CAPCO

DATA ETHICS IN FINANCIAL SERVICES

BACKGROUND

Technology has been developing at an increasingly rapid pace across the world, and along with it, data generation, storage, and integration into everyday business operations. As organizations continue to invest more in their data capabilities and AI regulations try to keep pace globally, financial institutions (FIs) need to ensure they have a framework in place to manage data ethically across its' lifecycle.

FIs have a fiduciary duty to their clients, and are expected by society and governing bodies to act in the best interest of their clients. However, as FIs operate across economic and cultural boundaries, data regulations can be expected to differ, and consumers across regions have begun raising concerns around how companies are collecting, using, and sharing their data. Consumers' increase in cautiousness can lead them to lobby for more stringent government policies on how companies can use client data and how transparent they must be when doing so.

Having large credit card and credit bureau companies headlining on their lack of proper storage and security of sensitive client data exasperates the desire for more stringent government regulations. Consumers now have heightened awareness of their rights and are increasingly wary of providing their personal information to firms they do not trust entirely to handle their data. There are specific laws in various geographic regions that outline fundamental privacy rights such as the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, the Privacy Act in the United States, and the General Data Protection Regulation (GDPR) in Europe. Although business leaders are likely inclined to want to use the highest volume of data along with the most advanced data analytics techniques to help them to make improved business decisions, they need to be cognisant not to misuse the data entrusted to them.

Consumers are becoming increasingly wary of the potential pitfalls of big data companies (Nichols, 2018), and as a result are becoming more demanding that their data be managed ethically and with transparency. This has led to an increasingly common trend; consumers threatening to discontinue doing business with organizations that do not meet their data ethics expectations. This could lead to significant and unrecoverable losses if FIs are not careful with how they are treating their clients' data with respect to both customer desires and regulations. Therefore, FIs must be aware of what their customers expect when it comes to data privacy, and must follow regulations to avoid potential lawsuits and damage to their brand.

However, the pace of technological advancement means that laws and regulations can quickly become obsolete, leaving grey areas for day-to-day operational decisions. FIs are now capable of making better use of customer data to help them increase customer satisfaction and profitability, and how they collect and use customer data has changed drastically. Therefore, FIs must ensure they are not only up to date with the policies within the regions they are operating, but have a robust ethical framework in place that spans across the organization to mitigate the risk of running into ethical issues when and where the laws are out of date because they have not accounted for the latest technologies or associated capabilities.

Another consideration for FIs is coordinating the sharing of data with third-party vendors in 'open-banking' ecosystems. FIs must ensure that they are moving this data securely and that the organizations they share data with will follow equally stringent privacy policies with their clients' data.

This brings us to question how FIs can ensure that they are not only following data laws, but also instilling trust in their customers and the public at large, that they are guarding and using their personal information ethically.

PILLARS FOR ETHICAL DATA MANAGEMENT

To address this we have established five pillars that FIs should leverage at the top of the organization and across geographies and lines of business to ethically manage data. As such, these pillars for ethical data management are agnostic to the domain or business units to facilitate organization-wide adoption. The pillars are: transparency, regulations, fairness and reliability, security and privacy, and accountability.

Transparency

Being transparent with data means ensuring that how data is being collected, stored, transformed and used is thoroughly documented. This documentation (in a digestible and accurate format) must be accessible to both regulatory bodies and consumers whose data the FI uses to derive insights, make predictions, or decisions. Examples of poor data transparency have proven to have damaging consequences to the firms involved. The Facebook/Cambridge Analytica incident in 2018 is a prime example where personal user data of the social media platform was shared with the third-party (Cambridge Analytica), without user knowledge. This resulted in an overall loss of trust from Facebook users and a drop in stock price, while serving as a flag to global regulators that regulations around transparency required modernization (Kozlowska, 2018).

Because of this, we now know that transparency requires going beyond the raw data itself, and must include the processing and transformation of data, as well as the intent behind any analytics or AI model in which the data is leveraged.

Regulations

Thorough documentation of how data is being collected, stored, transformed and used is not only a pillar of good data ethics but a critical aspect of business strategy to not only improve performance, but also to avoid penalties from regulatory bodies. Regulatory standards should be considered a baseline, to which FIs should be proactive in surpassing to stay ahead of potential ethical or regulatory issues that could arise in the future.

For example, GDPR gives more control to consumers over their data and the processing of it (Directorate-General EU, 2020), while ensuring the individuals behind the data are not identifiable. GDPR serves as a reliable benchmark. However, as we have seen, technology is evolving at such a rapid pace that often regulations have a hard time keeping up. As new technologies become available that facilitate new ways of creating, accessing or leveraging data, FIs must ensure that those applications follow stringent internal standards as regulations may not yet have been developed to the same degree.

Fairness and Reliability

The proliferation of data across the world is one of the reasons FIs have begun to incorporate AI and machine learning across a wide variety of use cases to minimize human effort and augment human decision making. However, a historical lack of fairness in certain areas can cause a systemic issue in AI and machine learning models if trained on compromised data.

For example, consider the Massachusetts Institute of Technology who recently uncovered a systemic inability for AI to detect early signs of breast cancer in mammogram data from women in minority groups. This stemmed from the fact that these groups historically had less access to healthcare in the United States, resulting in there being far less data representing these groups during the training of the AI models (McGreevey, 2018).

It is important for FIs to be acutely aware of this potential pitfall across all advanced analytics and AI use-cases, however a few key examples include; credit decisioning on loan applications, pre-approval offers, and hiring and human resource decisions. FIs must carefully evaluate the data they are using and examine it for potential bias before training and deploying an AI model on it. This will ensure that models can make decisions without prejudice that can unintentionally lead to discriminatory decisions.

This is especially important to remember if models perform continuous learning. Continuous learning means that information is continuously put into the models as time passes (Lu et al.,2016). The continuous stream of data may cause the model to change over time and can lead to concept drift – meaning the model changes the way outputs are produced, affecting the fairness and consistency of predictions.

Privacy and Security

With the vast amounts of data being generated, stored, and processed, how do FIs ensure it is kept secure and that the privacy of customers, employees, and partners is maintained? The question becomes even more pressing as FIs begin to face regulatory pressures to ensure data assets are being made available to third-party providers from regulations such as PSD2. While at the same time, there has been a rise in number of data breaches of global FIs where large volumes of sensitive, personal consumer data were exposed. As a result, FIs are facing a difficult set of demands; to increase data sharing and openness, while also increasing security.

In order to meet these demands, FIs must establish built for purpose data architectures across their lines of business. Architectural design must provide both a high level of security to personal data while providing the organization with the ability to quickly access, manipulate, and join datasets to extract value from the data. FIs must also accompany architectural design with a clear data governance structure, including data stewardship and data lineage as well as robust master and metadata management. This not only support privacy and security concerns, but the previously discussed pillars as well.

Accountability

Ensuring FIs remain accountable for ethically managing data and following the previously discussed pillars is not a trivial task. It requires an on-going commitment from the top of the organization. To support this, we suggest FIs do three key activities:

1. Embed data ethics into use-case evaluation frameworks.
2. Establish a centralized Ethics Review Committee to review all use-cases before they move to implementation.
3. Understand how data flows across their organization, and apply the pillars of ethical data management across the data lifecycle.

Embedding data ethics criteria into use-case evaluation frameworks forces business decision-makers to consider:

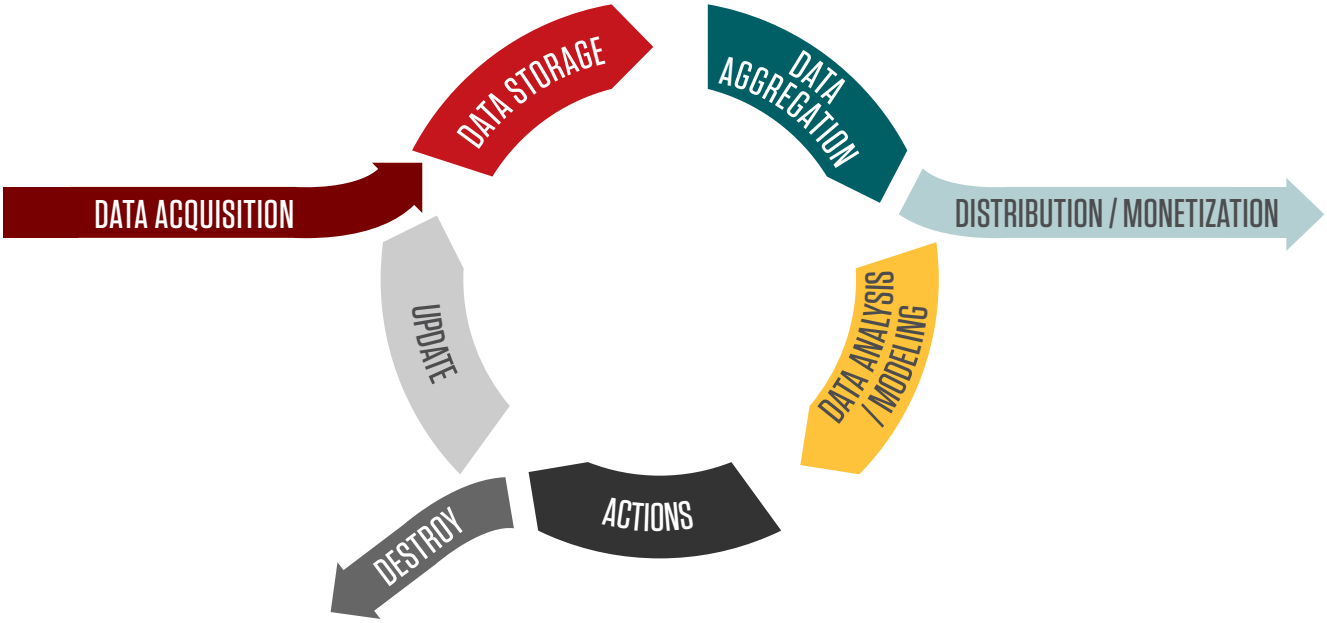
- How they will be transparent with their stakeholders.
- How they will ensure the data is being used and managed in a way that meets regulatory and internal standards.
- If the data to be leveraged in the use-case contains any inherent or unforeseen biases.

This evaluation will help ensure ethical decision making is engrained in use-case selection and prioritization. Establishing a central Ethics Review Committee to evaluate use-cases before going to production alongside the existing business review process will ensure that the use-cases follow an ethical framework. Specifically, for AI models, these reviews should occur at multiple stages from idea generation through data sourcing, model build, validation, and operationalization. Following these steps will help ensure the organization is applying data ethics consistently across lines of business and geographies.

IMPLEMENTING THE PILLARS OF DATA ETHICS ACROSS THE DATA LIFECYCLE

Frameworks can fall short or be short-lived in their implementation as they frequently are not tied directly to the actual day-to-day operational processes. To address this, we have mapped the pillars for ethical data management to the data lifecycle and provided context on how to connect the two.

To begin, let us first level-set on what the data lifecycle is. The diagram below lays out the stages of the data lifecycle. Note that the branches do not apply to all datasets or all organizations. Some organizations may not distribute any data assets for monetization externally. Some may do so with a specific subset of their data assets, while others core business model may center on data distribution and monetization.



Data Acquisition

FIs must be transparent with consumers on what data they collect, how they intend to use it, and how it will be kept safe. FIs must also ensure that the data they are generating or purchasing falls within regulatory guidelines, as well as stringent internal guidelines. FIs should evaluate the data to determine if it provides strategic value to the organization and improves outcomes for customers. If not, it should not be acquired.

FIs should examine if the datasets they generate through business activities or acquire from third parties are subject to a lack of fairness or poses any bias. For example, datasets containing lending decisions can reflect human bias of loan officers that may be discriminatory toward specific groups (Dobbie et al.,2018). Awareness of potential bias can shed light on operations that require changing, and ultimately prevent FIs from making biased decisions through the implementation of advanced analytics or AI models that contain an inherent bias due to the data they were trained on.

Data Storage

Establish the appropriate data architectures across lines of business rather than a 'one-size fits all' approach. Business units generate and acquire different data assets, with differing levels of sensitivity, different intended uses, and different regulatory requirements. The storage mechanism must minimize the risk of a breach (particularly for data that includes sensitive/personal information), while ensuring the organization has the right flexibility and scalability to support data access and aggregation as appropriate.

Data Aggregation

FIs must ensure consumers can be made aware through clear communication how they aggregate and transform their data by processing it, and/or joining it with other data sets that the organization may have purchased or created. Aggregated data sets can allow for masking or removal of personally identifying information; however, they can lead to aggregation bias. Aggregation bias stems from the assumption that a conclusion derived from an aggregated data set applies to all individuals or specific segments within the data set, while the underlying data usually has a distribution across a range. For this reason, FIs need to practice discretion when analyzing aggregated data as it can lead to overgeneralization.

Regulations are often less developed and clear-cut for aggregated data than raw data. However, regulations such as GDPR give more control to consumers over their data, and importantly, the processing of it.

As a result, FIs must have a clear understanding of what unit of measure is appropriate (such as customer, segment, or population) for a given analysis while considering the business question and the distribution of the underlying data to prevent aggregation bias. Further, FIs are required to have a clear step-by-step understanding of how they are processing and aggregating consumer data across the organization, with scalable processes in place to prevent it from happening at an individual level, if requested by the customer.

Data Distribution/Monetization

For organizations that partake in the external distribution and/or monetization of data, they must ensure consumers are aware of this and obtain informed consent before data is collected. Consumers must also be made aware of how the data will be stored, aggregated, and used within both the collecting organization and any that their data will be distributed or sold to. The collecting organization is referred to by GDPR standards as a data controller and is responsible for protecting the rights of the consumer. They must take responsibility to ensure partners (i.e. data processors) meet the same standards from a regulatory perspective and that the data they share will be secured to the same degree by the data processor as in the data controller. GDPR has clearly defined the roles and responsibilities of the data controller and the data processor so FIs can understand their legal obligations when handling the data. Organizations must establish a data processing agreement between the data processor and controller to ensure GDPR compliance from both parties. With regulatory and consumer pressure to make data available, this applies to an increasing number of FIs around the world.

Data Analysis/Modeling and Actions

FIs must be transparent with consumers on how their data will be analyzed, and what decisions or actions will be made based as a result of analysis or modeling. Regulatory bodies have certainly had a difficult time keeping up with the advancements in advanced analytics and AI. However, a gut-check FIs should do before putting a model into production, is:

> *"Does this model impact a customer in some way, if so, can we clearly understand and explain the model results?"*

For example, an FI looking to make credit decisioning should be able to clearly understand their model results, to be able to justify and explain why a decision was made, and to allow them to identify if the data was biased in anyway.

Destroy and Update

FIs must be able to communicate to their customers their data retention policies and destruction practices for personal information. Data that is retained will often be updated in several ways, including updates to personal information, changes in purchase patterns, as well as consumer responses to organizational actions taken. Exactly what data will be retained, and for how long, should align clearly to a strategic imperative for the organization, and these practices must fall within regulatory guidelines.

Mapping the Pillars for Ethical Data Management to the Data Lifecycle

	Data Acquisition	Data storage	Data Aggregation	Distribution/ Monetization	Data Analysis/ Modelling	Actions	Destroy	Update
Transparency	■		■	■	■	■	■	
Regulations	■	■	■	■	■	■	■	■
Fairness & Reliability	■				■	■		
Privacy and Security		■	■	■			■	■
Accountability	■				■	■		

The table above is meant to serve as a guideline for FIs. The Pillars for Ethical Data Management should be considered across the full data lifecycle, however, some stages of the data lifecycle require extra consideration from specific pillars.

SUMMARY

The technological landscape is evolving at an increasingly rapid pace, and with it, the volume of data and the opportunities to extract value from that data. However, so is the number of ways in which data can be compromised or misused. Because of this, consumers around the world have become increasingly aware of how important their data can be.

This has created growing public concern around how organizations are collecting, storing, utilizing, and sharing personal data. Although regulators are responding, they have not been able to keep up with the speed of technological innovation. This places the onus on FIs to hold themselves to a higher standard to manage data ethically than the standards that have been set by regulators.

To achieve this, we suggest FIs focus on instilling the Pillars of Ethical Data Management across the organization by understanding how each of the pillars connects to different aspects of the lifecycle of data, and use this as a guideline to ensure they ethically manage data.

REFERENCES

SOURCES

Directorate-General. European Union. EU Data Protection Rules. Accessed March 2020.

Dobbie, W., Liberman, A., Paravisini, D., & Pathania, V. (2018). Measuring Bias in Consumer Lending.

Lu, N., Lu, J., Zhang, G. & Lopez de Mantaras, R. 2016. A concept drift-tolerant case-base editing technique. Artificial Intelligence.

McGreevey, S. March 7, 2018. Race bias seen in breast-cancer screening. The Harvard Gazette.

Nichols, M. November 15, 2018. Big Data is so large, it's raising privacy & ethical concerns. Euro Scientist.Kozlowska, I. (2018, April 30). Facebook and Data Privacy in the Age of Cambridge Analytica. Retrieved from https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/

AUTHORS

Simon Campbell, Principal Consultant
simon.campbell@capco.com

Sonia Chau, Associate
sonia.chau@capco.com

Lavesh Bansal, Associate
lavesh.bansal@capco.com

Anna Chen, Associate
anna.chen@capco.com

Lyn Pham, Associate
lyn.pham@capco.com

Eddie Kim, Associate
eddie.kim@capco.com

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward.

Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and asset management and insurance. We also have an energy consulting practice in the US. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

To learn more, visit our web site at www.capco.com, or follow us on Twitter, Facebook, YouTube, LinkedIn and Instagram.

WORLDWIDE OFFICES

APAC
Bangalore
Bangkok
Hong Kong
Kuala Lumpur
Pune
Singapore

EUROPE
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto
Tysons Corner
Washington, DC

SOUTH AMERICA
São Paulo

WWW.CAPCO.COM

CAPCO