Chapter 7

The Criticality of Governance and Ethics in Augmented Intelligence

Introduction

When organizations begin to leverage their critical data, there is often a rush to quickly gain insights. Therefore, it is not surprising that data scientists and knowledge experts will focus on gathering and cleansing data and building models. Issues such as governance and ethics often take a back seat. The issues of governance and ethics only take center stage when there are incidents that land on the front page of the business press. Governance cannot be an afterthought when moving to augmented intelligence, since it must combine corporate and government-mandated processes and data tied to organizational requirements. Governance must also focus on the ethics of how a company behaves in terms of protecting its customers. It is the responsibility of both corporate and IT management to monitor activities to ensure that rules are followed. If governance and ethics are ignored, a company's reputation will be badly tarnished and financial consequences can be significant. In this chapter, we discuss the need to apply governance rules and ethical business practices to support your data.

Defining a Control Framework

To establish a consistent and predictable way to govern your data and your models, it is imperative to put a governance control framework in place. Simply put, you must focus on the controls you need to put in place to protect your business and your customers. A control framework creates the basis for formulating policies and procedures for working with data, machine learning (ML) and augmented intelligence. Your augmented intelligence control framework should be enforced throughout the business. Adherence to this control framework has to be designed in at every stage of implementation—from the pilot to full-blown implementation.

When you begin to think about controls over your data, you need to consider a number of different controls that apply to your entire data cycle. Managing the data cycle requires that controls apply across the entire data cycle:

- Controls should govern the way data is acquired, protecting data privacy by ensuring consumer opt in—both for the data items collected and any derived data.
- Controls should govern the breadth of the data collected, to limit the risk of training a machine to perpetuate the bias built into the record of current practice.
- Controls should ensure transparency in the model, sufficient to yield an explanation of the factors that drive the model's recommendation.

For purchased algorithms, a business cannot ensure that any of these principles were followed when the model was developed. In such a case, controls need to be applied before a model's recommendation is put into action. Likewise, when you purchase third-party data, you must ensure that proper controls were in place during the collection of that data.

To create your framework, you need to bring together teams from across your business. It is important to include both legal and compliance teams, but you should also include business executives. There are data governance and control rules that you are legally obligated to follow. For example, if you are a credit card company, you cannot include certain types of data in your creditworthiness algorithm. Certain data features about an individual, such as ethnicity, cannot be used. Your compliance and legal teams will outline data features that are barred from use. There are other data features that require your business to decide how they should be factored into an algorithm. Let's look at the issue of gender. Depending on your business, it may be prudent to factor gender into an algorithm. For example, it's well known that auto insurance rates are higher for young men than young women because young men have higher claims rates.

Likewise, nobody is going to argue that a health and beauty company should market products differently to different genders: Men and women obviously use very different grooming products.

But things are not as simple as they may seem at first glance. There are a number of nuances that organizations must consider in order to both market correctly and not run into compliance and ethical issues. For example, the biological differences between males and females would factor into the health and beauty market and necessitate not just different marketing campaigns but different products. However, a marketing plan designed to offer women toothpaste that is packaged in a pink box will likely be ineffective and deliver no real value. Although an insurance company may use statistical data to price insurance based on the fact that men may have more accidents than women, there are men with certain characteristics who actually have fewer accidents than women. If a man who has no accidents is charged more than women who have been involved in a number of accidents, this pricing scheme may be viewed as biased.

Let's revisit the original example of the credit card issuer. We can all agree that race should not be a factor when deciding an individual's creditworthiness. In fact, the Federal Trade Commission (FTC) enforces the United State's Equal Credit Opportunity Act (ECOA). The ECOA prohibits credit discrimination on the basis of race, color, religion, national origin, sex, marital status, age, or because you get public assistance. However, these factors might slip into an ML creditworthiness application through a back door. Let's look at age. If a prospective customer's high school or college graduation date is used as part of a company's creditworthiness algorithm, the system may begin to disfavor new graduates who also tend to be younger applicants. Similarly, if you begin to drill down into the credit applicants' four-digit zip code and include demographic data about that specific area, issues like race and national origin may creep into your determination on creditworthiness. These types of problems are why it is so important to have a cross-functional team create an AI control framework.

Technologists must be part of the team in order to fully explain the types of data that is being analyzed and why it is being included in the model. In many cases, machine learning algorithms are black boxes, and it is nearly impossible to understand exactly how a feature pertaining to a protected class is being factored into the model. One step that companies are implementing is to incorporate a testing procedure in which you check the output of the model to determine if the results favor a particular race, gender, or other category of people. This sort of testing is one of the steps that can be built into your governance framework before any model is put into production. Business and compliance teams can then question the need to include certain data factors and the possible pitfalls and concerns of including that information. Furthermore, the business teams and business executives will need to determine if they are comfortable

using certain data features in an algorithm. The business team needs to decide if including certain data features in its calculations goes against the corporate ethos, or corporate ethics. Even if a company puts profits above any sort of ethics, the business team must still consider the possible risks of including certain features in a model. For example, a restaurant loyalty program is not legally barred from including factors such as race, marital status, and gender into their rewards program. An analytics team might determine that a model can be made slightly more accurate by including such features about customers. However, if the public finds out that those factors are impacting their rewards points, there could be outrage. It is easy to imagine the headlines, protests, and boycotts that would surely follow the discovery.

YouTube's Engagement Model Can Spawn Ethical and Governance Problems

Like any ad-supported Internet platform, Google's YouTube has a goal of keeping users engaged and on its page for as long as possible. The longer a user is on YouTube, the more ads they can be shown. To further this goal, as somebody watches a video, YouTube suggests additional videos that the customer is likely to click on. As it turns out, YouTube viewers have a very high propensity to click on conspiracy-theory videos. For example, if someone is researching the moon landing, the YouTube algorithm would determine that the individual is likely to click on a video claiming that the moon landing was faked. This can lead to a spiral of being served more and more content claiming that the US government or other agencies faked all of the Apollo missions. Although the user may have originally just wanted to spend 10 minutes seeing footage from the moon, they may have spent two hours "learning" about how the moon landing was faked.

Of course, for YouTube's business, this is a great win: They were able to keep the user engaged on their platform for a prolonged amount of time (user engagement is one of the key factors that investors like to evaluate). On the other hand, YouTube has begun to recognize the problem of spreading disinformation. The company is changing its algorithms to help stop the spread of non–fact-based conspiracies.[1] In addition to altering its algorithms, YouTube is also using human reviewers to help better train the models that suggest videos and censor content. Although this decision might hurt the company's time of engagement, the decision was made at high levels because executives determined that the pushing of conspiracies conflicts with the company's ethics, or that the negative headlines were bad for business.

Creating Your Augmented Intelligence Control Framework

A good way to start off your framework is to think about the types of data you will include in your models and the possible use cases. This first step begins to create your *control environment*. The control environment provides the foundational internal rules and guidelines on how different types of data can be used. There is almost a limitless number of use cases and types of data, but why don't we look at how this assessment might look for a travel company. Figure 7-1 depicts a chart in which types of data are on the vertical axis and use cases are on the horizontal axis. Guidelines on how data should be used in an AI project are in the corresponding boxes. Of course, this is just a simple example, and

	Initial Marketing	Pricing	Upselling
Customer gender	Create campaigns targeting different activities.	Do not factor.	Promote different excursions or activities.
Does the customer have children?	Campaigns focused on family-friendly activities.	Perhaps you want to create options for less expensive travel for children.	Promote different excursions or activities.
Customer income	Campaigns based on the type of travel that the customer would likely book (i.e., budget vs. luxury travel).	Do not factor.	Offer unique, custom options for high-net customers, whereas more budget-focused travelers can be offered smaller upgrades for a low cost.
Weather where customer is located	Create a custom campaign when certain weather events occur.	Sales based on current weather where customers are located.	Opportunities to reach out to customers who have already booked but might be impacted by current weather. Example: Customer is experiencing a blizzard but has an upcoming trip to the Caribbean—offer them a half-day boat excursion to a deserted beach.
Customer loyalty to your brand	Treat your loyal customers wonderfully, and promote the perks of being loyal.	Loyal customers may receive discounts.	Maybe your most loyal customers get certain perks and discounts on special opportunities.

Figure 7-1 Factors to Consider in Governing Different Types of Data

your guidelines will be more complex as you bring together the opinions of data scientists and business leadership, along with the legal and compliance teams.

You can quickly see how this type of chart would be very different for a company that must comply with the Equal Credit Opportunity Act or other similar regulations. In addition, there are other factors that you might want to bar teams from ever considering (i.e., race, sexual orientation, political affiliation, etc.).

Steps in Your AI Control Framework

There are a number of steps required to control your data so that it meets both your governance and ethical requirements. These steps include conducting a risk assessment, creating control activities, and creating a monitoring system.

Conducting a Risk Assessment

The first step to create your control framework is to perform a *risk assessment*. You must determine the potential risks your organization faces as you continue down your AI journey. Even if you are following regulatory and compliance requirements, what are the social implications that you need to consider? Later in this chapter we talk about several use cases in which companies do not breach legal requirements, but their algorithms go beyond the comfort level of most people. You must also think about the level of notice customers and prospects receive regarding how their data will be used. Is the disclosure on how data will be used on page 4 of a terms and conditions agreement that nobody reads? If you sell customer or prospect data, do you have any controls over how that data is used in the future? If you do have policies regarding how customer data can be used by third parties, how do you ensure that those policies are being enforced? If you buy third-party data from other sources, are you sure that the data meets your own internal controls? Was that data lawfully collected? All of these risks must be considered by your cross-functional team.

Creating Control Activities

The next step in the creation of your augmented intelligence control framework is to create *control activities*. These are specific rules, policies, and procedures that are established to make sure that business units are following your agreed-upon data rules. For example, you may want to institute special rules when a group wants to use gender in a machine learning algorithm. The rule might

say that if gender is to be used when creating a model, the data team must get approval from a senior manager in the corresponding business unit. It is then up to that manager to approve, disapprove, or loop in compliance teams. This rule can be codified into the model-building process to create guardrails so that teams stay safe from violating established controls. If a manager receives a request to use gender in a marketing campaign, they will likely quickly approve of the use. On the other hand, if it is a human resources (HR) use case, the manager may need to disapprove of it or bring in legal and compliance colleagues to help determine the appropriateness of using gender.

Creating a Monitoring System

The final step in creating your control framework is to create a *monitoring system*. Many businesses that are experiencing successful implementations of AI have internal monitoring systems to make sure that internal controls are being followed. Monitoring should be overseen by executives who have two important mandates—(1) financial success; and (2) preventing avoidable risks, embarrassment, and legal problems. These executives should be well versed in the company's risk tolerance and the values that the company believes in: A family-friendly brand is going to have a very different risk tolerance than a company solely focused on males in the 18- to 40-year-old demographic. The executives that are overseeing the monitoring process do not need to understand the bits and bytes of your company's AI algorithms and models. Instead, they need to work with data teams who can explain the types of data that are going into a model, how data is being collected, and the impact of including different types of data in a model. If using age in a model only gives a very slight lift to a model's accuracy, the executive might determine that it doesn't make sense, from a risk perspective, to include age in the model.

Data Privacy Controls

If the business needs to acquire additional customer data, it can reach out to customers and obtain their permission to gather additional data about them. The company can also acquire additional attributes (i.e., features) about its customers. Having additional attributes available for model building can enhance the model's predictive power. It is increasingly easy for businesses to enrich their customer and prospect data with various third-party data sources. For example, they can match a customer's information with that customer's social media data.

The business can also shop for data from specialized data brokers who possess data on individuals who are not customers. It has been a common practice

The Ethics of Using Data to Modulate Pricing

Similar to any mobile application, users of Uber grant the company permission to access a lot of data. Buried in the terms and conditions of Uber's app is a condition that gives the company the right to see your phone's battery percentage. In a May 17, 2016, interview on NPR's podcast Hidden Brain, Keith Chen, former head of economic research at Uber and current Associate Professor of Economics at UCLA Anderson was interviewed.[2] During the interview, Chen revealed that Uber users are more likely to pay for surge pricing if the battery on their phone is low. Rather than waiting for surge pricing to go down, or trying to quit the app and restart it to see if the pricing has changed, the user has a high propensity to just accept the high pricing. Clearly, Uber could use this type of data to maximize its revenue and run tests to see how much they can charge before a large amount of low battery customers seek out alternative systems. When asked whether the company takes advantage of this data, Chen made it clear that Uber does not, stating that "We absolutely don't use that to kind of like push you a higher surge price, but it's an interesting kind of psychological fact of human behavior."

This is an example of corporate ethics and risks. Although it would look terrible if there was a news story about Uber gouging users whose batteries were low, few people would argue with a hotel incorporating status at other chains into their pricing model. If a traveler is a platinum member at a certain hotel brand, it makes sense for a competing brand to offer aggressive discounting, upgrades, or automatic platinum-status matching.

in marketing to acquire email lists or to access consumer data on social media sites such as Google, Facebook, or LinkedIn. Businesses now must be aware of enhanced regulatory controls on data privacy when reaching out to consumers via acquired data. Permission must be granted from the consumer to use the data via an opt-in process. But gaining such permission is difficult if there is no preexisting relationship between a business and the consumer.

There are now further challenges with the opt-in process that are becoming apparent. The very definition of "opt in" must be considered as well. Many individuals do not fully appreciate the implications of opting in. Additionally, if you want to use a service, or enter a store's premises, you may have no option but to opt in. There are several well-publicized cases that have raised questions on the adequacy of opt-in mechanisms to provide transparency to consumers on how a business is using their personal data. Here are two examples.

In one case, a predictive model was built for the retailer Target to determine whether a customer was likely to be pregnant. The model based its predictions on an analysis of retail transactions captured in point-of-sale data. Women who became pregnant tended to significantly change their purchase patterns. Armed with this knowledge, the retailer sent coupons for baby and maternity products

to each consumer flagged by the model. The marketing campaign caused a furor when it was reported that a father first learned that his teenage daughter was pregnant after Target's mail offers to his daughter raised his suspicions.[3] This chain of events was not envisioned by the consumer when she gave permission for the retailer to access her point-of-sale data.

A second case involved the collection of data on Facebook users (and their friends) via an app built on the Facebook platform for Cambridge Analytica and then downloaded by Facebook users. The personal and social data collected was used to target voters on behalf of the Trump campaign in the 2016 US presidential election.[4]

In both the Target and the Cambridge Analytica cases, customer opt in did not provide sufficient transparency to the consumers about how their personal data would be used by developers and advertisers:

- Consumers were not informed that additional personal data would be derived via algorithms from the data for which they had provided opt-in approval. In the Target example, this practice was especially problematic, since the derived data was protected health data.
- Consumers were not informed about which types of actions would be taken based on the derived personal data. And they were not informed about who would be acting on the data, whether it was a retailer or a political campaign.

Additionally, consumers were not informed that access to their personal data could be granted because a friend had granted access to his/her personal network. In the Facebook situation, access to a user's data brought access to data on the user's network of friends. This shift from personal data to social data raises additional privacy questions, which are being studied by government officials in Europe and the United States. New government regulations are establishing new standards on data privacy. The European Union (EU) leads this effort. General Data Protection Regulation (GDPR) was passed in 2016, and its enforcement phase began in May 2018. Likewise, in June 2018, the California Consumer Privacy Act, influenced by GDPR, was passed. This state regulation institutes the strongest privacy laws in the United States.

These new data privacy regulations are adding friction to the process of acquiring personal and social data and then acting on this customer data. The new regulations will require businesses to take several new areas into account:

- **Right to consent:** The opt-in process needs to be made more transparent and more specific to consumers. This calls into question today's practice of requesting a general opt in by a consumer for any and all uses of his or her personal data collected by a business. More granular permission

will be required in the future, getting an opt in for specific data and for specific actions that could be taken by the business. The consent request should be clear, and the consumer should have ready access to view all consent requests that have been granted. A growing number of consumers are demanding that businesses and websites show them all the personal data that has been collected.

- **Right to be forgotten and the right to erasure:** The right to be forgotten has a long history in the legal system as protection for people seeking to expunge their past records from public view. With today's widespread use of search engines, this has come to include the deletion of URLs from search results. But GDPR goes further to specifically demand a right to erasure. As a practical matter, businesses as well as Internet social media or search companies will need to respond to a consumer request to delete all personal data.
- **Right to an explanation:** When algorithms are used for making decisions about consumers, a consumer can demand to know the reasons for the decision. They can request to know which factors were considered by the algorithm when the decision was made. This principle is well established in the field of credit scoring, where there are long-standing regulations that require the credit-scoring company to list the key factors that impacted the score. More recently, GDPR speaks of a broader "right to an explanation" for a decision, though its scope and specific application is not yet settled. This right seeks to address issues of alleged algorithmic bias, sometimes referred to as "algorithmic redlining." This right extends to areas such as hiring, the granting of mortgage loans, and the sentencing of defendants. The implication for an organization is that it must be able to provide a level of transparency as to which factors drove the algorithmic recommendation or decision.

On an Organizational Approach to Controls

A variety of new governance and ethical requirements are shining a light on the role of data in today's organization. It's imperative for organizations to develop well-thought-out policies on how data is collected, managed, and used. At the same time, it's imperative for individuals to monitor what permissions they have granted for the use of their personal data and to demand greater transparency on future use. This awareness demands a response from organizations to develop policies that govern data use so that they can maintain the trust of their customers, as well as staying ahead of emerging legal standards.

GDPR Requires You to Rethink Your Control Framework

Large enterprises and multinational businesses have had to think about how GDPR will impact their business. Businesses have always had to walk a fine line between gathering and analyzing massive amounts of customer and prospect data and keeping private data and secure from both regulatory and reputational standpoints. What's different about GDPR is that it will impose substantial financial penalties if a company fails to comply with the regulation: Companies risk fines of up to four percent of their annual global revenue.

GDPR specifies eight rights that apply to citizens regarding the use of their data by external organizations:

1. Right to be informed[5, 6]
2. The right of access[7]
3. The right to rectification[8]
4. The right to erasure[9]
5. The right to restrict processing[10]
6. The right to data portability[11]
7. The right to object[12]
8. Rights in relation to automated decision making and profiling[13]

It is important to keep in mind that GDPR is not the only data protection regulation. The California Consumer Privacy Act that goes into effect January 1, 2020, will put greater restrictions around how companies can collect and use data. Regulations are constantly evolving. As compliance, legal, technical, and security teams work to ready their organizations for these rules, they must keep an eye on the future, anticipating new requirements that may impact the business.

Best Practices for Ensuring Data Privacy and Security

How can you balance the need for access to the right data while maintaining compliance with a changing regulatory and security landscape? Although there isn't one right answer, there are some best practices that can help turn the security officer into a business partner. Here are the top three:

1. **Work Together**
   Privacy, security, and project management offices must work together as a team. Many companies that proactively manage data privacy and security challenges embed privacy and security personnel within business units. Security by design should become a common strategy; this will help organizations build security and privacy provisions into projects from the outset.

2. **Assess Impact**
   Perform privacy and security impact assessments as part of a project's approval process. As a project moves forward, there should be continual checkpoints

to ensure that compliance, security, and protection requirements are met. A project should not move forward with funding until it has been reviewed and the risk levels defined. Continuous assessments allow teams to identify and address issues in early stages of the project.

3. **Identify the Data**
   Identify the data that will be used for a new project. Understanding the sensitivity of data being used will make it easier for companies to meet the requirements of regulations such as the GDPR, and it will reduce the risk of a breach. Give business leaders and executives oversight of data based on the sensitivity and risks associated with the information. These executives should sign off on a project only once they agree that the risks of exposure are worth the benefits.

These best practices should be the foundation of an organization's security and governance policy as it prepares for the GDPR. This foundation will help protect the business from costly fines and will help prevent future security breaches.

Combining both organizational change with technical solutions can help organizations overcome the risks posed by removing data silos, giving employees access to more data and exploring new, data-centric business models. A well-planned strategy can enable an organization to innovate safely and securely.

The bottom line is that businesses should plan for more restrictions on access to their customers' personal data. The restrictions will develop based on a combination of new government regulations along with new customer demands.

Summary

What does it mean to manage the cycle of data in a well-governed and ethical manner? It is clear that you need to understand the nuances of managing your data with the right set of controls in place. Business leaders must understand the types of data insights that the business's data is producing. How are those insights getting operationalized? What risks do you face? Even if you are 100% legally clear, should you use certain insights?

With the proliferation of augmented intelligence, the risks will become more intense. Businesses will be able to predict the future and gain greater insights into what the data is telling them. Therefore, it is likely that there will be more opportunities to violate both the privacy of customer data and to unintentionally break governance rules. Corporate policies pertaining to data privacy and data permissions should be managed by a Chief Data Officer or Chief Compliance Officer. That officer is responsible for establishing corporate policies that meet statutory requirements and are consistent with customer expectations of transparency and fairness in the use of personal data.

The executive office (whether the Chief Data Office, Chief Compliance Office, or Chief Risk Office) cannot work in isolation. The planning and implementation of policies and controls should be done jointly with line managers who have responsibility for each business function. Ethical controls speak to the continuing human responsibility to govern the development and deployment of machine intelligence in order to ensure fair and safe operation. In the future, consumers along with regulators will be looking to businesses to act responsibly and transparently in gaining permission for the use of personal data.