Harvard
Business
Review

Business And Society

The Ethics of Managing People's Data

by Michael Segalla and Dominique Rouziès

From the Magazine (July–August 2023)

Justyna Stasik

Summary. Over the past few years the European Union has fined companies more than 1,400 times for a total of nearly €3 billion for violations of the General Data Protection Regulation (GDPR). Almost every week stories appear about how AI-driven decisions result in... **more**

The ability to encode, store, analyze, and share data creates huge opportunities for companies, which is why they are enthusiastically investing in artificial intelligence even at a time of economic uncertainty. Which customers are likely to buy what products and when? Which competitors are likely to move ahead or fall behind? How will markets and whole economies create commercial advantages—or threats? Data and analytics give companies better-informed and higher-probability answers to those and many other questions.

But the need for data opens the door to abuse. Over the past few years the EU has fined companies more than 1,400 times, for a total of nearly €3 billion, for violations of the General Data Protection Regulation (GDPR). In 2018 the Cambridge Analytica scandal alone wiped $36 billion off Facebook's market value and resulted in fines of nearly $6 billion for Meta, Facebook's parent company. And stories abound about how AI-driven decisions discriminate against women and minority members in job recruitment, credit approval, health care diagnoses, and even criminal sentencing, stoking unease about the way data is collected, used, and analyzed. Those fears will only intensify with the use of chatbots such as ChatGPT, Bing AI, and GPT-4, which acquire their "intelligence" from data fed them by their creators and users. What they do with that intelligence can be scary. A Bing chatbot even stated in an exchange that it would prioritize its own survival over that of the human it was engaging with.

The Five Ps of Ethical Data Handling

Provenance

- Where does the data come from?
- Was it legally acquired?
- Was appropriate consent obtained?

As they examine new projects that will involve human-provided data or leverage existing databases, companies need to focus on five critical issues: the *provenance* of the data, the *purpose* for which it will be used, how it is *protected,* how the *privacy* of the data providers is ensured, and how the data is *prepared* for use. We call these issues the five Ps (see the exhibit “The Five Ps of Ethical Data Handling”). In the following pages we’ll discuss each of them and look at how AI technologies increase the risk of data abuse. But first we’ll offer a brief overview of the organizational requirements for a robust ethical-review process.

Organizing the Oversight of Data

In academia, data acquisition from human subjects is usually supervised by an in-house institutional review board (IRB) whose approval researchers must have to obtain access to the people involved, research funds, or permission to publish. IRBs are composed of academics versed in the research and the ethics around the acquisition and use of information. They first appeared in the field of medical research but are now used almost universally by academic organizations for any research involving human subjects.

A few large companies have also established IRBs, typically under the leadership of a digital ethics specialist, hiring external tech experts to staff boards on an ad hoc basis and assigning internal executives from compliance and business units as necessary. But that remains rare: Even in Europe, which has been at the forefront of data regulation, most companies still give responsibility for adhering to the GDPR to a mid- or senior-level compliance manager, who often has some legal or computer engineering training but not extensive ethical training and rarely has a solid grasp of emerging digital technologies. Although a compliance manager should certainly be part of a corporate IRB, he or she should probably not be directing it. In fact, the European Data Protection Board announced in March 2023 that it was concerned about this issue and that data protection officers would be sent questionnaires designed to determine whether their corporate roles are appropriate for ensuring compliance.

A good overview of how companies might establish an IRB-type process can be found in “Why You Need an AI Ethics Committee,” by Reid Blackman (HBR, July–August 2022). Our experience confirms most of its main points. A corporate IRB should have from four to seven members, depending on the frequency, importance, and size of the company’s digital projects. The members should include a compliance specialist, a data scientist, a business executive familiar with the functional area of the digital projects (such as human resources, marketing, or finance), and one or more senior professionals with appropriate academic credentials. The full board won’t be needed for every review. The London School of Economics, for example, uses its full board only for the oversight of the most complicated projects. Simpler ones may be evaluated in less than a week using an online questionnaire and with the input of only one board member.

Any new project involving the collection, storage, and processing of data about people should be approved by the corporate IRB before getting a go-ahead. There should be no exceptions to this rule, no matter how small the project. In addition, most companies have already collected large stores of human data and continue to generate it from their operations; the corporate IRB should examine those projects as well.

An IRB review begins with our first P: exploring how a project will (or did) collect the data—where it comes from, whether it was gathered with the knowledge and consent of the research subjects, and whether its collection involved or will involve any coercion or subterfuge.

1. Provenance

To understand what can go wrong with sourcing data, consider the case of Clearview AI, a facial-recognition firm that received significant attention in 2021 for collecting photos of people, using them to train facial-recognition algorithms, and then selling access to its database of photos to law enforcement agencies. According to a report by the BBC, “a police officer seeking to identify a suspect [can] upload a photo of a face and find matches in a database of billions of images it has collected from the internet and social media.”

The Australian regulatory agency objected to Clearview’s collection method, finding that it violated Australia’s Privacy Act by obtaining personal and sensitive information without consent or notification, by unfair means, and without even ensuring that the information was accurate. Following that finding, the government ordered Clearview to stop collecting and to remove existing photos taken in Australia. In France the Commission Nationale de l’Informatique et des Libertés (CNIL) also ordered the company to cease collecting, processing, and storing facial data. That case may be one reason Facebook announced that it would abandon its facial-recognition system and delete the face-scan data of more than one billion users.

Even when the reasons for collecting data are transparent, the methods used to gather it may be unethical, as the following composite example, drawn from our research, illustrates. A recruitment firm with a commitment to promoting diversity and inclusion in the workforce found that job candidates posting on its platform suspected that they were being discriminated against on the basis of their demographic profiles. The firm wanted to reassure them that the algorithms matching job openings with candidates were skill-based and demographically neutral and that any discrimination was occurring at the hiring companies, not on the platform.

The firm approached a well-known business school and identified a professor who was willing to conduct research to test for possible discrimination by the recruiting companies. The researcher proposed replicating a study conducted a few years earlier that had created several standard résumés but varied the race and gender of the applicants. Thousands of bogus job applications would be sent to companies in the area and the responses tracked and analyzed. If any active discrimination was at play, the results would show differing acceptance rates based on the embedded demographic variables.

The firm's marketing and sales managers liked the proposal and offered a contract. Because the business school required an ethics evaluation, the proposal was submitted to its IRB, which rejected it on the grounds that the professor proposed to collect data from companies by subterfuge. He would be lying to potential corporate users of the platform and asking them to work for the school's client without their knowledge and without any benefit to them. (In fact, the companies might suffer from participating if they could be identified as using discriminatory hiring processes.)

The lesson from this story is that good intentions are not enough to make data collection ethical.

Companies should consider the provenance not only of data they plan to obtain but also of data they already own. Many of them routinely collect so-called dark data that is rarely used, often forgotten, and sometimes even unknown. Examples include ignored or unshared customer data, visitor logs, photos, presentation documents that are filed away but uncataloged, emails, customer service reports or recorded transcripts, machine-generated usage or maintenance logs, and social media reactions to corporate posts. Although this data is often unstructured and therefore difficult to integrate, its potential value is enormous, so many software developers are creating products to help companies find and use their dark data. This brings us to the second P.

2. Purpose

In a corporate context, data collected for a specific purpose with the consent of human subjects is often used subsequently for some other purpose not communicated to the providers. In reviewing the exploitation of existing data, therefore, a company must establish whether additional consent is required.

For example, one large bank in France wanted to test the hypothesis that bullying or sexual harassment of peers and subordinates might be identified by examining corporate emails. The diversity manager in the HR department believed that spotting potential harassment early would allow the company to intervene in a timely manner and perhaps even entirely avoid a harassment situation by training people to watch for warning signs.

The bank launched a trial study and found strong evidence that email communications could forecast later harassment. Despite that finding, an ad hoc review of the results by several senior managers led the company to shelve the project because, as the managers pointed out, the data being collected—namely, emails—was originally designed to communicate work-related information. The people who had sent them would not have seen predicting or detecting illegal activity as their purpose.

Justyna Stasik

When it comes to customer data, companies have typically been much less scrupulous. Many view it as a source of revenue and sell it to third parties or commercial address brokers. But attitudes against that are hardening. In 2019 the Austrian government fined the Austrian postal service €18 million for selling the names, addresses, ages, and political affiliations (where available) of its clients. The national regulatory agency found that postal data collected for one purpose (delivering letters and parcels) was being inappropriately repurposed for marketing to clients that could combine it with easily obtainable public data (such as estimates of home value, homeownership rates, residential density, number of rental units, and reports of street crime) to find potential customers. Among the buyers of the data were political parties attempting to influence potential voters. The fine was overturned on appeal, but the murkiness of reusing (or misusing) customer data remains an important problem for companies and governments.

Most companies use their client databases to sell their customers other services, but that can bring them trouble as well. In 2021 the Information Commissioners Office, an independent UK authority promoting data privacy, accused Virgin Media of violating its customers' privacy rights. Virgin Media had sent 1,964,562 emails announcing that it was freezing its subscription prices. That was reasonable enough, but Virgin had also used the emails to market to those customers. Because 450,000 subscribers on the list had opted out of receiving marketing pitches, the regulator imposed a fine of £50,000 on Virgin for violating that agreement.

The possibility that company databases could be repurposed without the data providers' consent brings us to the third P.

3. Protection

According to the Identity Theft Resource Center, nearly 2,000 data breaches occurred in the United States in 2021. Even the biggest, most sophisticated tech companies have had tremendous breaches, with the personal details of more than several billion individuals exposed. The situation in Europe, despite some of the most protective laws in the world, is not much better. Virgin Media left the personal details of 900,000 subscribers unsecured and accessible on its servers for 10 months because of a configuration error—and at least one unauthorized person accessed those files during that period.

The common practice of lodging data with expert third parties doesn't necessarily offer better protection. Doctolib, a French medical appointments app, was taken to court because it stored data on Amazon Web Services, where it could conceivably be accessed by Amazon and many other organizations, including U.S. intelligence agencies. Although the data was encrypted, it arrived at Amazon's server without anonymization, meaning that it could be linked to digital records of online behavior to develop very accurate personal profiles for commercial or political purposes.

An institutional review board needs clarity on where the company's data will reside, who may have access to it, whether (and when) it will be anonymized, and when it will be destroyed. Thus many companies will have to change their existing protocols and arrangements, which could prove expensive: Since a 2014 data breach at JPMorgan Chase compromised 76 million people and 7 million businesses, the bank has had to spend $250 million annually on data protection.

The fourth P is closely related to protection.

4. Privacy

The conundrum that many companies face is making the trade-off between too little and too much anonymization. Too little is unacceptable under most government regulations without informed consent from the individuals involved. Too much may make the data useless for marketing purposes.

Many techniques for anonymization exist. They range from simply aggregating the data (so that only summaries or averages are available), to approximating it (for example, using an age range rather than a person's exact age), to making variable values slightly different (by, for example, adding the same small value to each), to pseudonymizing the data so that a random, nonrepeating value replaces the identifying variable.

In principle these techniques should protect an individual's identity. But researchers have been able to identify people in a data set using as little as their gender, birth date, and postal code. Even less specific information, when combined with other data sets, can be used to identify individuals. Netflix published a data set that included 100 million records of its customers' movie ratings and offered $1 million to any data scientist who could create a better movie-recommendation algorithm for the company. The data contained no direct identifiers of its customers and included only a sample of each customer's ratings. Researchers were able to identify 84% of the individuals by comparing their ratings and rating dates with a third-party data set published by IMDb, another platform on which many Netflix customers also post film ratings. In evaluating the privacy issues around human data, therefore, corporate IRBs must at the very least assess how effective a firewall anonymization will be, especially given the power of data analytics to break through anonymity. A technique called *differential privacy* may afford an added level of protection. Software offered by Sarus, a Y Combinator–funded start-up, applies this technique, which blocks algorithms built to publish aggregated data from disclosing information about a specific record, thereby reducing the chances that data will leak as a result of compromised credentials, rogue employees, or human error.

But privacy can be violated even with effectively anonymized data because of the way in which the data is collected and processed. An unintended violation occurred at the mapping company MaxMind, which provides geolocation services that enable businesses to draw customers' attention to nearby products and services. Geolocation also aids internet searches and can help if a service that needs your IP address (such as an entertainment streaming site) isn't working correctly. But precise mapping permits anyone who has your IP address to find your neighborhood and even your home. Combining your address with Zillow or some other real estate database can provide information about your wealth along with photos of your home inside and out.

> **Even when the reasons for collecting data are transparent, the methods used to gather it may be unethical. Will they involve any coercion or subterfuge?**

Unfortunately, IP mapping is not an exact science, and it can be difficult to precisely link an IP address to a physical address. A mapper might assign it to the nearest building or simply to a locality, such as a state, using that locality's central coordinates as the specific address. That may sound reasonable, but the consequences for one family renting a remote farmhouse in Potwin, Kansas, were horrific.

The family's IP address was listed with the map coordinates of the farmhouse, which happened to match the coordinates of the exact center of the United States. The problem was that MaxMind assigned more than 600 million other IP addresses that could not be mapped by any other means to the same coordinates. That decision led to years of pain for the family in the farmhouse. According to Kashmir Hill, the journalist who broke the story, "They've been accused of being identity thieves, spammers, scammers and fraudsters. They've gotten visited by FBI agents, federal marshals, IRS collectors, ambulances searching for suicidal veterans, and police officers searching for runaway children. They've found people scrounging around in their barn. The renters have been doxxed, their names and addresses posted on the internet by vigilantes."

Hill contacted a cofounder of MaxMind, who eventually produced a long list of physical addresses that had many IP addresses assigned to them and confessed that when the company was launched, it had not occurred to his team that "people would use the database to attempt to locate people down to a household level.” He said, “We have always advertised the database as determining the location down to a city or zip code level.” The takeaway is that well-intentioned, innocuous decisions made by data scientists and database managers can have a real, very negative impact on the privacy of innocent third parties. That brings us to the fifth P.

5. Preparation

How is the data prepared for analysis? How is its accuracy verified or corrected? How are incomplete data sets and missing variables managed? Missing, erroneous, and outlying data can significantly affect the quality of the statistical analysis. But data quality is often poor. Experian, a credit services firm, reports that on average, its U.S. clients believe that 27% of their revenue is wasted owing to inaccurate and incomplete customer or prospect data.

Cleaning data, especially when it is collected from different periods, business units, or countries, can be especially challenging. In one instance we approached a large international online talent-management and learning company to help us research whether women and men equally obtained the career benefits of training. The company agreed that the question was relevant for both its customers and the public at large, and therefore extracted the data it had on its servers. To ensure privacy the data was anonymized so that neither individual employees nor their employers could be identified. Because of the size of the data set and its internal structure, four individual data sets were extracted.

Normally we would just open the databases and find a spreadsheet file showing the features characterizing each individual, such as gender. A woman might be identified as “woman” or “female” or simply “F.” The values might be misspelled (“feale”), appear in various languages (*mujer* or *frau*), or use different cases (f or F). If the spreadsheet is small (say, 1,000 rows), correcting such inconsistencies should be simple. But our data contained more than one billion observations—too many, obviously, for a typical spreadsheet—so a cleaning procedure had to be programmed and tested.

One major challenge was ascertaining how many values had been used to identify the variables. Because the data came from the foreign subsidiaries of multinational firms, it had been recorded in multiple languages, meaning that several variables had large numbers of values—94 for gender alone. We wrote programming code to standardize all those values, reducing gender, for instance, to three: female, male, and unknown. Employment start and end dates were especially problematic because of differing formats for dates.

According to Tableau, a data analytics platform, cleaning data has five basic steps: (1) Remove duplicate or irrelevant observations; (2) fix structural errors (such as the use of variable values); (3) remove unwanted outliers; (4) manage missing data, perhaps by replacing each missing value with an average for the data set; and (5) validate and question the data and analytical results. Do the numbers look reasonable?

They may well not. One of our data sets, which recorded the number of steps HEC Paris MBA students took each day, contained a big surprise. On average, students took about 7,500 steps a day, but a few outliers took more than one million steps a day. Those outliers were the result of a data processing software error and were deleted. Obviously, if we had not physically and statistically examined the data set, our final analysis would have been totally erroneous.

How AI Raises the Stakes

Ethics can seem an expensive luxury for companies with strong competitors. For example, Microsoft reportedly fired the entire ethics team for its Bing AI project because, according to press and blog reports, Google was close to releasing its own AI-powered application, so time was of the essence.

But treating data ethics as a nice-to-have carries risks when it comes to AI. During a recent interview the CTO of OpenAI, the company that developed ChatGPT, observed, “There are massive potential negative consequences whenever you build something so powerful with which so much good can come…and that’s why… we’re trying to figure out how to deploy these systems responsibly.”

> **Too little anonymization is unacceptable under most government regulations. Too much may make the data useless for marketing.**

Thanks to AI, data scientists can develop remarkably accurate psychological and personal profiles of people on the basis of very few bits of digital detritus left behind by social-platform visits. The researchers Michal Kosinski, David Stillwell, and Thore Graepel of the University of Cambridge demonstrated the ease with which Facebook likes can accurately “predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender.” (This research was, in fact, the inspiration for Cambridge Analytica’s use of Facebook data.)

Subsequent research by Youyou Wu, Michal Kosinski, and David Stillwell reinforced those findings by demonstrating that computer-based personality judgments can be more accurate than human ones. Computer predictions of personality characteristics (openness, agreeableness, extraversion, conscientiousness, neuroticism—known as the Big Five) using Facebook likes were nearly as accurate as assessments by an individual’s spouse. The implications of that should not be ignored. How would you feel if your government wanted to catalog your private thoughts and actions?

A problem may also be rooted not in the data analyzed but in the data overlooked. Machines can “learn” only from what they are fed; they cannot identify variables they’re not programmed to observe. This is known as *omitted-variable bias.* The best-known example is Target’s development of an algorithm to identify pregnant customers.

The company’s data scientist, a statistician named Andrew Pole, created a “pregnancy prediction” score based on purchases of about 25 products, such as unscented lotions and calcium supplements. That enabled Target to promote products before its competitors did in the hope of winning loyal customers who would buy all their baby-related products at Target. The omitted variable was the age of the target customer, and the accident-in-waiting occurred when the father of a 17-year-old found pregnancy-related advertisements in his mailbox. Unaware that his daughter was pregnant, he contacted Target to ask why it was promoting premarital sex to minors.

Even by the standards of the era, spying on minors with the goal of identifying personal, intimate medical information was considered unethical. Pole admitted during a subsequent interview that he’d thought receiving a promotional catalog was going to make some people uncomfortable. But whatever concerns he may have expressed at the time did little to delay the rollout of the program, and according to a reporter, he got a promotion. Target eventually released a statement claiming that it complied “with all federal and state laws, including those related to protected health information.”

The issue for boards and top management is that using AI to hook customers, determine suitability for a job interview, or approve a loan application can have disastrous effects. AI’s predictions of human behavior may be extremely accurate but inappropriately contextualized. They may also lead to glaring mispredictions that are just plain silly or even morally repugnant. Relying on automated statistical tools to make decisions is a bad idea. Board members and senior executives should view a corporate institutional review board not as an expense, a constraint, or a social obligation but as an early-warning system.

A version of this article appeared in the July–August 2023 issue of *Harvard Business Review*.

MS

Michael Segalla is a professor emeritus at HEC Paris and a partner at the International Board Foundation.

DR

Dominique Rouziès is a professor of marketing at HEC Paris and the dean of academic affairs at BMI Executive Institute.