder for those sensors to quantify, like what kind of a block has that linemen just used? Is the receiver using his head or his body to fake out the defender? Or a critical strategic question which he really wanted to answer, like, is the quarterback getting rid of the ball quickly because that quarterback has really fantastic vision and really fast receivers who are getting open? Or is he getting rid of the ball so quickly because he's just scared to death of taking a sack? And that's something that Matt Patricia would really want to know, but that's data that he wasn't getting.

So look, wrap that up. What's really happening here? The NFL is spending a ton of money, time, people, resources, political capital getting beautiful data and creating beautiful data visualizations and some wonderful heat maps on the football field, but it's actually not changing the game. And the proposition here is maybe the NFL can afford to make those mistakes. But in most of the organizations that you're going to be working in and in the organization that I work in, UC Berkeley, we can't actually afford to make those kind of mistakes. We have to get the questions right upfront.

3.1.4. How to Reframe and Iterate

Part of our job is to help reframe the questions we are asked. We'll talk elsewhere more about how to mirror back and how to ask better questions. But here, I want to focus on the iterative nature of question formation.

It will often take a few tries to come up with well-formulated questions. Well-formulated questions have a clear outcome, and it has a clear outcome that we care about. Remember, most of the projects you all do will not be strictly academic. So instead, there's a specific decision you want to influence.

Well-formulated questions can also be answered with the time frame and the resources that are available. Question formation and the overall research process is an iterative thing, so let's go through an example.

Initially, we may come up with a question that we are content with. We may start looking at data sources or measurement, or we might start talking to people inside or outside of our organization. Then we might realize we're a little off the mark, so we go back to the question phase.

Maybe you're asking a question about how to predict a machine failure and what maintenance schedule we should implement. But then you start talking to the mechanics in the field and this is not how they think about this. They say the question doesn't pass the face validity test, or it doesn't pass the sniff test.

So you go back to question formation. Then you go to the data you have. And then you find out the data you thought you needed to answer the question doesn't exist, and you decide it's not worth the pain to gather that new information. So you might say, all right, initially, we wanted to answer question number one. We can't. Question number two is pretty close, but let's be clear that these are two distinct questions.

And so you might kind of go back to your manager and say, we're not going to be able to gather the new data and answer question one with the time constraints, but we can answer question two with the data we have and the existing time frame.

The headline takeaway is don't let the perfect get in the way of the good, but let me end with a final word of caution here. In this example, you wanted to pursue question number one, but you couldn't, and so you pursued question two. You must be very precise in this example. You don't want to unintentionally claim that the second approach is the same as the first. Be very precise about what evidence supports what arguments.

3.3.1. Introduction to Qualitative Approaches

Before I describe the qualitative approach, I want to briefly discuss the difference between a research approach, a research design, and a research method. As I go through the definitions, hopefully you'll see that we get more and more specific.

So a research approach-- this is your broad plan. It includes your philosophical world view about research, your design, and your methods. Design-- in the design, you articulate the type of inquiry, qualitative, quantitative, mixed, and the specific procedures in the study. You can think of this as the design as a recipe. This is what we plan to do, and this is why.

And finally, the method. The method is your planned data collection, your analysis, and your interpretation. And we'll continue to discuss approach design and methods throughout the course.

Let's turn to qualitative approaches. Qualitative approaches often derive theory in an inductive way during the research process. This approach often is focused on concepts that are less easily quantifiable such as social class, perception of products, user experience on a website.

Examples of methods underneath this kind of qualitative approach are observations, interviews, case studies, open-ended survey responses compared to the kind of close-ended responses in a lot of quantitative work.

Qualitative views often incorporate the participants fully. Oftentimes, researchers view individuals less as subjects to be done onto and more of participants in the research. They can be viewed as-- qualitative approaches can be viewed as a collaboration between the subjects or the participants and the researchers.

In the data science realm, qualitative approaches might be more applicable if you're embarking on a project where there's little guidance because there aren't clear best practices yet. Maybe it's unclear how we should think about the space. Maybe it's unclear what questions to ask. And maybe it's unclear how to measure things. In this case, a more open-ended approach might be best.

3.3.2. Introduction to Quantitative Approaches

Like qualitative approaches, some quantitative approaches take an inductive approach. Many quantitative coaches however take a more deductive approach where the study is a test of a broader theory.

Often, we're focused on concepts that are easily quantifiable or more easily quantifiable such as income, the number of products, the number of clicks. Examples of some methods might include structured surveys, close-ended responses, experiments. And we'll often use data from machines such as plane engines.

We'll employ statistical methods in the hope of generalizing from the sample to the population. In the realm of data science, this approach might be most appropriate where there are clear best practices. For example, we could use the quantitative approach to analyze the information we get from airplanes to better improve fuel consumption models.

3.3.3. Does Data Science Blur the Line Between Qualitative and Quantitative?

So really, the qualitative and quantitative approaches are a continuum. And I think data science further emphasizes the fluidity of these two approaches. We can apply traditionally quantitative approaches to what was in the past perceived as in the realm of qualitative. For example, we often employ really advanced methodologies to process text, sound, images. And all these were traditionally viewed as under the qualitative umbrella.

But we could, as I've already mentioned, increasingly apply these quantitative methods to analyze this type of data. In both the qualitative and quantitative approach, data is collected and analyzed. And there's some of the same data concerns in both of these approaches like sampling bias. And guess what? We don't have to just pick one.

We'll talk later about mixed methods. These various approaches can inform different parts of the research and can speak to different audiences. Qualitative approaches can also inform quantitative because it can help with measurement.

And it's important to remember that qualitative approaches aren't just there to add color to an otherwise quantitative study. Qualitative approaches can add color, but they could also help us better understand what's going on. We can do careful qualitative work about emotions and confidence in government leadership that can help us develop ideas that are more generalizable.

There's space for ethnography not only in the context of a postmortem analysis when the project goes poorly or when we get an undesirable outcome. But if we immerse ourselves early in the process, we may be able to produce better quality work. For example, we could help make sure that the metrics we capture are close to the on-the-ground context.

Maybe we do a little bit of both qualitative and quantitative and see if the answers converge. We could do a quantitative approach to identify patterns and then follow up with a qualitative. Some of the best projects bring together both qualitative and quantitative approaches.

3.4.1. The History of Artificial Intelligence and Machine Learning: Intro

Hi. I'm Mike Rivera. I'm the course coordinator for [INAUDIBLE]. In this segment, Steve and I will provide a short thematic history of the development and deployment of artificial intelligence

and machine learning over the past 50 years. But before Steve covers some of the specifics of the history, I want to first highlight three big themes.

Theme number 1. We've experienced pendulum swings between brute force computing and smart algorithms. The brute force approach leverages Moore's law to rely on improvements in computing capabilities and lower costs to complete computationally demanding tasks, while the smart algorithms approach leverages human creativity and statistical tricks to reduce the need for significant computer resources.

Theme number 2. There's a dynamic tension between human decision making and machine intelligence. What's the right division of labor between decisions we allow machines to make and decisions we rely on humans to make? And importantly, our perception on how much to rely on machines versus humans changes over time.

Theme number 3. There's an ongoing tension between efficiency and ethics. Since we can derive insight faster and faster, it may be easier and easier to overlook ethical questions. Another tension is between scientific fascination and ethics. Perhaps we get really excited about what we can do with technology, but we may forget to stop and think about what we should do.

Now, that I've introduced some of the core themes in the history of artificial intelligence and machine learning, Steve will provide us with a more detailed account.

3.4.3. The History of Artificial Intelligence and Machine Learning

OK, thanks for that introduction, Mike. In fact, we are going to be really short-- probably the shortest history of machine learning that anyone has ever created, but still, it's useful to think about what's happened over the course of the last, say, 60 years-- or really, 70 years, actually-- in roughly four periods. And the periods aren't entirely sequential, they're sort of more like mindsets and kind of periodicity that are the way in which people thought about the kind of problem that we're trying to solve and how they were going to solve it.

So let's jump in and start with period-- let's call it artificial intelligence 1.0. And actually, the word artificial intelligence was used at that time. I think of 1.0 as a period of sort of thinking about logic, symbolic reasoning, and search to a possibility space. And so everybody's first example-- everyone's sort of first exposure to the term, of course, comes with the Turing test. Alan Turing, and interestingly, people now forget that the Turing test was actually articulated all the way back in 1950.

And what Turing was trying to do was to create a sort of an aspirational goal for people working this field, and the goal was kind of a pragmatic definition of what intelligence actually would be. Of course, that isn't the way we think about intelligence today, but it was, of course, the test could a person distinguish between speaking with just another person or speaking with a computer in natural language terms?

And interestingly, even though many of the systems we work with today have sort of already exceeded the Turing threshold, the Turing test still has a kind of emotional resonance for lots of folks. And when they think about artificial intelligence, they think about does it actually pass the Turing test? So that was all the way back to 1950.

In 1951, so just really the next year right after, there was the first kind of brute force checkers program. You think about the game of checkers, of course, it's a program that is actually-- or a game that is actually kind of amenable to brute force computing. You can make an exhaustive search of the solution space and actually look at every single possible move in a checkers game. And that worked reasonably well for checkers.

The challenge, of course, and the roadblock is that as you move to different kinds of games or different kinds of problems, the space of solutions tends to grow exponentially. So checkers to chess, and then ultimately to, of course, to the game of Go, which is not possible to do an exhaustive solution space search.

But that was sort of a demonstration in 1951, that if the problem was bounded enough, you could build a program that could look at every single possible move, evaluate them, and then choose the best one. The other interesting thing that happened in 1951 was the creation of a system called SNARC and I never can remember actually what that stands for, so I'm going to look at my notes. It's Stochastic Neural Analog Reinforcement Calculator.

Of course, that was built by Marvin Minsky. And if you have a chance, it's worth going to look at a video of what this machine actually looked like. It was a physical build of a simple one-layer neural network. And yeah, it's called analog because it had motors and chains and like things we used to call potentiometers to connect what we're essentially synapses. And basically, all Minsky was trying to do here was to demonstrate that you could build the machinery and adjust the weights or the parameters, but it didn't really actually try to do anything with it. And it's kind of a cool device. Like I said, physical chains and motors and all that kind of stuff in 1951.

And this actually sort of set the terms for what people were trying to build in the 1950s. The second sort of like kind of iconic device was something called the perceptron, which was built in 1958 by Frank Rosenblatt. And if you saw the perceptron today, you'd probably think of it as supervised learning for a binary classifier.

Still, a one-layer neural network that was able to learn and actually do something-- you could learn linearly separable patterns. But of course, as a single layer, it didn't converge-- if the boundary was not linear, it would just fail. We didn't have the language at that time in 1958 of things like cost functions and J and convergence and all that kind of stuff.

So the way they talked about it was if it was a linearly separable pattern, the machine could do it, and if it was more than that, it would just fail. So there you are in 1958 with the sort of first one-layer neural network.

In '69, which is actually almost-- like little more than a decade later, Minsky again, and Seymour Papert published a kind of well-known and well-- widely-read criticism of neural networks that was interestingly-- the core argument was that a neural network cannot learn an exclusive or function if it's a single layer. And actually, that's true. It was widely quoted and misunderstood as saying that neural networks in general couldn't do that.

And so it was one of these cases where we created in a sense an AI winter, as the term went, by overhyping expectations and actually using science that wasn't quite right to describe a problem which wasn't as general as Minsky and Papert had been interpreted to say. So it was kind of like AI fake news 1969 version.

So that's kind of a summary of what I think of as the 1.0 period. I think that one other important element is to recognize that kind of this brute search approach doesn't actually entirely go away. Many of us may have encountered the term artificial intelligence first with the kind of 1997 Deep Blue IBM computer that kind of beat Kasparov and gained a lot of attention for doing so. It was a bespoke hardware that could search like 200 million moves per second, which is roughly six to eight moves down the chessboard.

But again, it was exhaustive search, and partly a supervised learning system and partly an expert rules-based system for then evaluating which of the moves were best, and I'll come back and say a little bit more about that as we move on to period 2.

OK, so let's move on and talk about what I roughly think of as artificial intelligence period 2.0. And 2.0 I think of as being characterized by what I call rules and expert systems. Basically, this was an attempt to create rule-based decision support. The idea was, can you extract from the minds of experts a set of rules that those experts implicitly or tacitly know that help you to sort of understand what the best-performing people actually do when they make decisions and kind of write down those rules in code so that they can be reproduced by a machine and kind of scale to everybody else who isn't quite as good as the experts?

Two examples of how this played out that are really interesting. The first one is the one that I encountered when I first sort of got interested in this world. It was a program called MYCIN-- M-Y-C-I-N-- built originally by Ted Shortliffe who was a computer scientist at Stanford. I believe the program was first launched in 1973. I encountered it in the mid-1980s.

And what MYCIN was an expert system for decision support in prescribing antibiotics in the hospital, which is actually a pretty well-bounded decision problem. And so what Shortliffe did was he went and interviewed the experts-- the people who were really known to be the very, very best diagnosticians and the very, very best prescribers who always seem to get it right

when they were dealing with sick patients, but couldn't exactly tell you why they did what they did. And he tried to extract that knowledge from those experts and code it into what ended up as 100 rules.

And the system, interestingly, when I encountered it in the mid-'80s, it worked really well. And relevant to some of the debates that we have today about explainability, with 100 rules, it was actually able to explain its decisions. It would tell you, here's why I'm recommending this particular antibiotic and not this other one.

But in practice, it was kind of a dead end. It was a dead end scientifically because it actually only worked for this very well-bounded problem in which there were a very actually kind of limited set of choices. There were maybe 12 or 15 antibiotics that were commonly used and you just had to choose between those 12 or 15.

But more interestingly, at least for me at the time, it felt like an organizational dead end, because even though the system actually worked pretty well, nobody wanted to use it. And they didn't want to use it because in a weird way, they didn't trust it, or at least that's what people said. But they also didn't want to use it because in some sense, it did legitimated what people felt of as a kind of ethereal expertise. They had worked very, very hard and gone to school for many years and seen many, many patients and gone through a many grand rounds to try to kind of accumulate for themselves, and this system made it look like it was just an easy decision problem.

So again, I think that experience-- it's a long time ago now-- could it foreshadow some of the debates that people have around the division of labor question that Mike mentioned earlier between humans and machines? There's another really interesting example of this period 2.0, which actually was done in 1980, it was called XCON. It was done for the sales organization at DEC Computer-- some of you may remember DEC Computer. Ken Olsen, the CEO, is famous for having said in 1979, there is no reason anybody would ever want a computer in their house. Boy was he wrong.

But what DEC tried to do was to essentially take its very, very best salespeople, again, and systematized the rules that they were implicitly using to select particular computer components for their customers' needs. XCON, unlike MYCIN, had actually 2,500 rules. So it's a much more complicated system. And it was said to save DEC a considerable amount of money, like $40 million a year. Ironically, it didn't actually save the company, which went bankrupt just a few years later, maybe because Ken Olsen thought people didn't need computers.

But it never really actually, again, created a huge amount of scientific or organizational excitement. I think both of these systems sort of felt like a direct attack on expertise, and at the time, people didn't really know how to manage that. So that sort of system 2 or period 2.0, rules and expert systems.

So let's jump forward to what I think of as artificial intelligence 3.0, roughly sort of where we are today in most of the work that we do and most of the things that actually you'll learn over the course in the MIDS program and I think sort of the core element of this period of thinking about artificial intelligence machine learning is really all about prediction and cost functions.

And I often think of like four elements coming together to make this possible-- and again, over the course of the courses in MIDS, I mean, all of these four elements will become really clear and you'll become practiced at working with them, but I'll just name them now so that they're in our heads.

First, the kind of uniform understanding of prediction as optimizing a cost function. What, it's called J. How far does the prediction deviate from the real value of the dependent variable, and the whole sort of effort being minimizing that cost function at scale with a very large number of parameters or theta if you kind of prefer that terminology. So that was the-- uniform understanding of prediction as a cost function.

Second big theme was-- I'll quote the intuition of gradient descent as a efficient way of minimizing the cost function in a computing environment. Now of course, there are lots of other ways to do that, but gradient descent was the first way that people really thought about how one could take computing power and use it to much more efficiently minimize that cost function.

The third element was a kind of a deeper understanding of the problems associated with over-fitting in very large data sets, and of course, that will come back in spades later on in the course, but I just wanted to name it now. And then fourth, of course, probably the most important part, which is obvious to everybody now, but if you think back 20 years, it wasn't so obvious, this massive increase in data and computing power. It's a combination of Moore's law, which Mike mentioned earlier-- really, the outperformance of most computing systems relative to Moore's law.

And I guess somebody's law-- I don't know if there is a name for it, but let's call it somebody's law for the massive explosion of data coming off very cheap sensors along with ubiquitous wireless that connects those sensors. And then, of course, as a result of that or in connection with that, machine learning optimized computing languages and libraries which everybody now uses to kind of short circuit-- you never have to build anything from scratch.

And so I think this kind of period of AI 3.0 is really characterized by a kind of shared deep recognition of something that we now all take for granted, that machine learning systems don't have to-- and actually probably shouldn't try to, quote, "solve problems the way humans do." That's-- it's kind of an irrelevant comparison. There's no real need to argue about is it intelligent or does it constitute intelligence or should we use the artificial intelligence as a word or anything like that? That's somebody else's concern, usually marketing and advertising departments. What matters is performance on a task.

And again, that may seem obvious to all of us today, but in the 1980s, the 1970s, even back to the 1950s, it kind of wasn't so obvious. And so we took a few turns and saw arguably some wrong turns into dead ends with these debates, which may have actually been best left to philosophers.

And then finally, I think it's worth mentioning a kind of AI 4.0, which isn't really kind of operational for most machine learning students, machine learning researchers right now, but I just want to mention it, because maybe in a few years, it will become something we talk about quite a lot. I would call it the debate around artificial generalized intelligence or AGI. Some people think about it as adaptive intelligence or the ability of systems to learn how to learn.

The irony of kind of this period or this kind of 4.0 artificial intelligence is that it kind of brings back in a weird way the comparison to human intelligence, combined with an element of I'll call it fear and loathing. And we'll talk about this later on in the course a little bit. Probably the iconic publication which kind of alerted people to this was Nick Bostrom's book, Superintelligence.

And probably some of you have encountered this argument about the possibility of AI systems exceeding the ability of human intelligence in generalized terms, and then sort of ratcheting up and getting smarter and smarter and smarter, and ultimately, in some people's view, taking over the world or causing disaster as a result of the general lack of goal alignment between those systems and what humans want.

Right now this debate-- Andrew Ng has said it's like arguing about overpopulation on Mars. At the same point, Elon Musk has said that he's very concerned that this is actually a short-term issue that human beings are worried about. I just wanted to name it now, we'll probably come back to it a little bit over the course of the semester, but it's actually not an operational part of most people's research agenda, and it's not going to serve a really important role in our conversations about how this history actually manifests in our work today.

3.4.6. The History of Artificial Intelligence and Machine Learning: Outro

Thanks, Steve. Now that we've laid out the history, let's revisit the three big themes I outlined at the beginning-- brute force versus smart algorithms, the tension between human decision-making and machine intelligence, and efficiency, a fascination with fancy tools, and ethics. I'll talk briefly about whether or not any of these themes apply today. And perhaps this is an opportunity to cue up discussion for a live session.

First, the pendulum swing between brute force and smart algorithms still persists. The continued desire to reduce data requirements, for example, through simulations will likely remain a concern. Think about the AlphaZero project. Even though computing resources are getting cheaper, there's still motivation to make algorithms more efficient.

Second, there still remains a tension between human decision-making and machine intelligence. In particular, there's broad concern about the labor market effects of automation and robotics. For example, how might machines influence business process automation?

There's also concern about the effect on skilled but routine operations like pathology. Importantly, there's no classifier that neatly predicts the evolving division of labor between humans and machines. And there probably won't be. Simply put, tech or AI can't solve its own problem. This is one of those wicked problems we discuss throughout the async.

Third and finally, there's a recent interest around new ethical themes like explainability, the meaning of human intelligence, and human dignity. But really, these ethical concerns aren't new, but there's a sense of urgency around these concerns, not only in advanced machine learning settings, but in diverse settings around the world.

3.6 Law, Ethics, and Social Obligations

So Annie and Steve asked me to come today to give you a little bit of a quick overview of some of the legal policy and ethical issues around data science. And I'm going to bring us back in time a little bit because it gives us a sense of the evolution of how policy and legal and ethical issues have become something that you probably, someone who doesn't think of yourself as a big data custodian, might be facing, and where those issues began.

Back in the late 60s, 1970s period where we had big databases that were generally under the control of large corporations or being used by the government, and people were beginning to get concerned that there were decisions being made about individuals that would influence the course of their life, and that might make decisions about whether or not they could have a job, or what sort of benefits they had access to. Corporations were using data to make decisions about access to things like credit, or banking services.

And out of that set of concerns, there emerged a range of safeguards that we associate with data protection. And data protection is a European word for what we call in the US privacy protection. And around the 1970s, there were activities in Europe and in the US that came to be a set of data protection statutes, some of which are in the US, a much more omnibus framework for privacy protection called data protection in Europe, that govern the collection and the whole life cycle of data. How you collect it, how you use it, and how you, importantly, reuse it.

Those frameworks are all evolved at a time when very few people held data. That it was an exception. That you had to actually really make a great effort in order to collect information. That when we flashed our driver's license to purchase liquor or we filled out a paper based form, if somebody wanted to keep that information and make it useful, get utility out of it later in the day, they had to undertake some effort. There was no swipe. There were no ubiquitous systems that were collecting data just as we walked down the street.

So fast forward to the mid 90s where all of the sudden we're looking at the commercial internet. And data protection emerges and privacy emerges, again, on both sides of the Atlantic as not just an important issue for government or for large corporations, but as an increasingly important issue as part of the bargain in a fair commercial marketplace. That now individuals are interacting online, they're leaving little clicks that relate to all of their activities that are collectible, and being managed, and archived, and stored, and reused by a whole host of entities-- backbone providers, ISPs. And every single commercial merchant, or newspaper, or nonprofit, or educational institution that's becoming part of this networked environment.

All of the sudden privacy isn't just a boutique issue. That those concerns that were historically part of large corporate concerns, large government concerns, are now the concerns of anybody who's handling information in this increasingly connected world.

So we fast forward to today, and now we not only have the internet, we have all of us carrying around mobile devices, we have connected appliances, we have cars that have more sensors than a very complex medical device did in the lab many years ago. And we have data that not only people are providing, we have data that's being collected in some ambient way. As we walk down the street, our devices are speaking to each other, our clients are speaking to each other. And we, the corporations, the government, and now increasingly individuals, have these collections of data. And so those privacy issues are pushing out further and further and affecting the decisions and questions that we make on a daily basis about how we're going to interact not just with commercial enterprises and the government, but with each other and in social context.

So we see a rise in privacy concerns and a complexity of privacy concerns as it's no longer about the rules for these large enterprises, but it's coming up in questions about, as an employer when you're making hiring decisions, what sort of information should you access? Sure, you want to mitigate your risk. You don't want to hire an employee that is going to cause you trouble down the line. On the other hand, is everything that people post on their Facebook page or the data that might be available on their device that they just brought into the workplace something that you ought to be gaining access to and being used to manage your risk in an employment context?

As individuals we're using social networking technology. We see our children, and our spouses, and our colleagues, and our friends sharing all sorts of information, both with these commercial providers that are providing the platforms on which individuals are sharing information, and that information is no longer little discrete bits. It's a pretty good picture of somebody's life when you begin to look at it, and it may, in fact, give those companies that are sitting on all their data the capacity to know things about individuals that the individuals don't know yet themselves.

With the classic recent example of an advertisement coming to a young girl whose parents didn't yet know that she had become pregnant. That the corporation was aware of that before the family was and that raises some really interesting privacy issues. Who knows whether or not the girl knew at the time.

But you can imagine an environment, and we see that today, where there may be a third party who's in a better position to know something about our future because they have better access to a richer set of data than we do ourselves. And that might sound kind of trivial in a commercial setting, but imagine take it into a health care setting.

Many of us are now using little Fitbit bands and other sorts of technology to monitor our health. We have electronic medical records, hospitals are sharing information, there is a whole slew of activities going around to try to use all that data that we pour into the health care system on a daily basis to manage both our personal health and our collective health and risk of disease in a very proactive way.

Well what does that mean? It means that we're benefiting in enormous ways. We're identifying new opportunities to use drugs in different ways. We're able to identify at earlier stages people's risk and propensity for disease. It also means, if you're a researcher, you're now dealing with a question about, when does your obligation kick in to notify somebody whose data you have who might never have been a patient of yours about the fact that they actually are a carrier of a marker for a disease that's just been identified, or some other sort of condition that you are in a position to know about and they aren't? And how do you do that? And what's your duty to warn? And what are your ethical obligations as a steward of that data, particularly where you might not have a direct connection to that particular individual? They were never your patient.

So with this great sea of data, we have both a growing complexity of privacy issues, and we also have increasingly the predictive power that's beginning to raise real questions about the duty of data stewards to the people whose lives they have information about and may both influence, but might also kind of assist in a fiduciary way. So that's a really interesting area where we're seeing a new set of issues that are arising because of the way in which we've begun to make use of data in a very powerful, analytic way.

Some of the other issues that have arisen in the use of data over the years that are increasingly important, and I think will be increasingly challenging, are issues of intellectual property. At what point do some of these data sets become things to which intellectual property interest might attach? We think about creativity, but some of the creativity that affords something copyrightable status can attach due to the selection and arrangement of information.

It's not just the author writing an original work. It can be the selection and arrangement of information. That curation activity can also give rise to some copyrightable interests. And so in this environment, we're going to be seeing increasing areas where there may be tensions about copyright, around selection and arrangement of information, and the presentation of information.

And on the other hand, there is an enormous push towards openness. Open data sets, open access to information, and technically the ability to scrape information. At what instance might those copyright protections not really matter so much because somebody is scraping all the

information out of a database that accidentally became accessible over the web, or is accessible through the standard interface despite the fact that there might be some legal protection for it?

So how, as a researcher a data scientist, are you thinking about intellectual property issues, property interests, and what's fair and not fair. Sometimes some of those uses that might appear fair first might actually have some contractual or potentially even some criminal implications, depending upon how you're accessing information.

And then the other issue that's becoming increasingly important in recent years is this effort of prediction related to data mining and the use of information to make decisions about people. We often think of the biases that individuals bring to their decision making. And at some level, people think, well, if the machine made the decision, well then it's got to be right because the information that went into the machine we assume is accurate. And of course the algorithm, well, it's an algorithm. It can't possibly have any biases.

But of course, algorithms are created for particular purposes. And lo and behold, they're created by men and women, and their interests, and biases, and their blind spots all end up being reflected in the data they've selected to analyze, as well as the algorithms that they're going to bring to bear on that data. And as individuals' lives are being impacted by that information, whether it's in the context of commercial interactions, or health care interactions, or the educational environment, or whether or not you can get on a plane, people are beginning to say, wait a second. We need to scrutinize those data sets. What information is it that led you to decide that my kid can't fly? What algorithm is it that told you that my son shouldn't get on a plane?

And as we begin to think about the due process considerations. If this data is as powerful as we hope it's going to be, and it's going to be brought to bear in very useful ways in our lives, we also have to have some concern about the fairness and the ability of individuals to think about whether or not that data is accurate. Is it fair? Is it broad enough? Was it the right data to consider? And the algorithms, similarly.

And this, of course, runs head on into another set of very important legal issues dealing with trade secrecy, sometimes intellectual property or copyright, around access to that sort of information. And then surrounding all of these, we have concerns about security.

When you're holding lots of information on lots of people or lots of information that might be strategically important trade secrets for different sorts of companies, the security of that information becomes exceedingly important. And if there's one thing we know about networks, particularly networks of networks, which is what the internet is, is that there is no such thing as perfect security.

What we have is managed security and lots of managed risk. And as somebody who's working with large sets of data, thinking about your fiduciary obligations to the individuals, or the companies, or to whoever it is that that data relates, and what your obligations are around security, and how you're going to deal with breaches, because they will happen, is a really, really important part of your ethical, legal, and social obligation.

3.7.1. Domain-Specific Policy

Many of you work in domains that have specific policies that govern what you can and cannot do with data. Some of those domains include finance, health, and education. Now, we don't expect you all to be lawyers, but there is an expectation to be responsible stewards of data.

Let's go through a scenario. Let's imagine you come up with an idea that seems great at first glance. But then part of the way through, someone tells you, eh, you can't really do that because a certain law prohibits us from using data in that way.

Especially in situations where there are significant time constraints, which is often, it's important to know at least the basics of the policies that govern the space we work in in order to minimize any setbacks.

How do you do this? Well, one way is to immerse yourself in the domain and begin to learn the inner workings of that space. This will come with time. But a more streamlined recommendation is to ask someone who has been working in this space longer than you.

Is there any policy that outlines the questions we ask, the data we use, and the analysis we conduct? That's the question you can easily ask a trusted colleague. Depending on the magnitude of the decision, you could also ask the legal team to get involved.

3.7.2. Different Levels of Policy

When we think about policy, we may automatically think about it in terms of federal laws or national laws. In a US context, this might be laws that the Congress passes. But it's more than that. If we take the definition from one of the foremost policy scholars Thomas Dye, public policy is, quote, "anything a government chooses to do or not to do," end quote.

This definition is pretty encompassing, so what are we talking about here? Examples at the federal level or at the national level include laws from the national legislator, executive orders, and administrative policy.

When Thomas Dye says that policy is also anything the government chooses not to do, this means that the lives of individuals and what organizations can or cannot do is also influenced by what the government chooses not to legislate in certain areas. The government can also choose

not to enforce certain policies. So this is what Thomas Dye means when public policy is things that government choose not to do.

The national government is important, but it's really important to remember that many countries operate in a federal system where states and municipalities exercise broad authority. So beyond policy at the national level, keep in mind any state, county, or local policies that might regulate what you and your organization may do.

And I want to cover two more types of policies before I close. One is international policy. Think about international regulations on data protection and privacy. Even if your company does not physically reside in the country in which these policies are implemented, if your company does business in those areas, they might be subject to the regulation.

Also, think about internal organization policy. This is distinct from the Thomas Dye definition of public policy. This is not public policy. This is private policy. Simply put, in addition to whatever international, national, state, and local policy that regulate what we can do in our organization, there are likely internal policies that govern our behavior.

And the idea here is not to intimidate you about a potential legal minefield. But rather, the more we are informed about the policies that govern what we do, the more efficient will be. We can avoid holdups if we know beforehand what is permitted and what will be a challenge to execute.

3.7.3. Policies Struggle to Keep up With Changing Technology

One of the many one-liners that I hope you remember from this class is that policy almost always lags behind technology. Think about the newest technology that has come out on the market. It's often the wild west until government policy can catch up to provide guidance on or regulate the use of those technologies. I think this matters for two reasons.

One, we may operate in a relatively unknown area. And we need to be open to the fact that governments may regulate what we do after we do it. This is just a reality. To be a first mover in a space has its benefits and its challenges. Number two, from a ethical perspective, if we only focus on what's legally permitted, we still might find ourselves in questionably-ethical waters. Legal does not always mean ethical.

3.9. Strengths, Weaknesses, Opportunities, and Risks

It's likely that you'll be asked throughout your career to advise on a particular topic. And it's likely that the situation won't be as clear cut as you would like it to be. Remember, overall, our goal is to help decision makers make an informed decision.

It's our task to lay out the options and articulate the strengths, weaknesses, opportunities, and risks in a data-driven way. And when we think about risk, it's important to think about it as a continuum. It's not a dichotomous assessment.

Examples of risk include legal issues, public perception, ethical dilemmas, internal organizational challenges, and decision risk. These risks might be around how you acquire the data or how you use the results. There may be a concern that you'll be unable to get the data.

Perhaps the cost of the project is unclear, or it's unclear whether or not you can deliver the desired insight in the timeline provided. There may be political or organizational pushback or concern that the idea is not in line with your organization's vision.

You might experience decision risk, such as recency bias or confirmation bias. We'll go into more detail elsewhere, but recency bias is when we unintentionally weigh information more heavily-- we unintentionally weigh recent information more heavily than historical information. Confirmation bias is when we unintentionally focus on information that conforms to our pre-existing beliefs, our priors, and dismiss information that is not in line with those ideas.

We think it's useful to think of the risk framework rather than a threat framework. Because unlike threats, we can't neutralize risks. There's some things you can prepare for. And you can mitigate them. And you may not be able to get rid of risks. And the overall goal is to be able to assess the various strengths, weaknesses, opportunities, and risks of a project and then provide a holistic recommendation.

3.10.1. The Responsibility to Understand Context Around an Ask

It's our responsibility to understand the context around a ask. And we could do this from both an ethical lens to help ensure that what is asked of us is acceptable. And we could also view our responsibility to understand the ask so we can better help our colleague or client.

Imagine your manager asks you to find data that shows x. They plan to make a presentation in a couple of weeks. And they're looking for data that supports a specific point they want to make. This approach is in contrast to, hey, go look at this and tell me what you find.

So what can you do if you're asked to take an approach other than one that is systematic and objective? If you anticipate that the data will show a different picture than what your manager is asking, what can you do?

So one approach is to ask for 10 minutes of their time to look at the numbers with you. Now, before you meet with them to make them part of the discovery process, plan the analysis you want to show them. Then once you meet, run the numbers in front of them. And if the data show something that is opposite of what they anticipated or wanted to see, hopefully they're more willing to accept this because they were a part of the discovery process.

Another plausible outcome is that you show them the numbers, and they don't like them. And they push back on the model or approach. They disagree on what you presented. And they say go run it again.

In this scenario, you could go run it again and present various modeling strategies and can try to present the most holistic perspective of the analysis. It may be easy for your decision maker to dismiss one model. But hopefully, it's more difficult for them to dismiss various models that all point in the same direction.

Let's go through a real world-- another real world example about how important it is to understand the context around the ask. I don't mention the company's name to preserve confidentiality. But I hope that you'll see that this example can apply broadly.

Imagine your manager comes to you and asks you to get data on users from Mexico and Canada. Now, thankfully, you work on a team that has a policy of asking why any time you get a request. And so you ask your manager, hey, why do you want the data? They tell you that they want to reduce the reliance on the US market.

Now, because you asked for this additional context, you were able to reframe the question. Instead of who are our users in Mexico and Canada, the question can be reframed to how can we reduce our reliance on the US market? So remember to ask for the context around the question. A decent dose of skepticism is healthy.

Now, I'll close with a subtle but important recommendation. I recommend you get used to asking what more than ask why. Instead of why do you want this data, you can ask, hey, to better help serve your needs, can you tell me what you plan to do with this data? Or what is the goal behind the inquiry? The reason you want to do this is that if you use why, it could be perceived as challenging someone's decision.

3.10.2. Just Because You Can Do Something, Doesn't Mean You Should

Data science does not have a Hippocratic oath, but maybe we should take a do no harm oath like doctors do. So what are some of the ethical frameworks we could use? You could take an entire class on ethics. And in the MDS program, there is a class on ethics that will provide a much more comprehensive framework than the one I will provide here. But let's give it a go.

Let's start with the idea that with great power comes great responsibility. Some of you may recognize this one liner from Spiderman. And so to continue with that, how do we hone our ethical Spidey sense? So one way is to practice how you would approach various scenarios with questionable ethics. Practice and ask yourself, what would I do in this scenario?

It's useful to practice because we may be challenged by people within our organization. And there may be political, organizational, and other power structures that may influence our

response. Furthermore, these ethical challenges may come from people that we care about and people that we respect. Our colleagues may be friends. And if someone you respect proposes something you disagree with on ethical grounds, you need to practice what you would say to them.

3.10.3. The "Newspaper" and the "Grandma" Test

Here are two additional ways to think about ethics. First, the grandma test. Imagine your grandmother or someone else whose opinion you care about. Imagine they asked you about a project you're working on and asked you to explain the nuances of the project. You talk about the design, what information you're gathering, and what you're doing with the data.

How would you feel about that? Let's assume that they understand what you're describing. Would this make you feel uncomfortable? Would this exercise make you feel uncomfortable that you're telling them exactly what you're doing? If so, think about the ethics of your work.

Second, the newspaper test. Imagine your project made its way to the front page of a major newspaper. Imagine internal communications or an internal report or your research design was written about in a major venue. Would this make you uneasy? Would you mind if the public knew exactly what you were doing? If so, perhaps you may think twice about the ethics of your project.

If you'd like to know more about these two frameworks that I've presented, I encourage you to look up the work of Eugene Bardach. He's a professor emeritus at Berkeley.

Now, before I close, let's talk about how the frameworks I just presented may be more or less applicable under certain circumstances. Imagine a national or global health scare where there's an incentive for governments to track people with the use of location data to help minimize the spread of the virus. This may pass both the newspaper and grandma tests.

During the crisis, there may not be any opposition to the partnership between private companies and government to use location data. Maybe prior to the crisis, society was not OK with the use of data to track residents. But this is an example of how ethics may shift in response to a situation we face. And public preferences may shift as we also develop new technological capabilities.

3.10.4. Belmont Report

The final Belmont Report was released in 1979 in response to the Tuskegee syphilis study conducted in the 1930s by the US government. In this study, black males were not aware that they were part of a research experiment. And those with syphilis were denied treatment so that researchers could observe the progression of the disease. The Belmont Report is one of the foundational works that has influenced human subjects research.

I'll briefly outline the ethical principles contained in the report. They are respect for persons, beneficence, and justice. Respect for persons-- people should willingly participate in the study and should be able to withdraw from the study whenever they want. Subjects should sign an informed consent document that outlines information about the study, which includes potential risks and benefits.

Now, this informed consent gets a little more complicated when we talk about vulnerable populations such as those who have different mental capacities, children, the elderly, and those who are incarcerated or in the military.

You might ask yourself, why are those who are incarcerated or in the military considered vulnerable populations in the same category as children and the elderly? Well, because there are either cultural or institutional rules that govern how individuals respond to people in authority positions. In other words, you may not feel comfortable saying no to the warden or to your commanding officer. And in these cases, extra care should be taken to make sure we're acting in an ethical way.

Now, what does this mean for industry? Who are your subjects? We could think of our customers or our users or our clients as subjects. It's important to ask yourself, did they willingly consent to participate? Did they give you blanket consent when they signed up for the platform? Is that enough?

The second principle we draw from the Belmont Report is beneficence. Do no harm and maximize possible benefits and minimize possible harms. That's the guiding principle here. The Hippocratic oath seems reasonable. Do no harm. But we often deal in terms of probabilities. There is some uncertainty with the potential and magnitude to do harm.

So ask yourself, do the potential benefits justify the potential harms? Is there a point when you may decide not to do a project because the potential benefits do not justify the risk you will expose people to? Ask yourself that. In your organization, you might ask what will we gain by rolling out this new initiative? And what are the risks to our customers? Or similarly, how might this affect our public image?

The third principle from the Belmont Report is justice. Let's start with a quote first from the Belmont Report. "We ought to receive the benefits of research-- excuse me-- who ought to receive the benefits of research and bear its burden?

This is a question of justice in the sense of fairness of distribution or what is deserved. An injustice occurs when some benefit to which a person is entitled to is denied without good reason or when some burden is imposed unduly," end quote.

So how do we treat people equally in this context? We could think of the distribution of benefits and burdens using kind of different formulas. One, to each person an equal share. Two, to each person according to individual need. Three, to each person according to individual effort. Four, to each person according to societal contribution. And five, to each person according to merit.

So think again about the Tuskegee syphilis study. Here there's an injustice because black men are not the only individuals who contract syphilis if the subjects in the study were only rural black men. Furthermore, because researchers did not want to interrupt the study, even when treatment was available, the subjects were not provided with medical care.

So in your work, you might ask yourself, who's bearing the burden of this project? And who bears the benefit? If you say those that are participating will also see the benefits after we make the product better, this may be enough depending on the design of the project. However, you may not expose all users to the same treatment or the same change.

Some users may benefit even though they were not exposed to the treatment. And that might be OK if you're testing out different web layouts. But what if instead you're testing out different customer support experiences? Are you withholding a potentially beneficial new support experience or feature? Does the potential benefit or discovery and innovation outweigh the potential for harm?

The main takeaway is that many of us will operate in a very legalistic framework. The legal team might tell us as long as you aren't breaking the law, then you're OK. But remember, just because something is legal does not mean it is ethical.

Furthermore, as technology improves and as we develop different methods to interact with others, we may be moving further away from people. And as a result, it may be easier to separate ourselves from the impact we have on humans because we're not forced to talk directly to them. They may simply show up as data points in our analysis. I urge you to keep in mind the human behind the data throughout your projects

3.10.5. Institutional Review Board

Institutions such as universities that receive federal funding have institutional review boards. The institutional review boards, also known as IRB, review the ethics of proposed human subjects research. Before one begins the researchers-- excuse me.

Before one begins the research, researchers must receive approval from this expert committee to ensure that the research will be conducted in an ethical manner. This review ensures that the many principles derived from the Belmont Report such as subject-informed consent and articulation of the potential risks and benefits are part of the study.

Now, there's various critiques of this review process, but here I want to highlight one. On paper, these are ethical review boards. Many researchers however feel that in practice, they are committees to help ensure that the research institute does not get sued.

Many researchers feel the IRB is simply operating by a very legalistic framework. It's a legal checklist. But remember, the one-liner I want us to remember is just because we follow what is legal does not mean we will sidestep any ethical concerns.

Now, many of us work in organizations that have some formal ethical review process. That's awesome. But many of us work in organizations that don't have that. And either way, we make judgments on a day-to-day basis that will never see the formal review process. We make decisions on how to sample, on how to design, on how to model, et cetera, that will never receive formal scrutiny.

We are on the front lines of the ethical debates. Inform yourself and have some guiding ethical principle such as those outlined in the Belmont Report. Because in some cases, if we leave it to others to deal with ethics, they won't.