

**College of Humanities**  
**Master of Financial**  
**Technology-**  
**Fin Tech**

**Artificial Intelligence for Financial Technology**

FNT 626

**Lecture 04**

**Prepared By:**

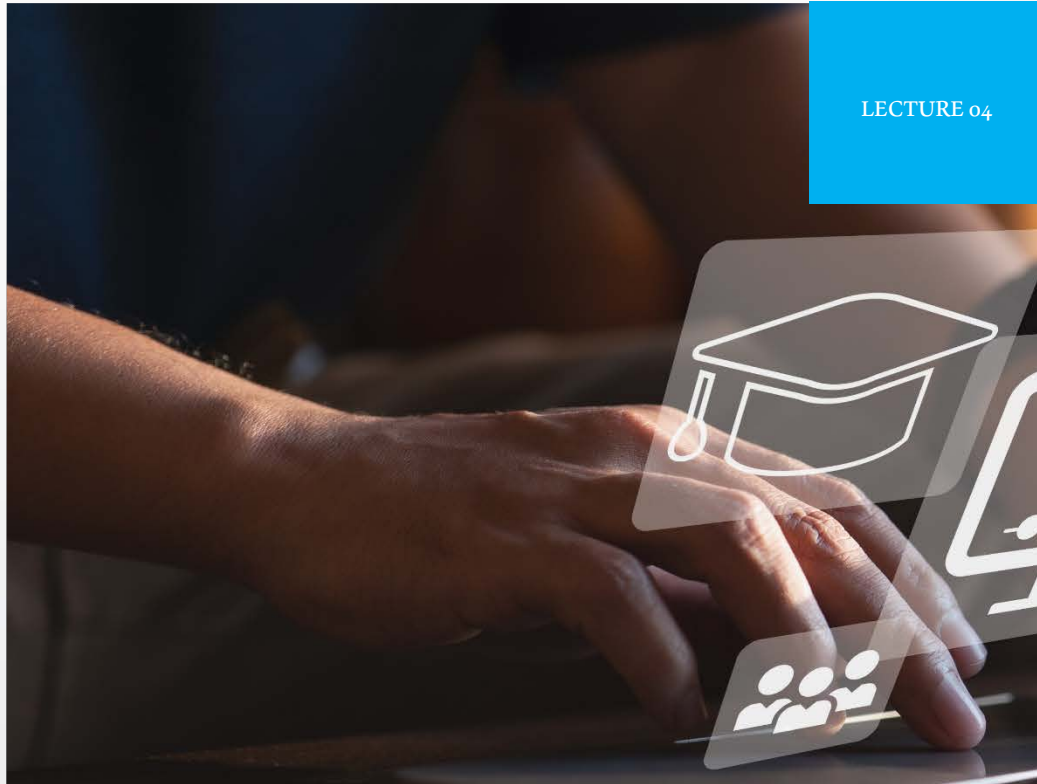
**Dr. Abeir Moussa**



*Artificial Intelligence for Financial Technology*

FNT 626- AI for Fin Tech

Lecture 04



## AI Governance Frameworks for the Banking Sector





LECTURE 04

## Ch.5: AI Governance Frameworks for Banking Sector

- **Learning Objectives**

1. **Understanding AI Governance in Banking**
2. **Explore the Ethical Principles and Regulatory Compliance**
3. **Understand the key components of AI governance frameworks related to explainability of AI-driven decisions,**
4. **Examine the critical aspects of data governance, data privacy, and data security within the context of AI governance in banking,**
5. **Assess the Model Validation and Testing reliability, fairness, and compliance with regulatory requirements.**
6. **Understand the Accountability and Oversight,**
7. **Understand the Adaptation to Regulatory Changes,**
8. **Address Mitigating Bias and Discrimination.**



LECTURE 04

## Ch.5: AI Governance Frameworks for the Banking Sector

### Chapter outline

1. INTRODUCTION
2. COMMON MISGIVINGS
3. FAIRNESS, TRANSPARENCY, AND ACCOUNTABILITY
4. LEGAL FRAMEWORK OF FAIRNESS FOR AI-DRIVEN DECISIONS.
5. THE IMPORTANCE OF AI GOVERNANCE
6. CONCLUSION

## **1. INTRODUCTION**

### **1.1 The Role of AI in Banking**

### **1.2 The Imperative for AI Governance in Banking**



LECTURE 04

## **Ch.5: AI Governance Frameworks for the Banking Sector**

## Ch.5: AI governance frameworks for the banking sector

### 1. INTRODUCTION

**Artificial Intelligence (AI)** has emerged as a transformative force in the banking sector, offering opportunities to enhance operational efficiency, customer experience, risk management, and decision-making processes. However, the adoption of AI in banking also presents unique challenges related to **ethics, transparency, accountability, and regulatory compliance**. As banks increasingly integrate AI into their operations, the need for robust AI governance frameworks becomes paramount to ensure responsible and ethical use of AI technologies

## Ch.5: AI governance frameworks for the banking sector

### 1. INTRODUCTION

#### 1.1 The Role of AI in Banking

AI technologies, including machine learning, natural language processing, and predictive analytics, have revolutionized various aspects of banking, such as customer service, fraud detection, credit scoring, and investment advisory. The ability of AI to analyze vast volumes of data, identify patterns, and automate decision-making processes has the potential to drive innovation and efficiency in banking operations.

## Ch.5: AI governance frameworks for the banking sector

### 1. INTRODUCTION

#### 1.2 The Imperative for AI Governance in Banking

The use of AI in banking introduces complex ethical and regulatory considerations, particularly concerning customer data privacy, algorithmic bias, explainability of AI-driven decisions, and compliance with financial regulations. As a result, the establishment of AI governance frameworks is essential to ensure that AI applications in banking adhere to ethical standards, mitigate risks, and align with regulatory requirements.

## 2. COMMON MISGIVINGS

LECTURE 04



## Ch.5: AI Governance Frameworks for the Banking Sector

## Ch.5: AI governance frameworks for the banking sector

### 2. COMMON MISGIVINGS

The apprehension that some people feel about AI is not entirely unfounded. As with the implementation of any new technology, there have been—and continue to be—notable instances where the implementation of imperfect, early-generation technology has illustrated the very problems that people fear. For example, Amazon's AI-powered recruiting engine was shut down by the company due to concerns about discrimination. The model was trained by observing patterns in résumés, most of which were résumés of male applicants in the technology industry.

## Ch.5: AI governance frameworks for the banking sector

### 2. COMMON MISGIVINGS

This led the model to conclude that male applicants were preferable and penalize résumés that includes words like “women’s”, such as “women’s chess club captain.” Another example is the credit scoring model for a new credit card offered by Goldman Sachs and Apple, the “Apple Card,” which was criticized for sexist decision-making. Opacity makes it difficult to explain how such disparate decisions were reached, making it appear that the AI models generated biased decisions. A 2018 UK House of Lords report highlighted that training AI with unrepresentative data sets may lead to unfair decisions and emphasized the need for developers to consider biases embedded in the algorithms themselves. In response to these kinds of concerns, the **Organization for Economic Co-operation and Development (OECD)** Council on Artificial Intelligence issued recommendations for the responsible stewardship of trustworthy AI. These form part of the first steps toward an international framework for AI governance.

## Ch.5: AI governance frameworks for the banking sector

### 2. COMMON MISGIVINGS

Fairness refers to AI actors respecting the rule of law, human rights and democratic values. Transparency/explain-ability is defined to mean a commitment to responsible disclosure concerning AI systems, including provision of meaningful information which is appropriate in the context. This information should enable those affected by an AI system to understand the outcome and enable those adversely affected to challenge the outcome. Accountability means being accountable for the proper functioning of AI systems with respect to the above principles.

Professor Ryan Calo argues that “it may not be wise or even feasible to pass general laws about artificial intelligence at this early stage.” Given that applications of AI and degrees of technological sophistication vary significantly from field to field, an overarching legislative approach is likely to be ill-adapted to specific contexts. While the increasing sophistication of AI models creates a layer of novel issues, the best approach to governance of AI at this stage is to enhance existing governance frameworks.

## Ch.5: AI governance frameworks for the banking sector

### 2. COMMON MISGIVINGS

The European Union, by way of contrast, has opted for a general approach to AI regulation. Act” in which it sets stringent regulatory requirements for the use of AI on the EU market. The proposed approached set forth by the EU is to concentrate regulatory requirements and thus regulatory enforcement of AI that is considered “high risk” by the regulation. It is important to note that AI does not equate to automation. Although AI includes automation of certain tasks, an AI model output may be one element of a larger, more complex strategy or decision-making process. Conversely, some end-to-end automated processes do not necessarily include AI.

## Ch.5: AI governance frameworks for the banking sector

### 2. COMMON MISGIVINGS

Although AI and automation may converge, it is important to differentiate the two concepts as part of a rigorous legal, ethical and policy analysis. This means that processes involving AI models also involve human action, meaning that AI models are not to be self-executing. Such an approach constitutes one way of countering AI opacity and exercising control over decisions. It is also important to differentiate the two for legal analysis—automated decisions, which as explained may or may not include AI-based output, are governed by privacy legislation when they use personal information. AI, as for now, is not specifically targeted by any active legislation, but by way of example, the EU's A.I. Act targets AI, specifically naming the underpinning technology in the AI definition.

### **3. The Key Components of AI Governance Frameworks for Banking Sector: FAIRNESS, TRANSPARENCY, AND ACCOUNTABILITY**

**3.1 Issues of Fairness**

**3.2 Issues of Transparency**

**3.3 Issues of Accountability**



LECTURE 04

## **Ch.5: AI Governance Frameworks for the Banking Sector**

### 3. The Key Components of AI Governance Frameworks for Banking Sector: FAIRNESS, TRANSPARENCY, and ACCOUNTABILITY.

The integration of **Artificial Intelligence (AI)** in the banking sector has ushered in a new era of innovation, efficiency, and customer-centric services. However, the adoption of AI in banking also brings forth complex ethical, regulatory, and risk management challenges. To ensure responsible and ethical use of AI technologies, robust AI governance frameworks are essential. These frameworks are designed to uphold ethical standards, mitigate risks, and align with regulatory requirements specific to the banking industry.

**Ch.5: AI governance frameworks for the banking sector****3.1 Issues of Fairness**

Traditional credit adjudication models entail a regression analysis—a statistical method that analyses the relationship between two or more variables based on past data in order to predict future outcomes. The data used to assess creditworthiness are usually payment and credit history, outstanding debts, debt-to-credit ratios, and a credit score from a credit reporting agency. A credit score is “designed to predict the relative likelihood of a negative financial event, such as a default on a credit obligation.” A customer’s credit score is one of the most important factors in a decision to issue credit. Concerns about fairness in credit decisions predate the use of AI in credit modelling. For this reason, credit reporting agencies are under an ethical obligation not to use certain personal information for credit scoring. It is reasonably foreseeable that if credit reporting agencies are negligent in the way they gather and publish information, it could cause harm to consumers because credit grantors may deny credit on that basis. Fairness concerns can stem from the fact that traditional metrics used for credit scoring may be under-inclusive, and thus may not present the full picture of an applicant’s credit-worthiness. For example, those who have not had access to formal credit systems (e.g., recent immigrants) may have shorter credit histories, which can make it more difficult to access credit.

### 3.1 Issues of Fairness

New creditworthiness assessment methods like AI have the potential to enhance understanding of both traditional and new data, thereby addressing issues of data under-inclusivity and broadening access to credit. It is acceptable for credit providers to differentiate between potential clients in good faith and based on what is reasonable under the circumstances. The data traditionally used to make credit adjudication decisions are based on longstanding industry best practice and have generally served economies well. The use of AI in credit models may introduce novel risks around fairness and discrimination. Performance of AI models is reliant, among other things, on the accuracy and diversity of the data used for development and deployment. According to the **Monetary Authority of Singapore (MAS')**, some AI techniques such as clustering and segmentation of data have the capacity to differentiate between groups more rapidly and on a larger scale than more rudimentary models or humans. An AI credit model that is trained with a bias is likely to replicate that bias, underscoring the importance of robust governance to avoid unintended unfair bias in AI-driven decisions.

### 3.2 Issues of Transparency

AI models and AI-driven decisions can give rise to concerns around opacity. The fear is that AI models, and more specifically ML, are a “black box” which inhibits the ability to explain how a decision was reached. Although this chapter focuses on fairness rather than transparency/explain-ability, it is worth mentioning that the degree of opacity of AI models varies considerably. Thus, to an extent, AI enablers can choose how opaque of a model to use, in light of the particular legal or ethical obligations that apply in a given context of AI use. For this reason, banks may choose to refrain from deploying AI trained with unsupervised learning. The issues that arise from a lack of transparency are context-specific, and vary according to whether an explanation is owed to the decision subjects and the degree of detail that the explanation must provide. However, in decisions that have a greater impact on people’s lives, such as the ability to get a mortgage, the transparency/explain-ability of the decision assumes greater importance.

### 3.2 Issues of Transparency

AI enablers should choose the appropriate level of opacity for a given AI model based on these factors. Banks, for instance, may deliberately choose to use types of AI that are less opaque based on regulatory requirements, or other legal and ethical obligations. The inability to explain exactly how an output was reached does not mean that AI enablers are unable to understand that output more generally. As an analogy, consider the role of brain chemistry in the context of human decision-making: an exact explanation of how the neurons worked in order to arrive at a given decision is not necessary to understand the decision. Rather, one can offer a meaningful explanation of the decision based on the human being's values, priorities, rules, and reasoning. Therefore, opacity does not necessarily impede or lessen the accountability of AI enablers. A meaningful explanation can be enough to conclude that a decision was fair and transparent.

### 3.2 Issues of Transparency

The relevant question is “What do we need to know about this algorithm to keep it accountable and functioning according to laws, rights and social norms?” For example, AI models used by the Houston, Texas school district to assess the performance of teachers attributed a score to each teacher. The teachers’ union challenged the use of the AI models in court and required an explanation as to how the models arrived at scoring decisions for teachers. A global explanation will discuss “system functionality,” meaning the design of the AI model in general. Such an explanation could discuss general functionality components, such as the requirements, criteria, and decision trees underpinning the decision-making. A local explanation will discuss the rationale for a specific decision. This targets the circumstances of a given AI-driven decision, the weight of the applicable features, and the precise information on the algorithmic rules that determined the output. However, not all decisions, such as decisions by banks with or without AI, hold this weight and significance. In the **European Union (EU)**, the **General Data Protection Regulation (GDPR)** establishes a consumer’s “right to be informed.” This refers to a right to an explanation of “the logic involved, as well as the significance and the envisaged consequences of automated decision-making systems.”

### 3.3 Issues of Accountability

AI-powered technology raises issues around the potential for unfair bias and the opacity of AI-driven decision-making. These two issues tend to exacerbate one another—with greater opacity fueling concerns about potential bias and vice versa. This, in turn, underscores broader questions about who is accountable for AI technologies and how they will be held accountable. Banks that are incorporating increasingly sophisticated AI into their day-to-day operations confront questions not only around refining and improving early-generation AI models, but also devising and implementing sound governance practices that will ensure compliance with relevant laws and, ultimately, help build trust in AI-powered technologies.

### 3.3 Issues of Accountability

A key concern about accountability for the implementation of AI is whether and how to hold the AI enablers, including organizations, legally responsible. Although AI models are increasingly sophisticated, the models underpinning a technology-based service or activity are not legally independent from the enabling person or organization. Therefore, AI enablers can be held legally responsible for breaking legal rules by extending existing legal concepts for holding organizations liable. For example, if AI was operated in such a way as to result in negligence, this could give rise to an action in tort against the AI enabler. In the EU, AI enablers are responsible for compliance with the GDPR restrictions on data processing and automated decision-making, including AI. We disagree with the result in this case, and argue that attaching legal liability to AI enablers that cause harm can help promote trust in AI-driven systems.

### 3.3 Issues of Accountability

In the past, the models used by the financial sector have been fairly rudimentary; however, we posit that the applicable legal frameworks extend to more sophisticated models as well. Responsibility for the use of models is tied to the particular activities of the AI enabler and therefore responsibility attaches regardless of the degree of technological sophistication with which the activities are performed. Thus, AI enablers must ensure that AI-driven decisions respect applicable legal rules. Credit adjudication processes that utilize AI models tend to have human involvement at a few points in the end-to-end process. This is known as having a “human-in-the-loop.” Human employees in the different teams are involved in the credit adjudication, review, and even override decisions generated by the AI model.

### 3.3 Issues of Accountability

In addition, AI credit adjudication models are subject to recurring back-testing, fail-safe mechanisms, safety short-circuits, fail-over (i.e., back-up models), and close monitoring to ensure that the models behave as intended. Holding AI enablers accountable may entail extending legal concepts and adapting approaches, along with developing a deeper understanding of how AI technology works. Therefore, the legal and governance implications associated with the adoption of AI are probably not as dramatic as some people think. Therefore, AI enablers must ensure that AI-driven decisions respect fundamental rights (**both legally and ethically**), including equality rights.

## **4. LEGAL FRAMEWORK OF FAIRNESS FOR AI-DRIVEN DECISIONS**

### **4.1 Applicable Law**

### **4.2 Fairness in AI Models**

#### **4.2.1 Direct discrimination**

**in AI models**

#### **4.2.2 Indirect discrimination in AI models**

#### **4.2.3 *Bona fide* justification**



LECTURE 04

**Ch.5: AI Governance Frameworks for the  
Banking Sector**

#### 4. LEGAL FRAMEWORK OF FAIRNESS FOR AI-DRIVEN DECISIONS

The increasing integration of Artificial Intelligence (AI) in decision-making processes across various industries has raised critical concerns regarding fairness, accountability, and transparency. In response to these concerns, the legal framework of fairness for AI-driven decisions has emerged as a pivotal area of focus. This framework aims to address the ethical and legal implications of AI technologies, particularly in ensuring that AI-driven decisions uphold principles of fairness, non-discrimination, and compliance with anti-discrimination laws and regulations. As AI technologies play an increasingly prominent role in decision-making processes, particularly in areas such as lending, hiring, and predictive analytics, the need to ensure fairness and non-discrimination has become paramount. AI-driven decisions have the potential to perpetuate biases, inadvertently discriminate against certain groups, and impact individuals' rights and opportunities. Therefore, the legal framework of fairness for AI-driven decisions seeks to mitigate these risks and uphold ethical and legal standards.

The legal framework of fairness for AI-driven decisions is underpinned by anti-discrimination laws and regulations that prohibit discrimination based on protected characteristics such as race, gender, age, disability, and other factors. These laws, such as the Civil Rights Act of 1964 in the United States, the Equality Act in the United Kingdom, and similar legislation in other jurisdictions, serve as the foundation for ensuring fairness and non-discrimination in decision-making processes.

## 4. LEGAL FRAMEWORK OF FAIRNESS FOR AI-DRIVEN DECISIONS

### 4.1 Applicable Law

As explained above, banks, as AI enablers, are accountable when it comes to complying with legal and regulatory obligations. This is the case whether banks perform tasks internally or through outsourcing – regardless of any private indemnity agreement. The only exceptions are instances where accountability specifically belongs to someone else. So which areas of law apply? AI governance in the banking sector has implications across a wide range of legal and ethical areas, from consumer protection, credit reporting regulations, prudential regulation, privacy law, as well as ethical obligations within organizations. AI governance in the banking sector has implications across a wide range of legal and ethical areas, from consumer protection, credit reporting regulations, prudential regulation, privacy law, as well as ethical obligations within organizations. There are differences from jurisdiction to jurisdiction in terms of the specific sources of applicable law (e.g., case law, legislation, codes, regulations, rules) and the entities responsible for enforcement (e.g., public agencies, private parties). In addition, the nature of a country's legal system (e.g., federal or unitary) may affect the form of legal and judicial oversight of AI. In the context of banking, **Basel III** is an international accord which addresses inter alia issues related to model risk that are relevant for AI.

## 4. LEGAL FRAMEWORK OF FAIRNESS FOR AI-DRIVEN DECISIONS

### 4.1 Applicable Law

In the context of fairness and concerns about discrimination, the **United Nations Declaration of Human Rights (UNDHR)** provides an international standard for human rights. Human rights law is central to discussions around the fairness of AI-driven decisions because it is often the most powerful legal mechanism through which one can challenge a decision on the basis of discrimination. While the general concept of fairness can have different meanings depending on the region or culture, international standards with respect to human rights and the UNDHR provide a common standard on which to base a framework for sound governance of AI. Regardless of whether a state has domestic human rights law or adheres to the UNDHR, the general principles of ethical and responsible AI apply to AI deployment. We argue that banks should be guided by these principles—specifically fairness, transparency/explain-ability, and accountability—as the guiding principles of AI governance. Additional, widely recognized ethical principles may come from noteworthy publications and organizations, public and private, operating in the AI governance space.

## Ch.5: AI governance frameworks for the banking sector

#### 4. LEGAL FRAMEWORK OF FAIRNESS FOR AI-DRIVEN DECISIONS

##### 4.1 Applicable Law

In addition to human rights law, the following fields of law may include requirements that pertain to the fairness of AI-driven decisions in the banking sector:

- **Prudential oversight of modeling:** Prudential regulators in some jurisdictions regulate the use of models by banks by establishing certain technical requirements, usually in a technology-neutral way. Although prudential regulation may not address fairness issues specifically, a guideline relating to models could impose requirements for transparency, bias testing, and/or force banks to validate their models.
- **Consumer protection (including fair lending):** Consumer protection law typically regulates many of the terms upon which businesses transact with consumers, with a view to prohibit a range of unfair practices. Restrictions on the use of personal information are significant for AI development and governance, as AI depends on large volumes of data in order to learn.
- **Privacy law may restrict the use of “sensitive data”**— data which corresponds to, or has a strong linkage to, prohibited grounds for discrimination.
- **Consumer credit reporting:** This regulation typically limits the type of information that can be used by credit reporting agencies to assess credit-worthiness and generate credit scores. It applies mainly to credit reporting agencies.

**Ch.5: AI governance frameworks for the banking sector****4. LEGAL FRAMEWORK OF FAIRNESS FOR AI-DRIVEN DECISIONS****4.2 Fairness in AI Models**

A general definition of “discrimination” is “The action of perceiving, noting, or making a distinction between things.” In this neutral sense, discrimination is neither illegal nor unethical.

In the banking industry, for instance, it is acceptable to discriminate when deciding whether to extend credit based on a borrower’s creditworthiness, as it is a fundamental duty of a bank to assess credit risk. In legal parlance, however, the word discrimination is often used in a specific sense to mean basing decisions on distinctions in a manner which is potentially illegal. At a conceptual level, equality rights laws—in line with the principle of fairness—prohibit private organizations from denying access to a good or service available to the public based on a prohibited ground of discrimination or sensitive data.<sup>80</sup> Prohibited grounds of discrimination include race, color, sex, language, religion, etc. In the USA, for example, fair lending laws prohibit reliance on certain protected classes, such as sex or marital status, and then, later, race, nationality, religion, age, as a basis for lending decisions.

Therefore, banks should be familiar with the prohibited grounds of discrimination and sensitive data in the jurisdictions in which they operate, as well as internationally accepted standards and ethics. For example, discrimination based on age is permissible for the purpose of establishing a minimum age to qualify for certain roles, such as that of a corporate director.

## 4.2 Fairness in AI Models

### 4.2.1 Direct discrimination in AI models

Direct discrimination occurs when a person discriminates against another person or a group because of, or based on a prohibited ground without a legitimate justification. For example, direct discrimination would result if disability itself was a factor in a decision about offering a good or a service customarily offered to the public. Thus, direct discrimination entails intent or knowledge insofar as a prohibited ground is an express factor considered by AI enablers in decision-making.

In the context of AI models, direct discrimination would probably not stem from the model itself, but from the underlying business rules or strategies decided upon by AI enablers. This is because for an AI model to perpetrate direct discrimination, its enablers would have to make a protected ground a data point on which decisions concerning goods or services were taken. It is unlikely that an institution would purposely design an AI model to discriminate in this way, as much effort generally goes into ensuring models are designed to respect legal rules, and that credit can be extended to as many customers as possible within the risk appetite and boundaries dictated by prudential rules.

## 4.2 Fairness in AI Models

### 4.2.2 Indirect discrimination in AI models

Indirect discrimination occurs when otherwise neutral policies hurt certain groups. Intent or knowledge is not a requirement for indirect discrimination. It is concerned with whether the effect of a policy or a decision results in systematic discrimination against a protected class. An example is Amazon's AI-driven recruiting engine. This AI model learned to prefer male applicants to female applicants because it was trained using a data set that contained more men than women even though this was not the intent of its enablers. In other words, discrimination arose due to the fact that the data set was biased in favor of men. As noted by the House of Lords in a 2018 report, the unrepresentativeness of data sets gives rise to concerns that AI models may "reflect the wider prejudices of societies past and present." This underscores the need for AI enablers to have robust governance in place to oversee and monitor the data sets used to train AI.

An AI model that engaged in indirect discrimination would likely do so by making new or unintended associations or correlations between data points in the course of making decisions. The initial data points included in the model may not be prohibited grounds or sensitive data. However, the associations made by the AI model may, in effect, link data with prohibited grounds or sensitive data, producing discriminatory outcomes.

## 4.2 Fairness in AI Models

### 4.2.2 Indirect discrimination in AI models

Other potential issues include inaccuracy of data, data sets not being representative of certain groups (over- or underrepresentation) or limited features regarding minority groups. If ethnic minorities tend to reside more in these areas, the AI model may infer that minorities are more likely to be in accidents, which could constitute indirect discrimination. The relationship between fairness and transparency is important when it comes to potential indirect discrimination through AI models. The more sophisticated the AI, the more challenging it may be to understand how the data was processed. Initial testing of the model prior to deployment and ongoing monitoring by model owners and users are essential to ensure that AI is behaving as intended.

## 4.2 Fairness in AI Models

### 4.2.3 *Bona fide* justification

A prima facie case of discrimination—whether direct or indirect—may be rebuttable if there is a bona fide justification for the decision. To justify the decision, the bank must have been acting in good faith in pursuit of a legitimate purpose and the requirement, qualification, or consideration on which the decision was based must be reasonable. While a decision based on a requirement, qualification, or consideration other than a prohibited ground or sensitive data will not give rise to direct discrimination, it may constitute a prima facie case of indirect discrimination. This prima facial case may be rebutted through a bona fide justification. “Equal” means “subject to all requirements, qualifications and considerations that are not a prohibited ground of discrimination.” For example, in Canada 52% of men and 49% of women are extended credit.

## 4.2 Fairness in AI Models

### 4.2.3 *Bona fide* justification

In other words, a decision based on a requirement, qualification, or consideration other than a prohibited ground will not constitute direct discrimination. It may, however, constitute prima facie indirect discrimination, which might be rebutted through a bona fide justification by showing the requirement, qualification, or consideration was used in good faith and was reasonable in the circumstances. Employing prudent loan practices based on the creditworthiness of potential customers therefore qualifies as a bona fide justification for discriminatory behavior. For credit decision-making, as well as decision-making regarding other products customarily made available to the general public by banks—whether AI-driven or human—this legitimate justification will generally be found in internal policies, procedures, business rules, and strategies. In the deployment of AI credit models, as in other bank business practices, it is essential that the model follow institutional policies, business rules and strategies. Banks have credit policies in place to prevent discrimination and avoid human rights complaints.

## 5. THE IMPORTANCE OF AI GOVERNANCE

LECTURE 04



**Ch.5: AI Governance Frameworks for the  
Banking Sector**

## Ch.5: AI governance frameworks for the banking sector

### 5. THE IMPORTANCE OF AI GOVERNANCE

Robust governance is key to developing and implementing AI models responsibly, to mitigate the issues discussed in this chapter. This chapter has argued that these requirements can be extended to AI-driven activities, and therefore they should serve as the basis for AI governance—along with the principles of fairness, transparency/ explain-ability, and accountability—within individual banks. At a high level, governance can be described as a workflow that has to be observed throughout the full lifecycles of AI projects. Despite the increasing sophistication of AI, it is submitted that the activities carried out by AI-driven models in the banking sector largely fall under existing laws and regulations, including equality rights. These laws and regulations “reach” the activities of banks and can be used to hold them accountable for their decisions, whether they are carried out by a human employee or an AI model. This underscores the importance of developing robust frameworks for governing AI models—including ethical principles, risk assessments, oversight, and monitoring—to ensure compliance with legislation, internal policies, and principles for responsible AI. The first part of governance occurs at the design and development stages of the AI lifecycle. During the design and development stages, governance should engage with both the strategy and scientific components of AI projects. From the strategy perspective, the choice of which AI models are deployed (whether built or bought) and how they are deployed must be informed by the FI’s codes of conduct. This will ensure that choices are compliant with internal policies, including those that pertain to equality rights law and the AI principle of fairness.

## Ch.5: AI governance frameworks for the banking sector

### 5. THE IMPORTANCE OF AI GOVERNANCE

The guiding principles of fairness, transparency/explainability, and accountability provide a foundation for developing sound AI governance practices at the institutional level. Therefore, banks should establish their own AI principles and codes of conduct, just as they would establish policies for other kinds of institutional activities. These principles can then be integrated directly into the workflow of the design, development, and deployment of AI projects. A thorough risk assessment should occur in the AI governance workflow, ideally at the design and development stages. To ensure that risks, such as discrimination risk, are captured, the risk assessment should consider the full lifecycles of AI models and projects. It suggests targeting and setting stringent requirements for AI that are considered “high risk.” Such AI models relevant to a financial institution’s activities may include creditworthiness and credit scoring, biometric identification, and human resources decisions. This will ensure that banks are aware of the degree of fairness risk posed by an AI model from the very early stages and enable the organization to prevent or mitigate this risk. This will foster meaningful accountability by the guiding principles of AI governance. The degree of oversight and monitoring should be informed by the criticality of the actions performed and the level of autonomy or sophistication of the AI model.

## Ch.5: AI governance frameworks for the banking sector

### 5. THE IMPORTANCE OF AI GOVERNANCE

Testing and monitoring should be aimed at both the AI models and the business decisions of which they are a part. Oversight methods may include regular stress tests, automated alerts, maker-checker decision processes, and the automatic stoppage of systems, activities, or products if AI models divert from intended purposes. Institutional oversight and monitoring should be aimed at ensuring that AI is deployed by that institution's AI principles and policies as well as applicable laws and regulations. From the standpoint of ensuring fairness, ongoing monitoring is useful to counter indirect discrimination, i.e., unintended bias arising from new associations and correlations inferred from AI models, as explained above. It is noteworthy that, to effectively monitor and test AI model outputs for discrimination, AI enablers do need to be allowed to use prohibited grounds of discrimination and sensitive data.

## Ch.5: AI governance frameworks for the banking sector

### 5. THE IMPORTANCE OF AI GOVERNANCE

The guiding principles of fairness, transparency/explainability, and accountability provide a foundation for developing sound AI governance practices at the institutional level. Therefore, banks should establish their own AI principles and codes of conduct, just as they would establish policies for other kinds of institutional activities. These principles can then be integrated directly into the workflow of the design, development, and deployment of AI projects. A thorough risk assessment should occur in the AI governance workflow, ideally at the design and development stages. To ensure that risks, such as discrimination risk, are captured, the risk assessment should consider the full lifecycles of AI models and projects. It suggests targeting and setting stringent requirements for AI that are considered “high risk.” Such AI models relevant to a financial institution’s activities may include creditworthiness and credit scoring, biometric identification, and human resources decisions. This will ensure that banks are aware of the degree of fairness risk posed by an AI model from the very early stages and enable the organization to prevent or mitigate this risk. This will foster meaningful accountability by the guiding principles of AI governance.

## 5. THE IMPORTANCE OF AI GOVERNANCE

The degree of oversight and monitoring should be informed by the criticality of the actions performed and the level of autonomy or sophistication of the AI model. Testing and monitoring should be aimed at both the AI models and the business decisions of which they are a part. Oversight methods may include regular stress tests, automated alerts, maker-checker decision processes, and the automatic stoppage of systems, activities, or products if AI models divert from intended purposes. Institutional oversight and monitoring should be aimed at ensuring that AI is deployed by that institution's AI principles and policies as well as applicable laws and regulations. From the standpoint of ensuring fairness, ongoing monitoring is useful to counter indirect discrimination, i.e., unintended bias arising from new associations and correlations inferred from AI models, as explained above. It is noteworthy that, to effectively monitor and test AI model outputs for discrimination, AI enablers do need to be allowed to use prohibited grounds of discrimination and sensitive data.

Overall, AI governance should be a multidisciplinary affair, breaking down silos that can typically be found in banks. Therefore, the governance architecture should entail a collaboration amongst all relevant parties, such as business owners, scientific and development teams, teams that use AI solutions, model risk and data governance, legal departments, etc.

## Ch.5: AI governance frameworks for the banking sector

### LECTURE 4

#### Conclusion

The chapter has provided a deep dive into the key components of AI governance frameworks, emphasizing the ethical principles, regulatory compliance, risk management, explainability, and accountability essential for responsible AI adoption in banking. Furthermore, the chapter has highlighted the imperative of continuous adaptation to regulatory changes, the mitigation of bias and unintended discrimination, and the need for training and awareness programs to equip banking professionals with the knowledge and skills to navigate the ethical, regulatory, and risk implications of AI technologies. As AI continues to evolve, the governance of AI in banking will face ongoing challenges related to the dynamic nature of AI, the need for continuous adaptation to regulatory changes, and the ethical dilemmas arising from AI-driven decision-making. In essence, the chapter serves as a foundational guide for banking professionals, regulators, and stakeholders, providing insights into the development and implementation of AI governance frameworks that uphold ethical standards, mitigate risks, and align with regulatory requirements, thereby fostering trust, transparency, and responsible AI adoption in the banking sector.

## Artificial Intelligence for Financial Technology

### References:

#### Required Reference:

Remolina, N. and Gurrea-Martinez, A. eds., 2023. Artificial intelligence in finance: Challenges, opportunities and regulatory developments. Edward Elgar.

#### Recommended Readings

Chishti, S., 2020. The AI book: the artificial intelligence handbook for investors, entrepreneurs and fintech visionaries. John Wiley & Sons.



*The End of Lecture*

**Thank You**