PROJECT SAPIENS:

THEMATIC REVIEW ON THE USE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN THE PHILIPPINE FINANCIAL SERVICES

2024

TECHNOLOGY RISK AND INNOVATION
SUPERVISION DEPARTMENT

TABLE OF CONTENTS

PROLOGUE

What is the future of the financial services with the rise of AI and machine learning?

The financial services industry is on the cusp of a significant transformation driven by the rise of Artificial Intelligence (AI) and Machine Learning (ML). Over the next 5-10 years, these technologies will redefine key aspects of finance, from customer service to investment strategies, risk management, and fraud detection. The integration of AI and ML promises to enhance operational efficiency, provide personalized services, and enable data-driven decision-making, but it also introduces new risks that require careful regulation.

AI and ML will revolutionize how financial institutions interact with customers. Personalized services, powered by AI, will enable banks and investment firms to offer tailored financial advice, automate transactions, and create customized financial products. AI-driven chatbots and virtual assistants will provide 24/7 support, improving the customer experience. Robo-advisors, which use machine learning algorithms to analyze market trends, will make wealth management services more accessible and affordable to the masses, democratizing investment opportunities that were once reserved for high-net-worth individuals.

AI will also significantly enhance risk management and fraud detection. Traditional systems rely on static rules to flag suspicious activities, but AI can process vast amounts of data in real-time, identifying potential fraud faster and more accurately. ML algorithms will improve over time, adapting to new threats and enhancing security measures. In lending, AI will help assess risk more precisely, analyzing alternative data sources and offering more accurate credit scores, especially for those without traditional credit histories.

However, the widespread use of AI and ML in financial services raises several risks. One significant concern is the potential for job displacement. Automation may replace many routine tasks currently handled by human workers, particularly in customer service and data entry roles. Financial institutions will need to invest in reskilling their workforce and creating new job opportunities in fields like data science and AI management.

Another major risk is data privacy and security. AI systems require vast amounts of data to function effectively, and mishandling or exposure of sensitive customer information could lead to significant breaches of trust. Financial institutions will need to implement robust cybersecurity measures to protect customer data and comply with privacy regulations.

Ethical concerns also arise with the use of AI in decision-making. If AI models are trained on biased data, they may perpetuate inequalities in lending or investment decisions, leading to discrimination. To mitigate these risks, financial institutions must ensure that their AI models are transparent, fair, and regularly audited for bias.

To regulate the use of AI in financial services, policymakers must establish clear guidelines and frameworks to ensure ethical practices, data privacy, and security. Regulation should focus on creating standards for transparency, ensuring that AI systems are auditable and that consumers understand how decisions are made. Additionally, regulations should address algorithmic bias and mandate regular audits to ensure that AI systems do not discriminate against certain groups.

In conclusion, while AI and ML offer significant potential to improve the financial services industry, careful regulation is necessary to address risks such as job displacement, data security, and ethical concerns. By implementing robust oversight and ensuring transparency, we can harness the benefits of AI while minimizing its drawbacks, creating a more efficient, inclusive, and secure financial ecosystem.

The foregoing was written by an AI, but the following study is not. We are on the verge of a technological frontier that has profound benefits if properly used and consequences if poorly regulated. This paper seeks to delve into the future of the financial services industry bearing in mind the current state of AI adoption in the country.

IN A NUTSHELL

KEY HIGHLIGHTS

- While it can be argued that the basic risks of AI are covered by existing rules and regulations, it is apparent that some gaps and risks must be addressed (such as accuracy, hallucination, data quality, and ethical issues). Ultimately, the **accountability for decision-making processes lies not on AI systems but on humans**.

- The BSP may opt to issue non-binding/soft regulations on the use of AI that are not too restrictive and allow for a more balanced approach towards innovative explorations and structured experimentations.

- Out of the 48 respondents to the BSP AI survey, 21 (44%) have deployed at least one (1) AI system in production, while 29 (60%) explicitly included AI/ML in their institutions' roadmap. This shows that financial institutions in the Philippines are eager to join in the AI revolution.

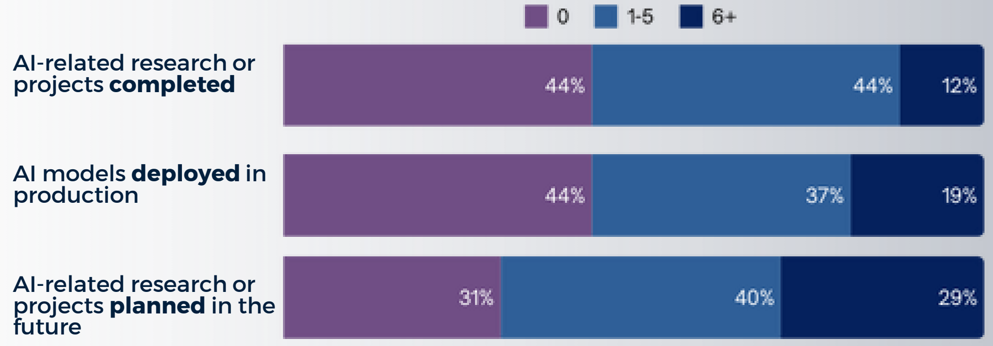
- For survey respondents, the benefits of using AI/ML contribute to business/process optimization, automation, and customer satisfaction. On the other hand, the challenges that institutions face are related to talent, data, models, governance, and budget.

- Ten (10) financial institutions were further assessed to reveal the average maturity levels based on five (5) key verticals, as follows: Data, Models, and Tools with the highest maturity level at 2.1, next is Talent and Structure at 1.6, followed by Risk Management at 1.5, and Consumer Protection and Ethics and Governance at 1.1 and 0.9, respectively.

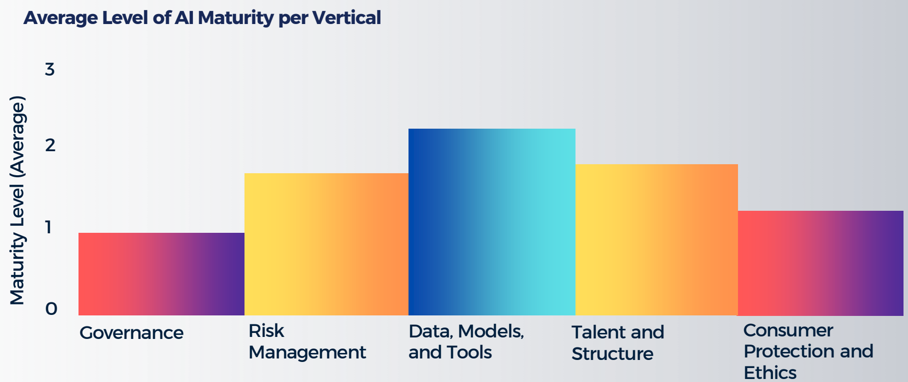
MAJOR USE CASES

- Next Best Offer/ Recommender System
- e-KYC/ Liveness check
- Fraud Management System
- Credit Risk Scoring
- Assistive AI/ GenAI

STATE OF AI ADOPTION



MATURITY ASSESSMENT

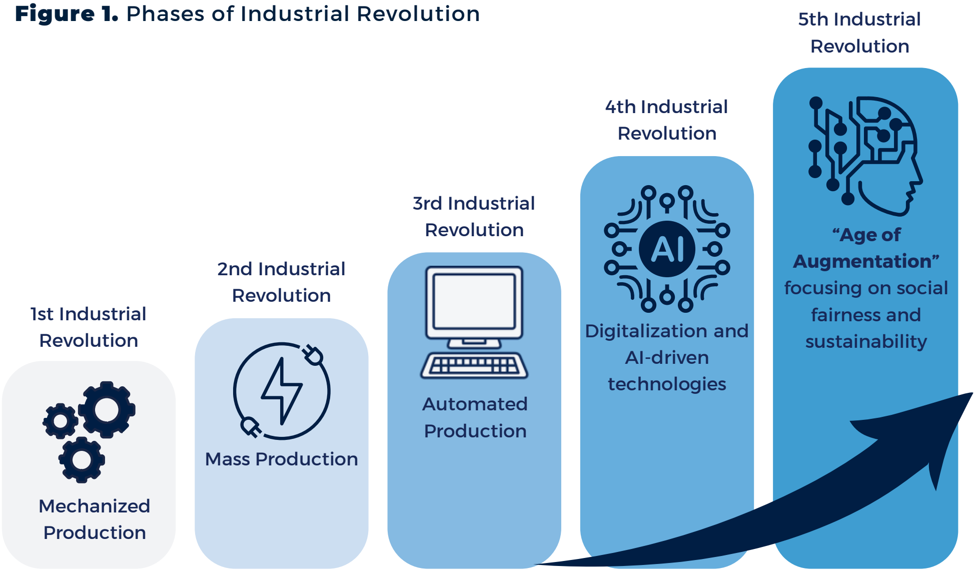


INTRODUCTION

Artificial Intelligence (AI) is considered one of the elements of the fourth industrial revolution, along with data and network, that could transform different industries and promote social progress. The World Economic Forum (WEF) described the latest revolution as "fusion of technologies" combining the physical, biological, and digital world (Schwab, 2016).

AI is now deeply embedded in the process of several industries. From mechanized production, we are now in the age of digitalization. Compared to the previous industrial revolutions which are technology-driven, the fifth industrial revolution is value-driven. Given the exponential rate of development in using AI, the fifth industrial revolution aims to be human-centric and sustainable.

Figure 1. Phases of Industrial Revolution



Sources: (Schwab, 2016) and (Xu, Lu, Vogel-Heuser, & Wang, 2021)

WHAT IS AI?

There is no uniform definition of AI. The Bank for International Settlements (BIS) claimed that AI "is a broad term, referring to computer systems performing tasks that require human-like intelligence". Different institutions defined AI with similar terms but not exactly the same.

BRITANNICA

AI is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings (Copeland, 2024).

FINANCIAL STABILITY BOARD (FSB)

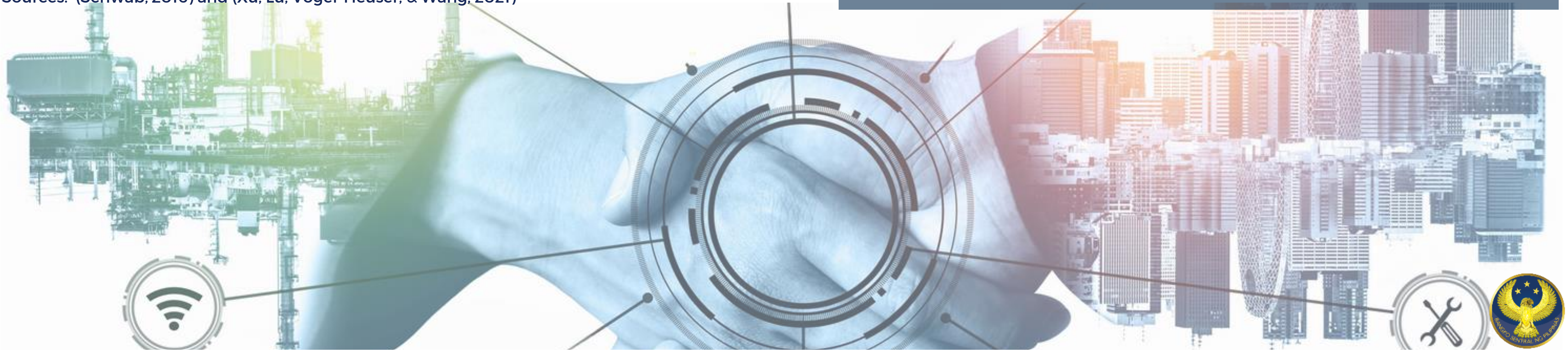
AI is the theory and development of computer systems able to perform tasks that traditionally have required human intelligence" (Financial Stability Board, 2017).

MONETARY AUTHORITY OF SINGAPORE (MAS)

AIDA or AI and Data Analytics are defined as technologies that assist or replace human decision-making" (Monetary Authority Of Singapore, 2018).

ASSOCIATION OF SOUTHEAST ASIAN NATIONS (ASEAN)

AI is the discipline of making analytical machines intelligent, enabling an organization to function appropriately and with foresight (Association of Southeast Asian Nations, 2024)



Focusing on the system rather than the field, more regulations are now using the term AI system. The European Union (EU) AI Act, the first comprehensive AI law enforced in August 2024, emphasizes the ability of an AI system to infer which makes it distinct from simpler automation using rules-based approach. While the lifecycle of AI system also starts with data processing and analysis, the capability to infer enables an AI system to learn, reason, or model patterns. Inference techniques include machine learning approaches[1] and logic- and knowledge-based approaches[2].

However, jurisdictions are converging to agree on the same definition to avoid the risk in regulatory arbitrage. To standardize the definition across its 38 member countries, the Organisation for Economic Co-operation and Development (OECD) released an updated definition of an AI system.

EUROPEAN UNION (EU) AI ACT

AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments (EU Artificial Intelligence Act, 2024).

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD)

AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

The term "Artificial Intelligence" can be traced back to 1950s, shortly after the invention of digital computers. In the '90s, rules-based decisions, personalization, and industrial robots were utilized. In early 2000s, bigger memory capacity and processing power made big data possible. This, in turn, opened the doors for natural language processing (NLP), deep learning, and image and voice recognition. Several Generative AI (GenAI) systems were released in recent years such as IBM's Watson, Jasper AI, and Google's Gemini (formerly Bard). In November 2022, OpenAI released the demo version of ChatGPT to the public and social media raved regarding how "human-like" the chatbot seems. The generated contents sound coherent on the surface, however, GenAI to date is still far from human intelligence and whether it can be sentient in the future is still up for debate.

"GenAI to date is still far from human intelligence and whether it can be sentient in the future is still up for debate.

This AI phenomenon is now evident across all industries. The financial services industry, being one of the richest sources of structured data, is an early adopter of AI. Financial institutions (FIs) can infer the hobbies and lifestyle of an individual based on their historical purchases and offer related promos to increase utilization. Institutions can also use an individual's payment patterns to determine the likelihood of default. Digital onboarding or electronic Know-Your-Customer (eKYC) process has also leveraged AI by using ID matching, face matching, and liveliness check. The use cases of AI are continuously evolving and the possibility is truly promising.

However, new technologies come with novel risks, and so does AI. According to OECD, many risks that are associated to AI can be classified as common risks being amplified by the use of AI (OECD, 2024). Examples of these are i) sophisticated phishing messages or deepfakes that could increase fraud or scam cases, ii) traditional herding effects in stock markets stimulated by AI systems, and iii) AI-powered cyberattacks[3].

Needless to say, the proliferation of AI should also be accompanied by proportionate risk management frameworks. Although general risks are covered by existing laws and regulations, there are gaps identified globally and locally. It is prudent for the BSP to monitor its impact on financial services and provide supervisory expectations to institutions that want to delve into AI.

[1] Machine Learning is a subset and application of AI, which focuses on the development of computer programs - designed to learn from experience without being explicitly programmed to do so. (Adapted from: Financial Stability Board:2017.)
[2] Logic-based and knowledge-based approaches infer from encoded knowledge or symbolic representation of the task to be solved. (Adapted from: EU AI Act: 2024.)
[3] Examples are AI-driven DDoS attack, intelligent self-learning malware, and password guessing/cracking (Guembe, et al., 2022).

AI POLICIES AND RISK MANAGEMENT FRAMEWORKS

There are a lot of similarities in AI regulations around the globe, however, there are still variations in terms of approach and focus. Several jurisdictions found it necessary to introduce binding/hard legislation on the responsible use of AI while some countries have opted for non-binding/soft legislation that does not penalize or mandate a specific approach to AI. Since the field is still developing, non-binding/soft legislation is being preferred to cover potential advances in the area.

Regulations around the world may vary but the risks identified with the use of AI are consistent. The annual report of the US Financial Stability and Oversight Council (FSOC) enumerates the possible risks in the area which include cyber risks, model risks, and consumer protection considerations (e.g., fair lending) (FSOC, 2023). In any regulatory approach, FSOC recommended two (2) things to ensure financial stability: i) continuous monitoring due to the emerging risks to the financial system that come with the rapid developments in the AI field; and ii) building expertise and capacity to support the said activity.

Binding/Hard Legislations

Non-Binding/Soft Legislations [4]

Main Source: (OECD, 2024)

PHILIPPINE AI REGULATIONS

National AI Strategy Roadmap (NAISR) 2.0

On 5 May 2021, the Department of Trade and Industries (DTI) released the country's AI Roadmap covering important dimensions for AI readiness. It was updated in July 2024 and outlines seven (7) strategic imperatives such as: i) establishing a robust, connected, and networked environment; ii) improving data access and value extraction; iii) transforming education and nurturing future AI talent; iv) upskilling and reskilling the workforce; v) fostering an AI ecosystem that prioritizes ethical considerations; vi) mastering and pushing the boundaries of AI research and development; and vii) accelerating innovation through the strategic application of AI technologies (DTI, 2024). DTI also launched the Center for AI Research (CAIR) which is the first AI hub of the country.

House Bill No. 10457 or the "An Act Establishing a National Strategy for the Development of Artificial Intelligence And Related Technologies"

DTI's National Strategy for AI is also supported by HB No. 10457 that promotes the establishment of the National Center for Artificial Intelligence Research (NCAIR) which is intended to be "the primary policymaking and research body concerned with the development of artificial intelligence and allied emergent technologies in the country" (House Bill No. 10457, 18th Congress of the Republic, 2021).

House Bill No. 7396 or the "Artificial Intelligence Development and Regulation Act of the Philippines"

The proposed law aims to support the development of AI to boost the economy. It also seeks to establish an AI Development Authority (AIDA) to oversee the developments and deployments that will be done through AI (House Bill No. 7396, 19th Congress, 2023). The authority is expected to provide risk assessments and standards to ensure consumer protection while enhancing cybersecurity (Quismorio, 2023).

[4] Includes supervisory expectations, positions, principles, guidelines, white paper, recommendations, code of conduct, blueprint, policy report, or executive order.

AI-RELATED
LOCAL LAWS

While there is no specific law for AI yet, there are laws in the Philippines that cover basic principles such as:

Existing Regulations	Description
Republic Act (R.A.) 10173 or the Data Privacy Act of 2012 (DPA)	In 2012, under the DPA, processing of sensitive and personal information and privileged information is prohibited unless a consent is given or under rare cases as stated. The law also covers the rights of the data subject as well as the penalty for violations.
R.A. 8293 or the Intellectual Property Code (IPC)	In 1997, IPC was signed into law which aims to protect and give exclusive rights to intellectual properties that may relate to a product, process, or an improvement.
RA 7394 or the Consumer Act of the Philippines	Signed into law back in 1992 with focus on protecting the interest and general welfare of consumers. The said law also establishes standards of conduct for business and industry.
RA 11765 or the Financial Products and Services Consumer Protection Act (FCPA)	The FCPA focuses on affording more protection to consumers of financial products and services with specific sections on privacy and protection of client data and information security standards.

Following a lot of multi-sectoral consultations and continuous collaboration across the globe, a set of principles for an effective AI governance framework was gathered (Personal Data Protection Commission (PDPC) & Info-communications Media Development Authority (IMDA), 2020). These principles are used to cross-check for the common AI principles already covered by existing BSP regulations.

While it can be argued that basic risks in AI are covered by the existing rules and regulations, it is apparent that some risks are still left untouched such as accuracy, data quality, and ethical issues focusing on transparency and explainability, fairness and bias, and hallucination. These risks are very specific to the field of AI and it would be appropriate to address them in a directed manner. This policy gap, if remains unchecked, can amplify the risks, and affect not only the reputation of the institution but also public's trust in technologies.

AI-RELATED
BSP REGULATIONS/BILLS

Circular	Principles Covered	Guidelines Discussed
Circular No. 808 **IT Risk Management**	• Accountability • Auditability • Robustness • Security • Transparency	• IT Audit • Information • Security • Project Management • Development Acquisition • Change Management
Circular No. 982 **Information Security Management**	• Accountability • Auditability • Robustness • Security • Transparency	• Information Security Management
Circular No. 1160/1048 **Financial Consumer Protection**	• Human rights alignment • Fairness • Transparency	• FCPA implementation in financial services
Circular No. 1140 **Fraud Management Systems and Consumer Education and Awareness**	• Fairness • Transparency • Explainability	• Electronic Products and Services (EPFS) • Consumer Awareness • Fraud Protection
Circular No. 989 **Guidelines on the Conduct of Stress Testing Exercises**	• Robustness and Security • Transparency	• Overarching governance standards • Risk management expectations on stress testing practices
Circular No. 971 **Guidelines on Risk Governance**	• Accountability • Auditability • Robustness and Security • Transparency	• Enhances corporate governance guidelines • Effective challenge process by independent and competent parties
Circular No. 900 **Guidelines on Operational Risk Management**	• Accountability • Auditability	• Identify and assess operational risks
Circular No. 855 **Guidelines on Sound Credit Risk Management Practices**	• Accountability • Auditability • Transparency	• Specific for credit-related methodologies

PROJECT SAPIENS: THEMATIC REVIEW ON

AI AND MACHINE LEARNING

In 2024, the BSP started to assess the landscape of AI in the local setting following the recommendations of international bodies to continuously monitor emerging risks concerning new technologies. To fully comprehend the regulatory gap or overlap within the realm of AI, the BSP deemed it necessary to undertake this problem through a collaborative and data-driven approach by conducting a thematic review.

OBJECTIVES

01 To determine the extent of AI adoption in the financial sector

02 To assess the readiness and maturity of the financial sector

03 To identify the benefits and challenges in AI development and implementation

04 To gather more information on existing and planned AI use cases

SCOPE

Traditional AI systems are being used by institutions for more than a decade. To clearly identify the scope, the study used the OECD definition of AI system to determine the coverage of discussion. Moreover, the assessment focused on the gaps and risks between the existing regulations and the current implementation and process related to AI of financial institutions in the industry. While not all financial institutions were included in the phase two – deep dive sessions, the chosen participants sufficiently represented the state of AI adoption in the financial industry.

METHODOLOGY

PHASE ONE: AI Adoption Survey

The team deployed an online AI adoption survey from 08 April to 03 May 2024 to 51 BSP-supervised financial institutions with questions related to the nature of their AI projects and initiatives. The chosen survey respondents are considered as top players in their area of business and/or technology-heavy institutions. Out of the 51 BSP-supervised financial institutions (BSF) who were invited to participate, 48 institutions (94%) responded.

PHASE TWO: AI Deep-dive

The team conducted walkthroughs of AI system deployment and implementation to deep dive on the current processes and best practices in the industry. Institutions that are either heavy users of AI, with interesting use cases, or considered as Domestic Systemically Important Bank (DSIB), including digital banks, are selected for Phase Two of the thematic review.

STATE OF AI ADOPTION

IN THE PHILIPPINE FINANCIAL INDUSTRY

Out of the 51 BSP-supervised financial institutions (BSFI) who were invited to participate, 48 institutions (94%) have responded.

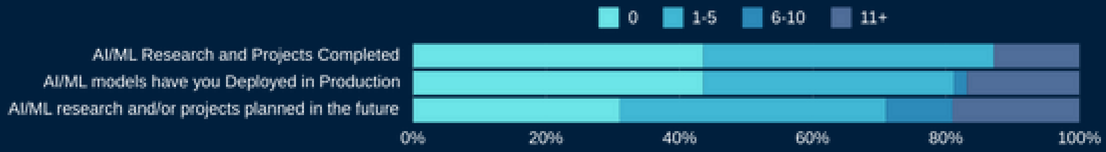
60% institutions explicitly **include** AI/ML in their present **roadmap**

10% Ranked advanced AI/ML as their **top priority**

33% have a **dedicated or centralized unit** for AI-related projects and systems

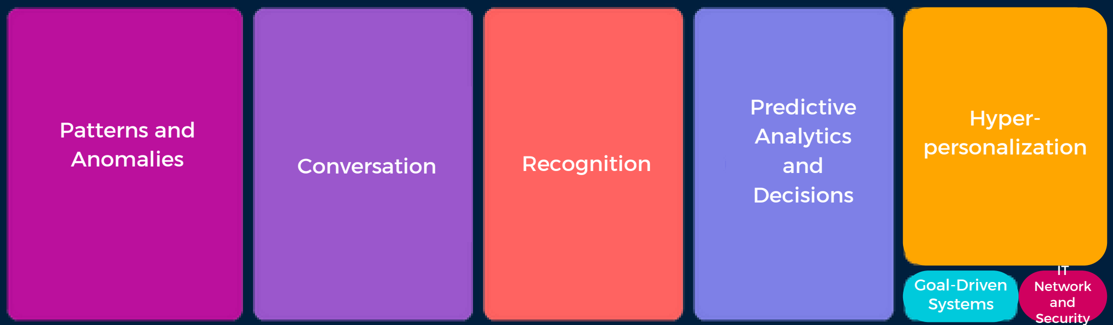
56% respondents have already deployed AI/ML models in production.

AI ADOPTION OF BSP-SUPERVISED FINANCIAL INSITUTIONS



USE CASES

Top use case is **Patterns and Anomalies**, which is usually used in Fraud and Anti-Money Laundering systems. Use cases that can be classified as Conversation, Recognition, Predictive Analytics, and Hyper-personalization are also widely used in financial services.



BENEFITS

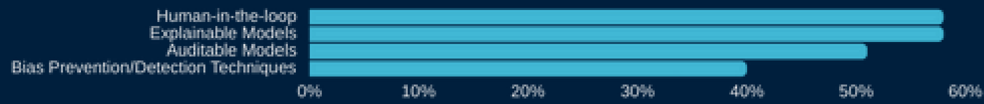
mostly used words to describe the benefits they reap when they started using AI/ML in their institutions:

CHALLENGES

mostly used words to describe the challenges they encounter when they started using AI/ML in their institutions:

MODEL CHARACTERISTICS

Majority of the institutions considers their models to be auditable, explainable, and with integrated human-in-the-loop mechanism.



Institutions are using external AI/ML models developed and/or maintained by **Outsourced Service Providers (OSPs)**

52%
25 out of 48

Institutions **anticipated a revenue increase** of more than 5% of their annual budget due to the adoption of digital technologies

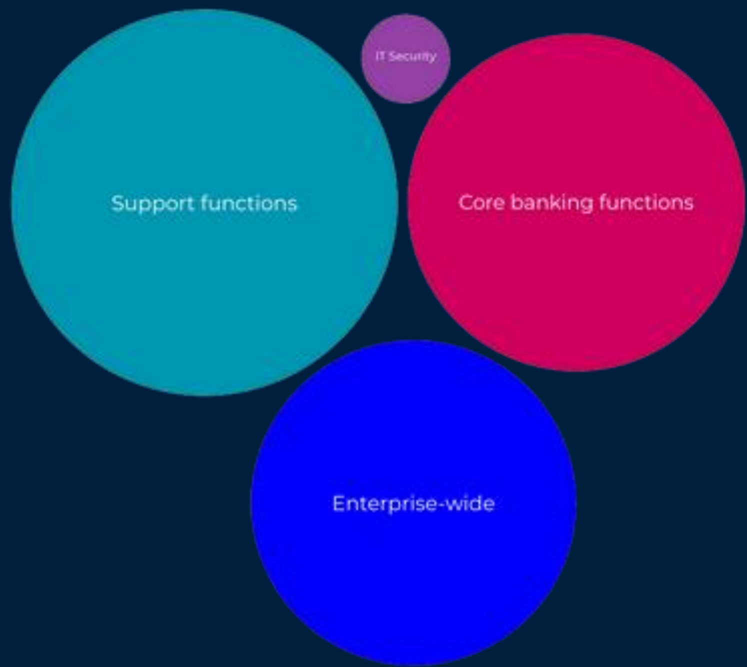
46%
22 out of 48

40%
19 out of 48

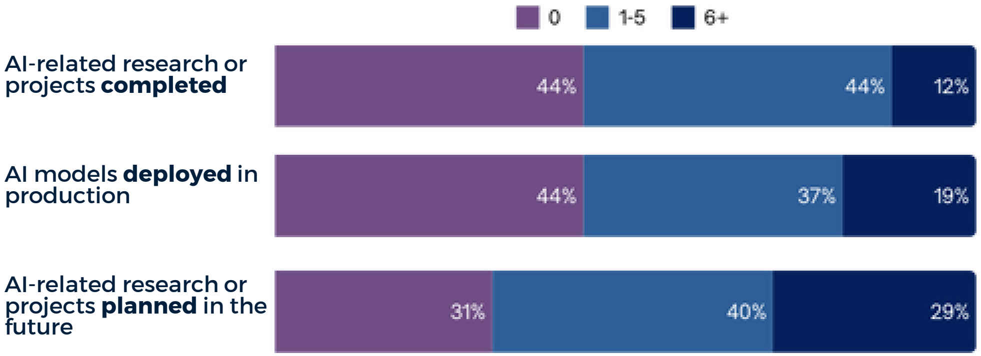
Institutions **anticipated a cost savings** of more than 5% of their annual budget due to the adoption of digital technologies

BUSINESS PROCESS SUPPORTED

AI initiatives mainly focus on support functions followed by core banking functions and enterprise-wide involvement.



Phase 1: STATE OF AI ADOPTION
IN THE PHILIPPINE FINANCIAL SERVICES



The initial results of the survey proved to be valuable in understanding the current AI landscape of the country. Out of 51 target respondents, 48 respondents (94%) participated in the survey as of 24 May 2024.

Top use case is "Patterns and Anomalies" which is usually used in Fraud and Anti-Money Laundering systems. Specific use cases under this category include suspicious transaction monitoring and network security analytics. AI has improved the ability of financial institutions to detect fraudulent transactions more swiftly and accurately. Further, 46 percent (46%) of the respondents identified "Conversation" as the next most utilized type of use case in their organization which is mainly used in customer service and sentiment analysis initiatives. Use cases that can be classified as Recognition, Predictive Analytics, and Hyper-personalization are also widely used in financial services.

The survey showed extreme sides on the state of AI adoption where 21 (44%) participants indicated that they have not completed any research or project related to AI/ML, 21 (44%) have completed at least one, and six (12%) have completed more than 10. Meanwhile, 29 FIs (61%) explicitly included AI/ML in their institutions' roadmap, 15 FIs (31%) explicitly excluded AI/ML in their institutions' roadmap, and four (4) financial institutions (8%) were not sure.

OTHER SURVEY RESULTS

60% AI is explicitly included in the institution's present **roadmap**

10% Ranked advanced AI/ML as their **top priority**

33% have a **dedicated or centralized unit** for AI/ML-related projects and systems

56% have already deployed AI/ML system in **production**

52% are using **external AI system** developed and/or maintained by outsourced service providers

46% anticipate an **increase in revenue** of more than 5% of their annual budget due to AI adoption

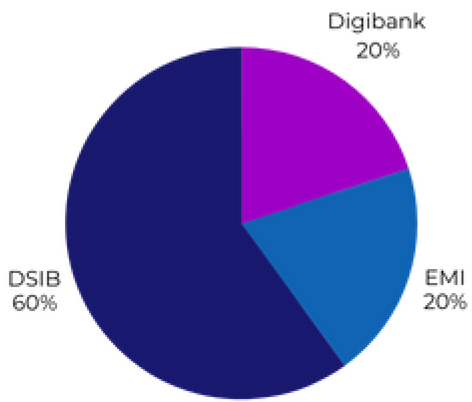
Phase 2: AI Deep-dive

Ten (10) financial institutions were invited for the deep dive sessions discussing the following

i) present and future planned use cases
ii) talent and organizational structure for AI-related activities
iii) process overview and risk management
iv) consumer protection and ethics
v) AI governance
vi) learnings in the field

This phase aims to know more about their current process, implementation and use of AI, the pertinent risks, best practices, and the challenges faced in the industry.

Before crafting any strategy or governance framework, it is crucial to determine the current maturity of the institution to effectively analyze the road ahead. Hence, during the second phase of the thematic review, the team focused on how the selected institutions developed and implemented their AI systems. This phase highlighted the best practices and the challenges that institutions are facing in the area.



MATRIX ASSESSMENT SUMMARY

Governance	Risk Management	Data, Models, and Tools	Talent and Structure	Consumer Protection and Ethics
• AI Roadmap • AI Policy • Cybersecurity • Data Security and Privacy • Outsourcing Process	• Risk-Based Assessment • Human-in-the-Loop Technique • Bias Detection and Mitigation • Model Drift	• Data Management and Architecture • Tools and Platform • AI System Lifecycle	• Organizational structure • Workforce Development • Training Plan	• Consumer protection mechanisms • Explainability and Transparency • Ethics and Discrimination

Five (5) verticals were used to assess the overall maturity of the institutions in the current state of their AI systems: 1) Talent and Structure, 2) Governance, 3) Data, Process, and Tools, 4) Risk Management, And 5) Consumer Protection and Ethics. Each vertical has several topics under it. The maturity of institutions in AI/ML development and adoption was assessed by using an internal matrix developed by the BSP review team[5].

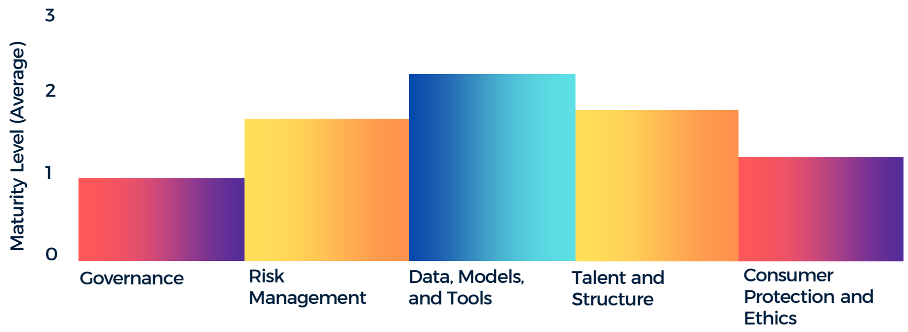
LEVELS OF MATURITY

- 0 No usage
- 1 Primitive/Manual Process - Initial adoption
- 2 Automated Systems - Defined and Managed
- 3 Advanced Systems - Optimized

Financial institutions were evaluated based on their current state of AI adoption in four (4) stages that can be generally described above. The criteria from previous stages should be satisfied before graduating to the next stage. The matrix serves as a guide for reviewers to assess the maturity level, systematically and objectively, across all institutions. By this classification method, the result of the review can be easily digested to reflect the holistic landscape of the Philippine financial system regarding AI.

MATURITY ASSESSMENT

Average Level of AI Maturity per Vertical



The maturity assessment summary reflects that most institutions have started to invest in data, talent and tools to jumpstart their AI journey. However, the AI governance aspect is lagging with almost all of them not having a policy specifically designed to assess or tackle AI risks. The reviewed institutions' AI journey has already started but the approach and maturity differ.

Organizations tend to incorporate AI in their current structure by choosing a setup that will introduce minimal changes. There are still institutions that are lacking formal governance structures leading to potential inconsistencies in oversight.

Workforce development appears to be sufficient with no reported job displacement due to AI. One major challenge, not just in AI but also in the whole technological field, is talent shortage.

Most institutions are still considered conservative in terms of their risk appetite towards AI. These institutions stated their intentional disinclination regarding customer-facing or fully automated AI to keep the risks manageable. This strategy may also be a mindful way to keep AI-related laws in practice.

Although there are a lot of questions concerning AI ethics, some institutions have initiated projects and activities addressing issues of explainability and fairness.

Quantifying the benefits of AI in the business is also another matter that institutions have yet to figure out.

[5] The thematic team have developed the framework by incorporating relevant criteria from the MITRE AI Maturity Model (Bloedorn, et al., 2023), the DELTA Plus Model and Five Stages of Analytics Maturity (Davenport & Harris, 2017), and the ASEAN Guide on AI Governance and Ethics (Association of Southeast Asian Nations, 2024).

MATRIX ASSESSMENT SUMMARY

USE CASES

Reviewed Use Cases	Cross-selling/ Next Best Offer/ Recommender Systems	Digital Onboarding/ e-KYC/ Liveness Check	Fraud Management System	Gen AI/ Assistive AI	Credit Risk Scoring
PARTICIPANT 1		✔	✔	✔	✔
PARTICIPANT 2	✔	✔	✔	☑	✔
PARTICIPANT 3	✔	✔	✔	✔	✔
PARTICIPANT 4	✔	✔	✔	☑	☑
PARTICIPANT 5		✔	✔	☑	
PARTICIPANT 6	☑	✔		✔	☑
PARTICIPANT 7	✔	✔	✔	✔	✔
PARTICIPANT 8		✔	✔	☑	✔
PARTICIPANT 9		✔	✔		
PARTICIPANT 10	✔	✔	✔	☑	✔

Legend: ✔ use cases are utilized; ☑ use cases under development or in progress

AI is deeply embedded in the reviewed institutions processes. The reviewed financial institutions have common use cases. However, some of them have a few distinct features than the others.

There are several institutions developing or enhancing their credit decisioning and recommendation engine through AI. Some institutions are planning to launch these use cases in a customer-facing system. There were also observed use cases in forecasting/analysis to optimize banking assets. For cross-selling and next-best offer use cases, some institutions are sticking to the traditional way of leads generation. The list of customers who are likely to avail loan offers is generated and submitted to marketing people for targeted campaigns. However, for sophisticated players, product advertisements are already embedded in their mobile app/website by integrating AI recommender systems to generate fresh marketing offers.

Digital onboarding process integrates AI in their e-KYC (Know-Your-Customers), facial or ID matching, and liveness check. These AI use cases are usually integrated into tools provided by vendors or outsourced service providers (OSPs).

Outsourcing is also popular for fraud management systems with built-in fraud risk score. It is important to note that the underlying models from these outsourced tools could provide a different perspective from the limited data that institutions have. Using external data, third-party AI systems offer a proactive approach to targeting unusual patterns or customer behavior.

Third-party AI systems also offer the capability to provide a more holistic view. This is the case for credit risk models developed by OSPs or credit bureaus. There are ongoing Proof-of-Concepts (POCs) to explore using alternative data (i.e., cash flow, mobile app activity) to target the unbanked or underserved segment of the society. While there is a continuous search for alternative data, institutions remain skeptic on the use of social media data for credit scores.

Targeting the unbanked or thin file segment[6] is a persistent issue in credit scoring that hinders financial inclusion. Hence, the use of alternative data to prove the creditworthiness of an unbanked individual can be a great contribution of AI in this field.

Since the launch of ChatGPT, organizations can't wait to unleash the potential of GenAI that has been the topic of several talks in the financial industry field. However, to date, the GenAI use cases of institutions remain limited to chatbots and assistants.

Overall, the use cases vary and are prioritized based on business objectives and business models. Maturity cannot be attributed to sophisticated use cases alone, but to the institutional understanding and use of AI, its potential, its risks, and the management thereof.

Controls for AI systems are similar to the established Software Development Life Cycle (SDLC), except for the model validation and human oversight aspects.

AI model validation procedures, both during the development stage and as a regular monitoring activity, are being done by majority of the institutions. Majority of the institutions also perform AI-specific test activities on their AI systems, on top of the regular test activities. Common exercises include AI back-testing, stress testing, and discriminatory or bias testing.

In these early stages of AI systems in the financial sector, human oversight is an essential element to manage both known and unknown AI risks. However, only three (3) institutions have established comprehensive and well-integrated human oversight mechanisms, involving manual checks and validations throughout the model lifecycle.

Other controls and supervisory expectations will be discussed and clarified in the following sections.

[6] Thin file segment refers to unbanked customers which are applying for a credit product but has no historical credit data. Hence, the only data that could be used to infer the likelihood of a customer to default is their application information. Thick file segment refers to customers which already have ample historical credit data.

GOVERNANCE

AI Roadmap ensures that AI initiatives are aligned with business objectives, supports efficient resource allocation, and provides a structure for tracking progress and adapting to technological changes.

AI Policy pertains to the overall plan by which the Board, Senior Management, owners, and users of AI systems agree while ensuring alignment with ethical principles, legal requirements, and the institution's organizational culture and values.

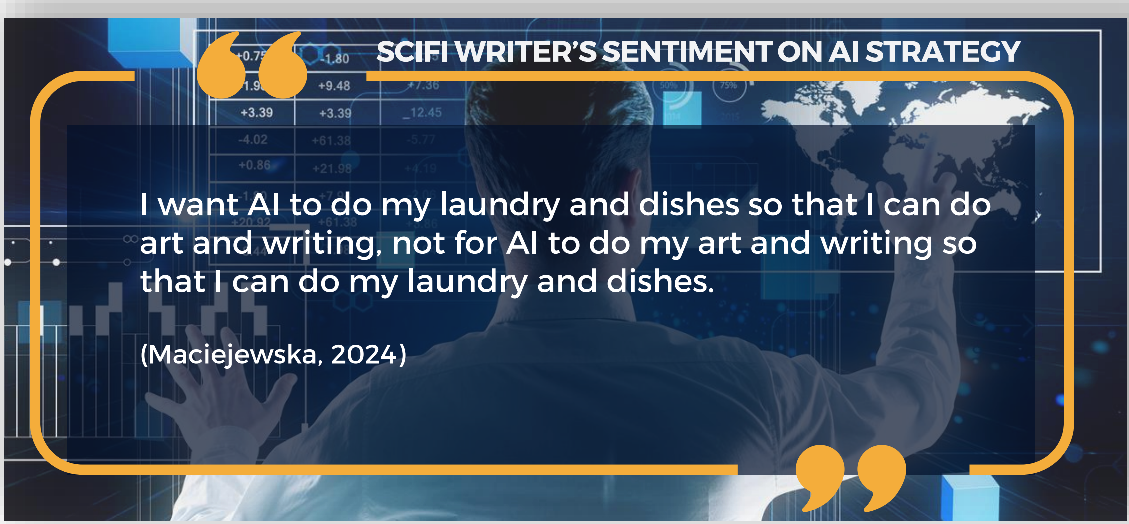
Budget Allocation reflects an institution's investment in AI, encompassing resources dedicated to AI infrastructure, talent development, and ongoing project costs.

Outsourcing in AI systems involves leveraging third-party vendors to develop, manage, or maintain AI systems

Cybersecurity and Privacy for AI systems is crucial to safeguard against threats that could compromise data integrity, confidentiality, and system availability.

Cognizant of the need to enhance trust and transparency among consumers and stakeholders, institutions are increasingly developing and adopting comprehensive AI governance frameworks that align with existing regulations and internationally accepted recommendations and standards. Among the 10 institutions, only one (1) has demonstrated Stage 2 maturity in AI governance. This means that most participants have no clear and working AI governance in place. As a result, existing and ongoing AI initiatives may lack development, deployment, maintenance, and usage guidelines.

Other challenges remain, such as the lack of formal governance structures or comprehensive AI policies. Some institutions have recently approved AI policies, but their implementation and effectiveness are still not fully assessed.



AI ROADMAP

- Risks abound when AI roadmaps are absent or misaligned with broader organizational objectives. This misalignment can lead to poor resource allocation, inefficient AI deployments, and missed growth opportunities. Additionally, the lack of cross-functional involvement poses a significant threat. When AI initiatives are developed without input from key areas such as compliance, risk management, or operations, they may fail to align with broader business goals, leading to inefficiencies and an inability to scale AI efforts effectively.

- Two (2) institutions were able to develop comprehensive AI roadmaps extending beyond 2024, showcasing a strategic focus on AI integration. However, two (2) Fis have not integrated AI into their strategic roadmaps, leading to ad hoc AI initiatives that could result in resource wastage. Additionally, in some institutions, AI initiatives are developed in isolation. This siloed approach can cause inefficiencies and difficulties in effectively scaling AI efforts.

- Across the Philippine financial industry, smaller institutions like rural banks face significant challenges in adopting AI technologies due to resource constraints and heavy reliance on third-party providers. This reliance amplifies risks related to outsourcing, including data security breaches, regulatory non-compliance, and inconsistent quality of AI solutions. Larger institutions, meanwhile, grapple with maintaining flexible AI roadmaps that can quickly adapt to evolving technologies and market demands.

SUPERVISORY EXPECTATION ON AI ROADMAP

In such cases that institutions plan to adopt AI, the BSP expects that these plans are integrated in roadmaps within their overall business and IT strategies to ensure alignment with long-term objectives. Sound technology risk management and strategic planning of institutions include:

- Developing comprehensive AI roadmaps that outline their strategic plans for AI adoption, including clear goals, timelines, and resource allocations, especially for high adopters of AI.
- Regularly updating AI roadmaps to reflect technological advancements and changing market conditions and are reviewed and vetted by senior management and the board.
- Involving various departments—such as compliance, risk management, and operations—in the development and implementation of AI roadmaps to promote cohesive and effective AI integration.
- Tracking and reporting the return on investments (ROI) of significant AI investments to ensure that resources are used efficiently.

Refer to:
i. BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions
ii. BSP Circular No. 971 dated 22 August 2017 on the Guidelines on Risk Governance on the enhanced requirements on risk management systems of institutions.

BUDGET ALLOCATION

- In theory, AI creates value in one of three ways: 1) by reducing costs, 2) by increasing revenue, or 3) by enabling new business opportunities (United Nations Industrial Development Organization, 2021). However, some institutions are still struggling to identify or quantify the benefits of using AI. Though some entities are willing to wait to reap the rewards, others are not.

- Smaller institutions often face underinvestment in AI, leading to underperforming AI systems, failed implementations, and an inability to compete with larger, better-funded institutions. Conversely, larger institutions risk overinvesting in AI technologies without a clear ROI analysis, which can result in wasted resources and strained budgets. Additionally, failing to invest in AI talent creates a significant skills gap, forcing institutions to rely on external solutions. This reliance limits their ability to customize AI systems effectively, with smaller institutions particularly vulnerable due to their limited ability to attract and retain skilled professionals.

AI POLICY

- Approach to emerging technologies, specifically for AI, can be classified into three: 1) top to bottom, 2) bottom-up, or 3) mixed approach. The top-to-bottom approach, usually observed from established institutions, means that the management gives enough power and resources to developers to be able to explore use cases. However, ROIs to these exploratory projects are still not evident in the early stages of implementation. Bottom-up approach is more evident from experienced or digitally oriented institutions where talents could provide POCs and steer the organization's AI path. Projects from this approach are more intentional since the development comes from a business need. A mix of both appears to be more beneficial.

- Some challenges observed include the absence of formal AI policies in other institutions which hampers consistent and ethical AI implementation. Resource constraints in smaller institutions may limit their ability to develop and maintain robust AI policies, increasing the risk of ethical issues and non-compliance.

- Institutions without formal AI policies might violate existing laws. Inconsistent AI oversight could lead to unlawful use of customer data, resulting in regulatory sanctions and reputational damage.

SUPERVISORY EXPECTATION ON AI POLICY

The BSP recognizes that institutions are at different maturity levels in terms of AI governance and management. Nonetheless, the BSP expects institutions to put in place a proportionate AI policy framework depending on its risk appetite and set out fundamental controls pertaining to the management of AI systems. Institutions are also expected to establish comprehensive AI policies that govern the ethical, transparent, and accountable use of AI technologies ensuring the following:

- Compliance with all relevant laws and regulations, particularly those related to data privacy, cybersecurity, and consumer protection.

- When using AI heavily, institutions should cover the responsible use of the technology in their internal policies which can be enterprise-wide or IT-specific. This would ensure that there is accountability and auditability embedded in the process.

- Establishing clear lines of authority and accountability for AI-related activities to enhance oversight and reduce operational risks.

- The board and risk management committee should be able to effectively challenge the performance of AI systems. Business metrics for AI system monitoring should be comprehensive enough to cover its overall output, impact, and performance.

- Institutions should keep proper documentation of their AI system processes together with any change, modification, or decision that were made during its lifecycle to promote proper change management.

Refer to:
i. BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions
ii. BSP Circular No. 900 dated 18 January 2016 on the Guidelines on Operational Risk Management
iii. BSP Circular No. 971 dated 22 August 2017 on the Guidelines on Risk Governance on the enhanced requirements on risk management systems of institutions; and
iv. BSP Circular No. 989 dated 04 January 2018 on the Guidelines on the Conduct of Stress Testing Exercises provides the overarching governance standards and risk management expectations on stress testing practices in the banking industry, including the supervisory expectations on the use of models for stress testing.
v. BSP Circular No. 1160 dated 28 November 2022 on the Regulations on Financial Consumer Protection to implement Republic Act No. 11765, otherwise known as the "Financial Products and Services Consumer Protection Act"

CYBERSECURITY MEASURES

- A key risk lies in exposure to AI-specific cyber threats, such as adversarial attacks and data poisoning[7]. Institutions without AI-specific cybersecurity measures are particularly vulnerable. This risk is further magnified when AI systems handle sensitive customer data without proper security controls, making institutions susceptible to both regulatory and legal consequences. The lack of regular, rigorous independent security testing adds to this vulnerability, as undetected system flaws can lead to significant breaches.

- Vulnerabilities include a lack of AI-specific testing and evaluation standards, leaving AI systems potentially exposed to unique threats such as adversarial attacks. The evolving threat landscape introduces vulnerabilities that existing cybersecurity frameworks may not fully address, requiring continuous updates and specialized expertise. Smaller banks may lack the technical expertise and resources to implement advanced cybersecurity measures, increasing their risk exposure to cyber threats.

- Institutions exemplify best practices by implementing robust security measures, including strict access controls and encryption, to protect AI systems from threats, similar to any other system around it. Conducting regular security assessments on AI systems, such as Vulnerability Assessment and Penetration Testing (VAPT), helps identify and mitigate potential vulnerabilities. By following standardized security protocols and adhering to guidelines like those from the National Institute of Standards and Technology (NIST), institutions establish multiple layers of security controls for comprehensive protection.

SUPERVISORY EXPECTATION ON CYBERSECURITY

Depending on the AI system risk-based assessment, institutions must implement robust cybersecurity frameworks specifically addressing the unique risks associated with AI systems with the following expectations:

- Develop and implement **cybersecurity defense mechanism** that address AI-specific threats such as adversarial attacks and data poisoning.
- Conduct regular **security testing** of AI systems, including vulnerability assessments and penetration testing.
- Establish **protocols for** timely identifying and reporting of **AI-related security incidents** to regulators. Incident Response procedures should also apply to AI-related security breaches that may require specialized approach.
- Employ robust **data protection strategies**, including encryption, anonymization, and access controls. Establish appropriate access controls at different levels of the AI systems and datasets by restricting access only to authorized personnel.

Refer to:
i. BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions
ii. BSP Circular No. 1019 dated 31 October 2018 on the Technology and Cyber-Risk Reporting and Notification Requirements

[7] Data poisoning intends to inject special samples to the training dataset to influence the result of the AI model. Adversarial attacks are generally techniques on how to manipulate AI outputs. (Adapted from: Boukherouaa, El Bachir, Khaled AlAjmi, Jose Deodoro, Aquiles Farias, and Rangachary Ravikumar. 2021. "Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance." Departmental Papers 2021 (024).

DATA SECURITY AND PRIVACY

- All reviewed financial institutions rely on general information security and privacy policies rather than AI-specific standards. While measures like role-based access controls, stringent controls on data access, anonymization protocols, and minimizing the use of Personally Identifiable Information (PII) are being implemented, these approaches may not fully account for the specific risks associated with AI technologies such as data poisoning, data leakage during inference, evasion, and model extraction which are more likely to be done by trusted insiders (U.S. Department of the Treasury, 2024). Furthermore, there are also an increased threat of AI-enabled attacks or "Adversarial AI" which include misinformation and deepfakes (U.S. Department of the Treasury, 2024).

- Established banks integrate privacy considerations from the design phase of AI systems, ensuring data protection is embedded[8] throughout the AI system lifecycle—a concept known as "Privacy-by-design".

- Reliance on third-party vendors for AI solutions introduces risks if vendors do not adhere to the same privacy standards, potentially leading to data breaches or regulatory breaches.

- Data minimization [8] presents a challenge in i) building enough training data to explore use cases or generate insights and ii) implementing bias detection and mitigation techniques. In some use cases, balancing data utility and privacy presents difficulties. When most customers opt out on data collection and processing, training data will be scarce. This will result in a potential decline in AI system accuracy or even feasibility.

- Best practices in security and privacy involve employing data protection measures such as encryption, data minimization, and regular privacy impact assessments to safeguard sensitive data. Compliance with data protection laws, including adherence to the Data Privacy Act of 2012 and other relevant regulations, demonstrates a commitment to legal compliance and customer trust.

- Advanced anonymization techniques are being explored to protect personally identifiable information (PII) during AI processing, enhancing data security.

SUPERVISORY EXPECTATION ON DATA SECURITY AND PRIVACY

Institutions should conduct a mandatory Privacy Impact Assessment (PIA) for all AI systems that handle personal data. Following this, a risk analysis should be performed to evaluate potential privacy risks to individuals arising from the system's operations. Institutions must develop mitigation strategies to minimize identified risks effectively. Additionally, it is crucial to conduct a compliance verification to ensure that all aspects of the AI system conform to the Data Privacy Act of 2012. Institutions should also assess their current terms and conditions, processes, and systems for compliance with National Privacy Commission (NPC)'s Guidelines on Deceptive Design Patterns.

Refer to:
i. Republic Act No. 10173 or the Data Privacy Act of 2012 (DPA)
ii. BSP Circular No. 989 dated 04 January 2018 on the Guidelines on the Conduct of Stress Testing Exercises
iii. BSP Memorandum No. M-2014-032 on the Guidelines on the Electronic Submission of the Stress Testing Reports
iv. BSP Memorandum No. M-2021-043 on subject: Data Breach Prevention and Control
v. BSP Memorandum No. M-2024-019 on the subject: Reminders on the Handling of Personally Identifiable Information (PII) and Other Sensitive Data

OUTSOURCING

- All reviewed financial institutions deal with outsourced service providers (OSPs) for AI-related initiatives mainly for i) data storage, ii) cybersecurity, iii) analytics tools, and iv) advanced models (e.g. GenAI). Four (4) financial institutions have developed a structured approach to outsourcing with clear contractual protections, such as Non-Disclosure Agreements (NDAs), right-to-audit clauses, and strict Service Level Agreements (SLAs), to safeguard against potential risks.

- Best practices in outsourcing AI systems involve rigorous vendor evaluation involving multiple departments before integrating third-party AI solutions. Some institutions include clear contractual protections, such as Non-Disclosure Agreements (NDAs), right-to-audit clauses, and strict Service Level Agreements (SLAs), to safeguard against potential risks. Ongoing monitoring and governance also ensure that vendors align with institutional policies and regulatory requirements, with regular performance evaluations to maintain standards.

- Strategic use of external expertise allows institutions to leverage specialized skills for specific AI functionalities (e.g., eKYC and fraud detection) while maintaining control over critical aspects of AI deployment.

- Observed challenges include ensuring control over data, and compliance when third parties have access to sensitive information. Moreover, dependence on vendors' stability and performance may introduce operational risks if vendors fail to deliver as expected or cause service disruptions. The lack of formal governance structures, especially in smaller institutions, to manage and evaluate third-party service providers can lead to inconsistencies in AI implementation and difficulties in adapting solutions to specific institutional needs.

SUPERVISORY EXPECTATION ON OUTSOURCING

When outsourcing AI systems, institutions must effectively manage third-party risks to ensure security and regulatory compliance, as outlined in BSP Circular No. 1137 on outsourcing guidelines. The perceived heightened risks of third-party dependency and market concentration can be mitigated as follows:

- Conducting comprehensive evaluations of third-party vendors' capabilities, security practices, and compliance with legal requirements.

- Including clauses in vendor contracts that mandate adherence to Philippine laws and BSP regulations, including data protection and cybersecurity standards.

- Establishing governance structures to oversee outsourced AI services, including regular performance evaluations and audits.

- Implement the shared responsibility model across all AI Systems to emphasize the responsibility of the entity and the third-party vendors.

- Maintaining transparency about their outsourcing arrangements and report material outsourcing activities to the BSP.

Refer to:
i. BSP Circular No. 1137 on the Amendments to Regulations on Outsourcing and IT Risk Management
ii. BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions
iii. ISACA Shared Responsibility Model

[8] Data minimization should be practiced by collecting only the minimum amount of personal data as needed - implemented by the Philippines Data Privacy Act.

RISK MANAGEMENT

- **Risk-Based Assessment:** Evaluating the comprehensiveness of the risk-based assessments conducted.
- **Human-in-the-Loop:** Assessing the presence of an approved and consistent approach to integrating human oversight with machine processes.
- **Bias Prevention and Detection:** Identifying mechanisms for detecting, preventing, and correcting bias.
- **Drift Analysis:** Examining tools for monitoring model drift or decay and ensuring regular retraining of AI/ML models.
- **Interoperability, Evolvability, Modularity:** Investigating the interoperability, scalability, and plans for the evolvability and modularity of AI/ML models.

Risk management is a fundamental step in the AI system lifecycle. Of the 10 institutions interviewed, five (5) are at Stage 2, implementing automated systems and comprehensive processes for managing risk of their AI systems. In contrast, four (4) institutions are still in the early stages of adoption and are working on developing a comprehensive risk management framework for their AI systems. Only one (1) institution is in the exploratory phase (Stage 0), with no established practices for AI-specific risk management or no AI systems in production.

RISK-BASED ASSESSMENT

- Majority of the institutions treat their AI systems the same as any IT model or project and applies existing controls and risk assessment procedures for AI system development. Only three (3) institutions have an existing risk assessment framework specifically for AI systems which they are proactively updating. Other institutions are currently in the process of developing or formalizing their frameworks.

- Risk assessment processes mostly involve quantitative aspects. Moreover, heavy emphasis was given to operational risk issues and model materiality (i.e. Profit and Loss (P/L), opportunity costs. and business/success metrics). For institutions that have a centralized AI committee (or equivalent), final decision for model deployment/decommissioning ultimately rests with the committee in close coordination with the business owner. While most institutions have clear and well-defined roles and responsibilities for AI system development, there is an observed lack of internal controls especially for cybersecurity assessment.

SUPERVISORY EXPECTATION ON AI SYSTEM RISK-BASED ASSESSMENT

Since AI is a broad field and initiatives in the area range from simple models to complex AI systems, the BSP emphasizes the need for risk-based assessment per project.

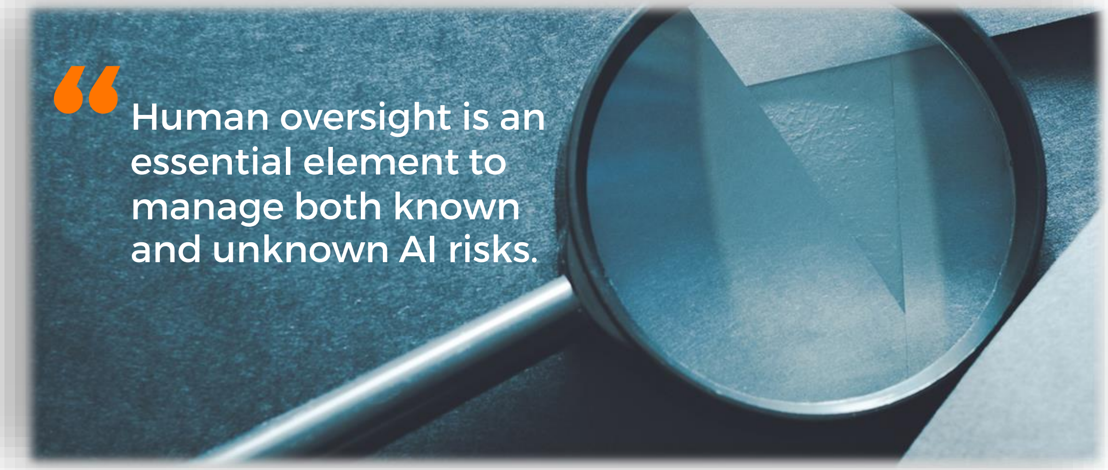
When AI-related initiatives are foreseen to increase, risk-based assessment through risk tiering should be implemented to uniformly determine if proposed projects are within the risk appetite and tolerance of the entity. Similarly, identifying high-risk AI system is beneficial for prioritization and risk management.

Refer to:
i. BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions
ii. BSP Circular No. 971 dated 22 August 2017 on the Guidelines on Risk Governance on the enhanced requirements on risk management systems of institutions.

HUMAN-IN-THE-LOOP

- All 10 institutions acknowledge the critical role of human oversight in managing AI systems. However, the implementation of human-in-the-loop practices varies significantly. Only three (3) institutions have established comprehensive and well-integrated human oversight mechanisms, involving manual checks and validations throughout the model lifecycle. Conversely, other institutions are either still developing or inconsistently applying these practices, with some still in the early stages of formalizing their frameworks.

- None of the critical AI systems are directly customer-facing; instead, their results serve as reference or guide data for banking functions. There is a particular case of credit score AI system wherein the outputs are showed directly to customers to serve as a guide in maintaining a healthy credit score. However, credit application is still not fully automated.



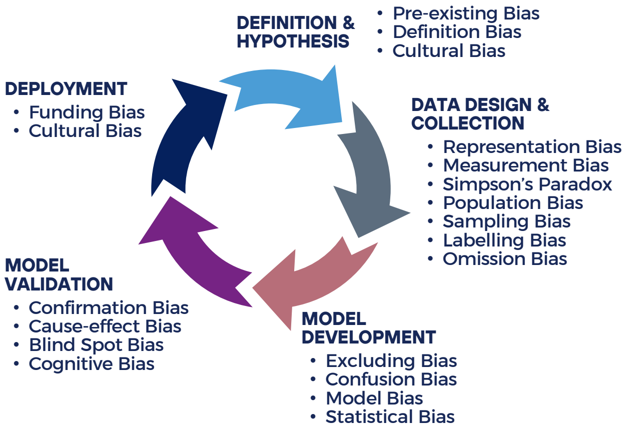
SUPERVISORY EXPECTATION ON HUMAN-IN-THE-LOOP

Since AI system is considered an emerging technology, proper stress testing procedures apply.

Pre-deployment. AI system users should be thoroughly informed and prepared by the management to interpret and detect questionable output, especially for critical or decision-making processes. The extent of involvement should depend on the associated risks of the specific process linked to the AI system.

Post-deployment. Users should be able to review and reverse any questionable output in a specific time period before any risk arises.

Refer to:
i. BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions
ii. BSP Circular No. 989 dated 04 January 2018 on the Guidelines on the Conduct of Stress Testing Exercises provides the overarching governance standards and risk management expectations on stress testing practices in the banking industry, including the supervisory expectations on the use of models for stress testing.



TAXONOMY OF BIAS

Bias is the degree to which a reference value deviates from the truth (Organization for Standardization (ISO), 2006)

According to the European Parliamentary Research Service (EPRS), there are different types of bias arising from different factors, external or internal. Each type of bias can occur at specific stages of the model lifecycle. Being aware of each type of bias can better equip developers and users to detect and monitor its harmful effects.

Source: (European Parliamentary Research Service, 2022)

BIAS PREVENTION AND DETECTION

- Bias can lead to mistrust, reputational risks, and even legal risks. A common approach among institutions is to exclude sensitive and discriminatory data (e.g., gender, race) from their AI systems to mitigate bias. Despite these efforts, there is an ongoing acknowledgment of the inherent bias in data. While explicit bias-imbued data can be avoided, there remains the risk that other data points may inadvertently serve as proxies for such biases.

- Most institutions implement techniques to prevent bias in their AI systems, but only a few incorporate bias detection algorithms or tools into their risk assessment processes. One (1) institution has noted that bias checks are not a mandated requirement for their AI systems, while another is exploring external providers for specialized bias detection tools. The institution has explicitly expressed the need for regulatory guidance on managing bias-imbued data, highlighting the necessity for clearer standards and practices to address this critical issue effectively.

SUPERVISORY EXPECTATION ON BIAS

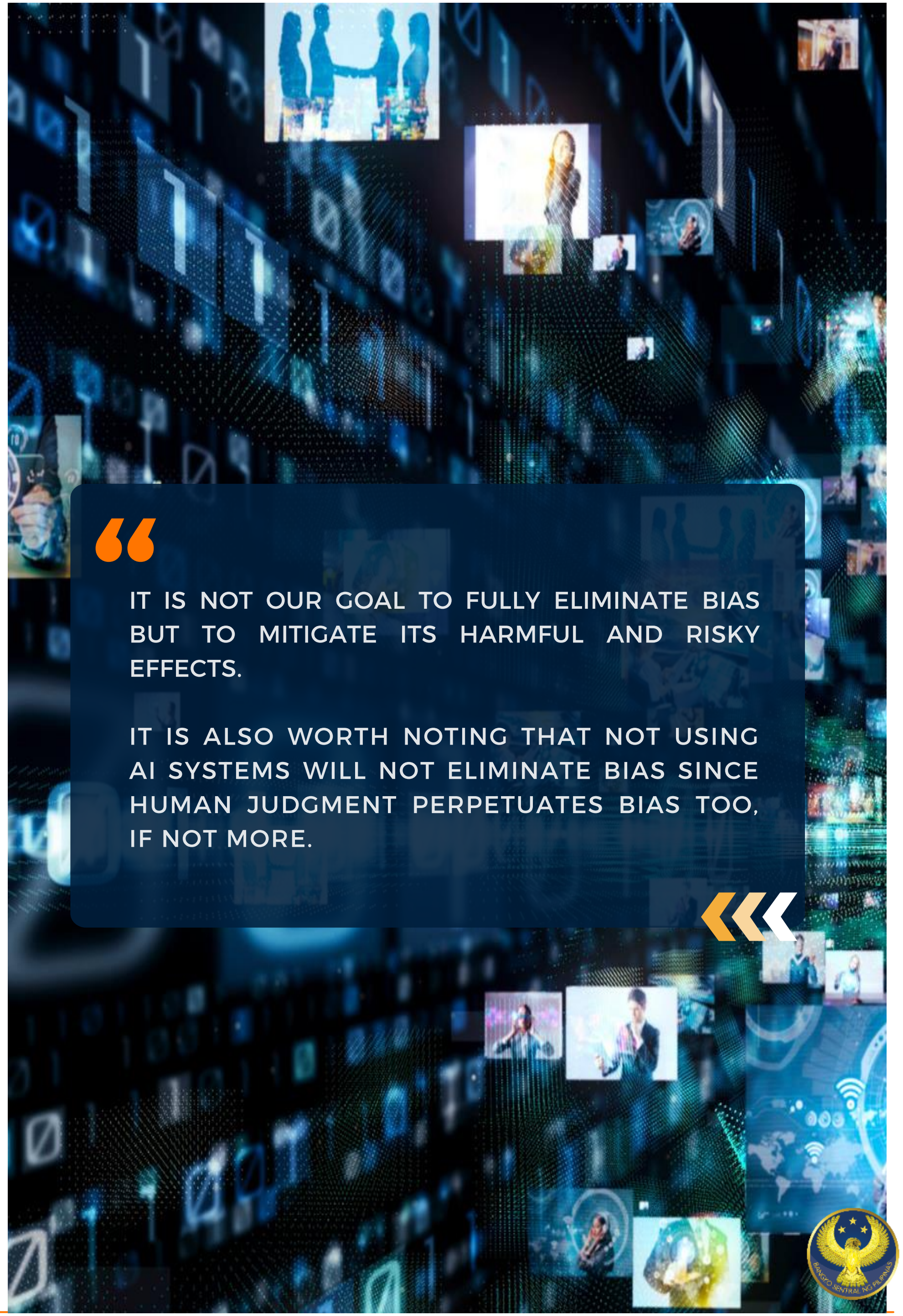
The types of bias that we want to mitigate are those with harmful and discriminatory impact. In the absence of concrete approach for hallucination and biases, institutions must pro-actively monitor business and statistical metrics of the output.

- The main source of bias is on the data – from collection, labelling, to processing. Ensuring the high quality of data can mitigate bias early on.

- Analysis should be performed to identify key points where there is a risk of harm or discrimination.

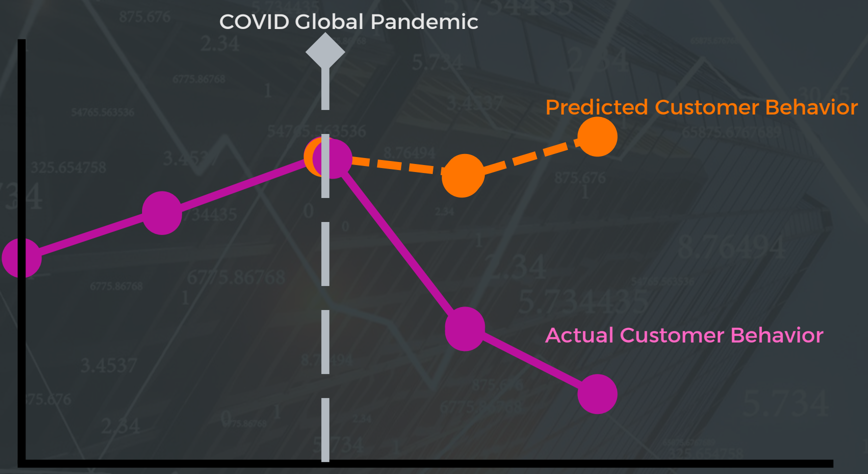
- Human-in the-loop technique could also be an effective practice when there is a diverse set of skills during the development process that could vet on the fairness of the AI system output.

Refer to:
i. BSP Circular No. 544 dated 15 September 2006 on the Guidelines on Market Risk Management
ii. BSP Circular No. 855 dated 29 October 2014 on the Guidelines on Sound Credit Risk Management Practices
iii. BSP Circular No. 900 dated 18 January 2016 on the Guidelines on Operational Risk Management
iv. BSP Circular No. 1160 dated 28 November 2022 on the Regulations on Financial Consumer Protection to Implement Republic Act No. 11765, otherwise known as the "Financial Products and Services Consumer Protection Act"



> "The risk of not monitoring model drift can lead to disruption of operations resulting in financial losses.

EXAMPLE OF SUDDEN MODEL DRIFT



Model drift can be caused by concept drift, data drift, or by upstream data change (IBM, 2024). A prominent example from recent years is the effect of the COVID global pandemic to existing financial models. Customer behavior radically changed during these years or even post-pandemic. However, model drift can also happen gradually over time.

MODEL DRIFT

- Model Drift is the decrease in a model's predictive accuracy over time. Common metrics used for detecting model drift include the Population Stability Index (PSI) and various performance metrics like accuracy and recall. Some institutions are advanced in model drift monitoring and utilize real-time dashboards and alert mechanisms for drift/decay detection in AI systems. The frequency of monitoring ranges from daily to monthly, with some institutions focusing on continuous real-time monitoring and others on periodic evaluations. Daily or real-time monitoring helps in the timely detection of issues, whereas periodic reviews allow for more comprehensive assessments.

- Institutions are employing diverse strategies for monitoring and managing AI systems, with a notable emphasis on automated tools for drift detection. While some institutions have well-established and systematic approaches, one (1) institution is still developing its processes or is in the exploratory phase. Institutions have varying approaches to model retraining, with some relying on automated tools and predefined thresholds, while others use manual reviews and business rules. The process and trigger for model retraining are subjective across institutions. However, the ultimate decision rests with the model owner upon consultation with business units.

SUPERVISORY EXPECTATION ON MODEL DRIFT AND REGRESSION MECHANISM

To fully protect consumers from risks arising to model drift, the following practices are expected:

- Business and model metrics to detect model drift should be determined prior to deployment.
- In accordance with Business Continuity Plan, there should be an established process once these metrics are triggered, depending on the severity and impact on critical processes.
- Once ready for re-deployment, the steps in the AI system lifecycle should be followed without skipping steps (i.e., validation should be done after recalibrating or revamping the model or data)
- Risk mitigation plan for model/data drift or AI system retirement/decommissioning should also be in place to ensure business continuity.
- Breached metrics should be properly reported to the risk management team. The report should also include a specific action plan and timeline to effectively address the issue, especially for critical systems and processes.

Refer to:
i. BSP Circular No. 951 dated 20 March 2017 on the Guidelines on Business Continuity Management
ii. BSP Circular No. 1160 dated 28 November 2022 on the Regulations on Financial Consumer Protection to Implement Republic Act No. 11765, otherwise known as the "Financial Products and Services Consumer Protection Act"

INTEROPERABILITY, EVOLVABILITY, MODULARITY

- The adoption of comprehensive monitoring frameworks and data management practices among institutions is on the rise, although substantial gaps persist, particularly in vendor management and legacy systems integration.
- Implementing a unified data architecture can significantly boost efficiency and transparency, suggesting that investing in such platforms would be advantageous for many institutions.
- Integration of new AI systems remains challenging due to existing legacy systems and the ongoing development of overall infrastructure.

DATA, MODEL AND TOOLS

AI Tools & Platform cover the type and amount of AI-specific tools and platforms the institution uses for its AI/ML development, management, and testing. Further, patch management and updating are checked in terms of the service provider support and capability on technology refreshment.

Data Management and Architecture pertains to the AI and data architecture in place. This includes the data storage and management set-up, level of centralization of AI data, data preparation procedures and the amount of automation involved.

AI System Lifecycle looks into the development lifecycle specific for AI systems. Focus is given on the type, scope and frequency of validation procedures for AI systems, both during the development stage and after implementation.

Among the 10 institutions, five (5) have demonstrated Stage 2 maturity in their AI/ML data, model and tools implementation and management. This means that 50% of the participants have defined AI/ML data management and model development procedures in place, complemented by AI-specific architecture and tools. Only three (3) institutions showed Stage 3 maturity or having an advanced data and model governance system that is integrated with the rest of its IT operations and support functions.

Risk management and compliance expectations on AI/ML Data, Models, and Tools are generally covered by existing rules and regulations. Proper AI systems management (e.g., data preparation, model planning, model testing, etc.) is expected to adhere to existing guidelines on project management, data privacy and security, outsourcing, and IT risk management.

TOOLS & PLATFORM

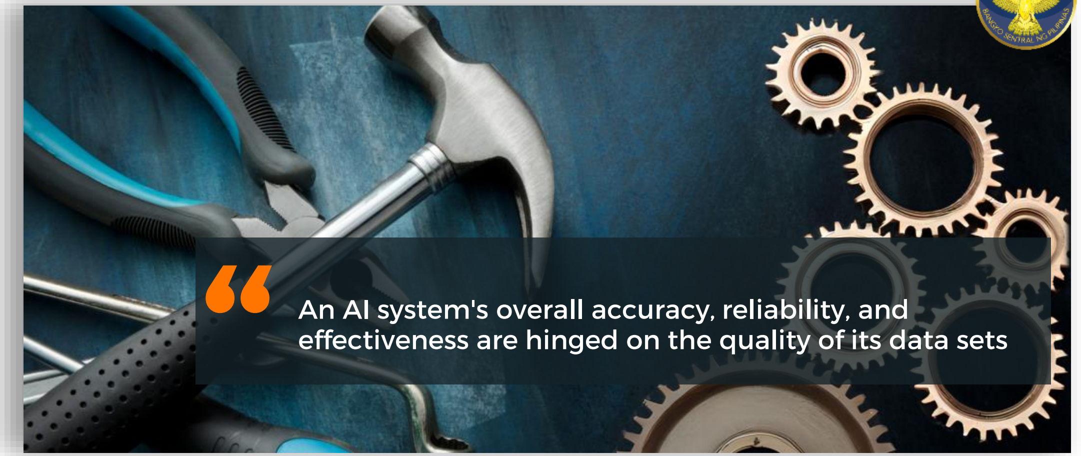
- Institutions' utilization of AI-specific tools, platforms, and systems depends on their risk appetite statement towards AI developments. Institutions with more model development projects utilize a variety of AI tools such as AWS, Databricks, Gitlab, Python, Microsoft Azure, among others. Institutions with slower AI development pacing, are still in the initial adoption phase with limited AI tools mostly concentrating on cloud or programming-related suites. Nonetheless, all institutions ensure that the platforms and tools utilized are regularly patched and updated.

DATA MANAGEMENT AND ARCHITECTURE

- Majority of institutions with significant AI system developments or AI implementations have employed some form of data architecture. These range from data warehouses as a centralized repository of transformed data to data lakes that have a larger storage capacity to cater to both raw and structured data. Only two (2) banks have a working data lakehouse architecture, which combines the data lake's centralization of diverse data types and the data warehouse's management and analytical capabilities.

- Data preparation (i.e., data extraction, sanitation, and transformation of raw data into useable data sets) is done either manually or semi-automated by most institutions. This ensures that human oversight and intervention are constantly present, particularly for high-risk/critical AI use cases such as credit risk and lending. The same human-in-the-loop requirement is also present for institutions with fully automated data preparation.

- One of the main challenges in AI system development is still data quality. Data quality and integration were not prioritized by most institutions due to the global pressure to catch up. Although enhancements are being made to address issues related to data and infrastructure, institutions will tend to rely on third-party vendors if these issues persist.



SUPERVISORY EXPECTATION ON DATA QUALITY

It is worth noting that all risks surrounding AI could stem from data quality. An AI system's overall accuracy, reliability, and effectiveness are hinged on the quality of its data sets. If the training data is inaccurate, erroneous or biased, then the AI system may return incorrect or irrelevant output. This also explains the concept of "Garbage in, garbage out" or GiGo for AI systems.

- Data quality should be evaluated at the input level. Accordingly, the methodology for labeling (target/dependent variables' value) must be systematic and must also be documented. By doing so, institutions could prevent bias stemming from their training dataset.

- A robust data management and transformation framework should be in place to ensure that datasets fed into AI training models are accurate, relevant, and unbiased.

- As also recommended by the Monetary Authority of Singapore (MAS) in their comprehensive report on data governance and management practices, data issues and quality, along with other relevant metrics and KPIs, should be reported to the risk management group for a holistic view and alignment on the overall roadmap. It should be a detailed breakdown of the data source, quality issues, common root causes, and risk implications.

Refer to:

i. BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions

ii. BSP Circular No. 971 dated 22 August 2017 on the Guidelines on Risk Governance on the enhanced requirements on risk management systems of institutions.

AI SYSTEM LIFECYCLE

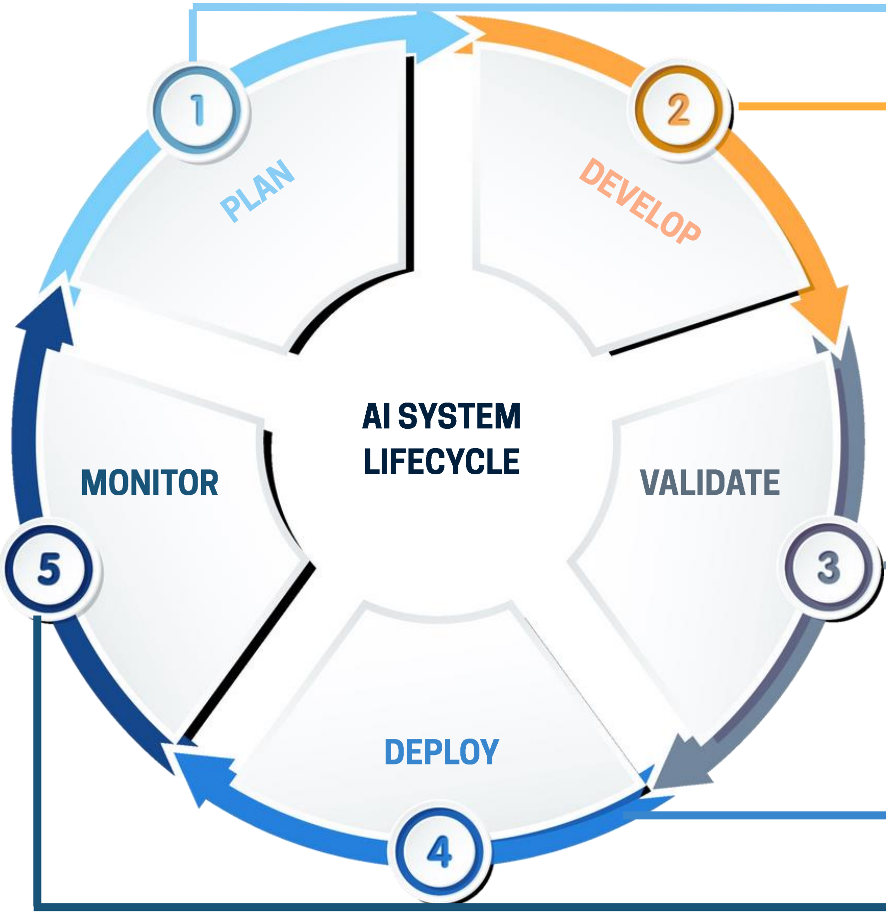
- Six (6) out of the 10 institutions perform AI model validation procedures both during the AI development life cycle and as an ongoing basis post-deployment. This typically involves qualitative (e.g., documentation, data quality, AI model design and purpose, etc.) and quantitative (e.g., Mathematical and Statistical Techniques) analysis. However, there are a handful of the institutions that have yet to formally include the performance of periodic model validation for deployed AI systems in their policies.

- Nonetheless, the majority of the institutions perform AI-specific test activities on their AI systems, on top of the regular test activities under the project management framework or software development life cycle. Common exercises include AI back-testing, stress testing, and discriminatory or bias testing.

- It was also determined that only a few institutions have included AI model development and data management in their Internal Audit and Compliance programs. Some banks only include select critical AI use cases (e.g., credit operations and lending) for IA and Compliance reviews.

SUPERVISORY EXPECTATION ON AI SYSTEM LIFECYCLE

The AI lifecycle should include all significant checkpoints as follows:



PLAN

- Conduct risk-based assessment taking into consideration the alignment of purpose with intended use
- Clearly defined roles and responsibilities especially for AI model/system owners, users, developers, validators, reviewers, and deployment team. This should clearly state the associated accountability as well expected from each role.
- Preparation of data availability together with data quality evaluation and feature relevance
- Options and decisions made throughout the AI system lifecycle should be documented. It is also crucial to identify whether the model is fundamentally suitable for the specific application.

DEVELOP

- Ensure transparency with proper documentation and version control. Methodology should be documented to justify the soundness of each decision or step.
- AI model validation, testing and monitoring framework should help spot model errors early on, while maintaining model fairness, usability, and relevance.
- Sufficient testing (i.e. back-testing, out-of-sample test) should also be performed when choosing the final model, version, or AI system.

VALIDATE

- Independent validation, aside from the usual developer validation, should have a degree of independence from the team of developers to ensure the integrity of the results.
- To demonstrate the robustness of the model, the results of an algorithm should also be reproducible. This reproducibility ensures a minimum transparency that AI systems should be built upon. Reproducibility also requires proper documentation to follow best practices.
- Management should also be capable to effectively challenge any decision or output during this stage.
- Evaluation should also include stakeholder metrics and success criteria.

DEPLOY

- Ensure transparency and accountability through independent deployment and proper documentation of each step.
- Sufficient testing, in any appropriate form, should also be conducted.

MONITOR

- Ensure Business Continuity Plan (BCP) is kept updated and tested.
- Continuous or regular monitoring should be in place observing business and stakeholder metrics.
- Regression process should be ready in case of system or output irregularity.

Refer to:
i) BSP Circular No. 808 dated 22 August 2013 on the Guidelines on Information Technology Risk Management for All Banks and Other BSP Supervised institutions
ii) BSP Circular No. 900 dated 18 January 2016 on the Guidelines on Operational Risk Management
iii) BSP Circular No. 971 dated 22 August 2017 on the Guidelines on Risk Governance on the enhanced requirements on risk management systems of institutions.
iv) BSP Circular No. 989 dated 04 January 2018 on the Guidelines on the Conduct of Stress Testing Exercises provides the overarching governance standards and risk management expectations on stress testing practices in the banking industry, including the supervisory expectations on the use of models for stress testing.

TALENT AND STRUCTURE

Organizational Structure refers to the current functional setup of AI development and implementation in the organization.

Workforce Training Plan refers to a development plan tailor-fitted to the level and needs of employees, board members, and key officers.

Workforce Development Rate refers to skill-building activities (e.g. workshops, seminars, courses) being completed by employees, board members, and key officers.

Among the 10 institutions covered, seven (7) have demonstrated Stage 2 maturity on talent and structure related to AI. The assessment indicates that majority of the reviewed institutions have a functional structure and supports its talent adequately. All reviewed institutions reflect a broad spectrum of sophistication and formalization in terms of their organizational structure surrounding AI. Meanwhile, three (3) out of 10 institutions still lacked formal governance structures leading to potential inconsistencies in oversight. Decentralized approaches, such as mesh-type organizational structure, allow for rapid innovation but may pose challenges in maintaining consistent oversight. These challenges present potential risks such as misalignment with strategic objectives leading to inefficiencies.

ORGANIZATIONAL STRUCTURES

- Some financial institutions exhibit high levels of maturity with well-established specialized groups such as dedicated risk management and data science teams. This enables effective governance and oversight of AI/ML initiatives. Some observed best practices for overseeing the use of AI include establishing dedicated committees and adopting an interdisciplinary approach that involves several departments such as data science, risk management, and compliance. Some banks have adopted proactive strategies by establishing Centers of Excellence and dedicated training programs that support skill development across various levels of the organization.

- Other institutions have integrated AI within broader technology departments or are still in the early process of developing their frameworks. Some banks operate with collaborative or decentralized nature, which, while can be innovative, has its own challenges.

- One major challenge, not just in AI but also in the whole technological field, is talent shortage. The scarcity of skilled AI professionals can severely constrain a bank's ability to develop, implement, and advance AI initiatives. This shortage further delays the recruitment of qualified individuals who possess the necessary expertise to handle complex AI systems.

- As regards foreign-owned institutions (fintechs and digibanks), it is challenging to assess AI initiatives, governance, and risk management as this is usually handled by regional counterparts; hence, local subsidiaries do not necessarily have visibility on the mechanics and underlying structure.

WORKFORCE TRAINING PLAN

- Approaches to workforce training differ widely among the reviewed entities. While some banks offer a large volume of training programs, others have taken a proactive approach by creating well-rounded training plans. This has been achieved by continuous optimization of their current workforce strategy.

- Several banks are implementing targeted upskilling strategies and exploring diverse training options, including international programs. This reflects a strategic effort to enhance skills relevant to AI and adapt to evolving industry demands. Having these plans illustrate a forward-looking approach to workforce development, aimed at ensuring that employees remain competitive and proficient in the rapidly changing financial landscape. This will help ensure that banks are well-prepared to meet the demands of this dynamic field.

- Developing a comprehensive workforce plan which aligns to specific organizational remain an ongoing challenge for many banks.

WORKFORCE DEVELOPMENT RATE

- AI-related training among the institutions varies significantly, reflecting different extents of effort to enhance AI skills of their talent pool. Some institutions have implemented numerous skill-building initiatives, yet with low completion rates which suggests that these programs may need to be more engaging or accessible. While some entities do not at all track the completion rates of their training offers, they maintain that the nature of their materials is self-paced thus requiring lesser monitoring.

- On the other hand, some banks have developed comprehensive training plans that feature extensive certifications and workshops, in which selected talents are mandated to complete leading to notable employee engagement set-up.

- Most of the observed entities have also begun conducting general AI awareness programs as their first step in integrating AI within their culture.

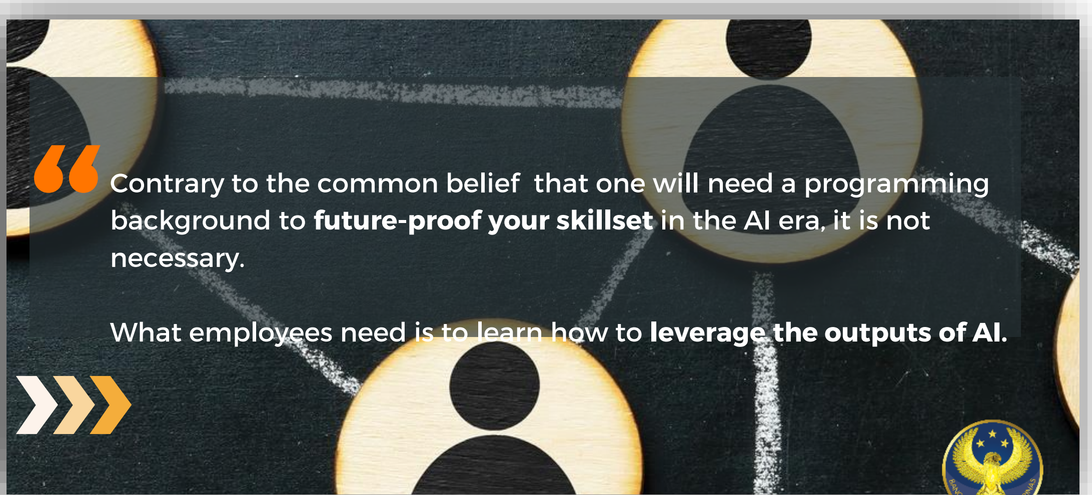
- Despite these efforts, a common issue is the lack of detailed tracking and metrics, which hampers the ability to fully assess the effectiveness and reach of training efforts. Additionally, the lack of performance metrics may prevent the identification of areas for improvement, perpetuating skill deficiencies and hindering progress. This inadequacy can lead to a workforce that is not fully equipped with the necessary skills to leverage AI technologies effectively.

SUPERVISORY EXPECTATION ON WORKFORCE DEVELOPMENT AND TRAINING PLAN

According to talent experts in the Singapore Fintech Festival 2023 who talked about the topic of building a future-proof skill set, "AI users" are the ones who will thrive in the future AI era. Contrary to the common belief that one will need a programming background to future-proof your skillset, it is not necessary. What employees need is to learn how to use as leverage the outputs of AI. This also supports the call that AI should be used to augment human capabilities.

Hence to support the continuous development of the institutions' workforce, proper information drive and training plan should be prioritized. Creating a training plan fit for each level of AI system participation allows for a targeted approach that could result in better appreciation and improved efficiency.

Refer to: BSP Circular No. 900 dated 18 January 2016 on the Guidelines on Operational Risk Management



CONSUMER PROTECTION AND ETHICS

Responsible, Governable, and Contestable use of AI refers to practices ensuring basic principles such as accountability and contestability are in place

Transparency and Explainability are qualities of a robust AI system which means the users, developers, and management knows how it works on different levels.

Human-centric, Fair, and Equitable refers to adherence to ethical practices.

Among the 10 institutions, only two (2) have demonstrated stage 2 maturity with approved guidelines and enforced practices to ensure the responsible use of the technology. Moreover, seven (7) were assessed as stage 1 with a standardized methodology while standardized policies are being considered.

RESPONSIBLE, GOVERNABLE AND CONTESTABLE USE OF AI

- Almost all visited institutions have practices in place on how (a) accountability is observed, (b) compliance with regulations including consumer protection is adopted, and (c) initiatives are overseen or governed. However, among these institutions, many still do not have approved AI policies/frameworks. On the other hand, for those who already have relevant policies/framework, implementation is still in the early stages.
- It was also found that institutions have limited attention to compliance with consumer protection regulations. This may be associated with the functions of deployed use cases which are mostly for operational efficiency or supporting processes. When asked about consumer protection, the main consideration is compliance with Data Privacy regulations. Therefore, the risk that other regulatory requirements on consumer protection might be overlooked is present.
- Institutions, especially those active in AI system development, showed proactiveness and resourcefulness by referring to AI regulations or frameworks issued by or available in other jurisdictions.

EXPLAINABILITY AND TRANSPARENCY

- Nearly all the subject institutions have mechanisms to ensure that their respective AI initiatives adopt transparency and auditability, but only a few have approved standardized processes on this aspect. Generally, institutions utilize available tools or techniques to check model explainability and ensure that the outputs of AI systems can be explained and justified.
- Conservative institutions benchmark their sophisticated models' output to a challenger model's output (or external model's output) before committing to the resources that sophisticated models require. Previous studies suggest that the accuracy of using neural networks versus traditional models is comparable or at par, depending on the problem at hand (Kumar, 2005).
- Challenges in this area largely relate to the required level of transparency of its AI system vis-à-vis nondisclosure of proprietary information to maintain business competitiveness, and the lack of local regulations or clarification on how to approach AI.

ETHICS AND DISCRIMINATION

- On the observance of human-centricity, fair and equitable AI principle, most institutions have implemented measures or controls despite the absence of formalized procedures through AI-related policies or frameworks.
- These measures and controls vary in nature. Examples are feature analysis (i.e., justification of top model features or exclusion of unnecessary data that could induce bias), establishment of an AI Ethics committee, and development of explainability toolkits, among others.
- Fairness is a complex and crucial concept to define. Individuals may draw different lines on what may seem fair and what might not be.

SUPERVISORY EXPECTATION ON EXPLAINABILITY AND TRANSPARENCY

There is a tendency to give high levels of autonomy to AI blackbox models in exchange for higher accuracy. This, in turn, can increase other risks as well, especially if transparency is not met. This lack of transparency could also influence the decision-making process of the management, stakeholders, and even customers – affecting the overall safety and soundness of the institution.

- Transparency promotes trustworthiness. Institutions should give relevant information to AI systems' users (internal or external), stakeholders, and management such as the features/components which are weighted the highest for specific outputs. AI system output is only as good as the interpretation of the end-user. Hence, without the proper knowledge of the internal process of the system, the end-user might not maximize its potential or, worse, might misinterpret the output.
- Institutions should conduct proper workshops and maintain adequate documentation to AI systems' users (internal or external) to ensure proper usage and handover to business units (BUs). Risk of misinterpretation from users might lead to incorrect usage – due to the highly technical and complex nature of AI models, outputs can be correct but might be misinterpreted by the users due to lack of clear understanding about the model and the output.

Refer to:
i. BSP Circular No. 900 dated 18 January 2016 on the Guidelines on Operational Risk Management
ii. BSP Circular No. 1160 dated 28 November 2022 on the Regulations on Financial Consumer Protection to Implement Republic Act No. 11765, otherwise known as the "Financial Products and Services Consumer Protection Act"

Transparency refers to the "extent to which information about an AI system and its outputs is available to individuals interacting with such a system" (NIST, 2023). **Transparency** can answer the question **"what happened"** when the need arises. **Interpretability**, on the other hand, "refers to the meaning of AI systems' output in the context of their designed functional purposes" and aims to **answer the question "why"** a decision was made (NIST, 2023). **Explainability** refers to "a representation of the mechanisms underlying AI systems' operation which should answer the question **"how" a decision was made** in the system (NIST, 2023).

"**Blackbox" model** is the blanket term used for complex models that internal working is hard to trace or explain (e.g. neural network, gradient boosted trees). There is a particular fear in the area of AI and machine learning for blackbox models in the topic of transparency. However, some institutions argued that blackbox models can still be explained using the new techniques in the field of explainable AI (XAI). But in general, there is a tradeoff between model predictability versus the interpretability of the output.

SUPERVISORY EXPECTATION ON ETHICS AND DISCRIMINATION

Addressing these ethical considerations could mitigate potential risks to internal or external users.

- Generative AI systems oversight. Aside from checking the quality of the output of conversational agents, checking the quality of input statements/questions should also be included for regular monitoring.
- The ultimate accountability for decision-making processes lies on humans. AI systems' output cannot replace human responsibility.
- Ensure "Fair Treatment" of AI system outputs (e.g., not discriminate against clients on the basis of race, age, and similar factors) aligned with FCPA.

Refer to:
i. BSP Circular No. 1160 dated 28 November 2022 on the Regulations on Financial Consumer
ii. Protection to Implement Republic Act No. 11765, otherwise known as the
iii. "Financial Products and Services Consumer Protection Act"

CONCLUSION

GAP	RISK	MITIGATING ACTION
Absence of appropriate **risk management controls** throughout the AI lifecycle	• Exacerbate bias and discrimination • Prone to cyber-attacks and fraud/scam schemes amplified by AI • Exposure to unique data privacy risks associated with AI systems • Exposure to outsourcing risks and market concentration	• Ensuring compliance with all relevant laws and regulations • Implement AI-risk tiering • Apply human-in-the-loop and proper stress testing procedures • Mandatory Privacy Impact Assessment (PIA) for AI systems
Absence of **AI roadmap and policy**	• Misalignment with institution's overall strategy • Resource wastage	• Developing comprehensive AI roadmaps and policy • Involving various internal stakeholders
Lack of **workforce development plan** and metrics to prepare the employees in the so-called AI revolution	• AI systems replacing or displacing employees	• Continuous workforce development with clear goals
Lack of **transparency and explainability**	• High level of autonomy given to AI systems • Risks on the overall safety and soundness of the institution by relying on AI systems' output or recommendation without sufficient evidence or challenge	• Give relevant information to AI systems' users • Maintain adequate documentation to ensure proper usage • Implement transparent and user-friendly consent processes for data collection in AI systems
Gaps in **AI system lifecycle** foregoing crucial checkpoints	• Operational risk arising from unmonitored machine or human error • Cybersecurity and data privacy risks might arise when gaps are exploited by bad actors.	• Ensure all checkpoints in an AI system lifecycle are present
Data management and data quality issues	• Risks of intensifying bias stemming from different phases in AI lifecycle	• Ensuring the high quality of data through the implementation of a robust data management and transformation framework • Pro-actively monitor business and statistical metrics

The results of the preliminary intersectoral consultations and extensive review found that common AI principles such as accountability, auditability, robustness, security, transparency, fairness, and explainability are covered by existing BSP regulations. However, risks specific to AI are yet to be addressed. The first step to fully realize the potential of an emerging technology is to determine the pertinent risks and mitigate each aptly.

Currently, AI adoption in the Philippine financial sector is confined to narrow usages. However, there are several use cases that are in development that cover an array of use cases interlinked with institutions' critical processes. Starting with the pertinent risks identified, institutions should implement mitigating controls to allow for a safe and responsible innovation.

The BSP recognizes that institutions are at different maturity levels in terms of AI adoption and governance. Nonetheless, the BSP expects institutions to put in place a proportionate AI policy framework depending on their risk appetite and set out fundamental controls on the management of AI systems. Accordingly, to address the risks discussed in this paper, further strengthening or expanding of existing rules appears to be necessary specifically in key risk areas such as consumer protection, cybersecurity, anti-money laundering and combating the financing of terrorism (AML/CFT) that should be in place in every instance of AI system implementation.

What does the future of financial services look like with AI at the forefront? The financial services industry is particularly exposed to AI higher than the global average (WEF, 2025). While 47% of businesses plan to transition AI-disrupted roles to other positions, 41% still plan to downsize their workforce due to the increasing capability of AI to replicate roles (WEF, 2025). Specifically for the Southeast Asia region, 91% of the businesses are leaning towards reskilling and upskilling their existing workforce rather than hiring new people (WEF, 2025). The time is here for financial institutions to carefully plan and execute their AI transformation journey keeping in mind to produce a positive and meaningful impact on society.

RECOMMENDATIONS

The above discussion covers the risks and supervisory expectations that institutions must put in place when adopting or exploring AI systems. The following discusses the recommended items for BSP-supervised financial institutions (BSFIs) to uphold the standard across the industry.

FOR BSFIS

1. BSFIs should **identify and implement AI-specific policies and frameworks** concerning the risks and supervisory expectations identified in this paper applying the concept of proportionality.

2. BSFIs should check their **level of compliance against existing regulations and laws** pertinent to their use of AI and improve their existing processes as necessary.

3. BSFIs should solidify their AI strategy by **upholding human-centric values to foster further trust** in the digital banking ecosystem.

With the rising adoption and appreciation of AI, recent developments in the area, and the fruitful use cases in the field, continued adoption is seen. Though there are still a lot of open questions and pending alignments to do, the future of AI in the financial system is promising.

This thematic review hopes to bring BSP closer towards developing a trusted AI ecosystem across industries, where AI is harnessed safely and confidently to uplift the society.

REFERENCES

(2024). Retrieved from EU Artificial Intelligence Act: https://artificialintelligenceact.eu/article/3/

AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity. (2024, 01). Retrieved from International Monetary Fund (IMF) Blog: https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity

Association of Southeast Asian Nations . (2024). ASEAN Guide on AI Governance and Ethics Contents. Retrieved from asean.org: https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf

Balaji, N., Bharadwaj, A., Apotheker, J., & Moore, M. (2024, April 8). Consumers Know More About AI Than Business Leaders Think. Retrieved from Boston Consulting Group Global: https://www.bcg.com/publications/2024/consumers-know-more-about-ai-than-businesses-think

Bloedorn, E., Kotras, D., Schwartz, P., Chaney, C., Chaney, C., & Patsis, J. (2023, November). The MITRE AI Maturity Model and Organizational Assessment Tool Guide: A Path to Successful AI Adoption. Retrieved from MITRE: https://www.mitre.org/sites/default/files/2023-11/PR-22-1879-MITRE-AI-Maturity-Model-and-Organizational-Assessment-Tool-Guide.pdf

Boston Consulting Group. (2020, October 20). Annual MIT Sloan Management Review—Boston Consulting Group Study Finds Significant Financial Benefits with AI Are Noticeably Amplified When Organizations Develop Multiple Effective Ways for Humans and AI to Learn Together. Retrieved from Boston Consulting Group: https://www.bcg.com/press/20october2020-study-finds-significant-financial-benefits-with-ai

Copeland, B. (2024, October 29). Artificial Intelligence. Retrieved from Encyclopedia Britannica: https://www.britannica.com/technology/artificial-intelligence

Crismundo, K. (2024, July 3). DTI Wants AI Research Center to Be Revenue-Generating. Retrieved from Philippine News Agency: https://www.pna.gov.ph/articles/1228244

Davenport, T., & Harris, J. (2017). Competing on Analytics: The New Science of Winning. Boston, Massachusetts : Harvard Business Review Press.

Department Of Information and Communications Technology, and Philippine Civil Service Commission. (2024, April 18). PRINCIPLES and GUIDELINES for an ETHICAL and TRUSTWORTHY USE of ARTIFICIAL INTELLIGENCE (AI) in the GOVERNMENT. Retrieved from Department Of Information and Communications Technology: https://dict.gov.ph/ictstatistics/wp-content/uploads/2024/05/DRAFT-COPY_JMC-on-the-Principles-for-an-Ethical-and-Trustworthy-Use-of-Artificial-Intelligence-AI-in-Government.pdf

Department of Trade and Industry. (2024, July 5). DTI Drives AI innovation with National Roadmap 2.0 and CAIR Launch. Retrieved from Department of Trade and Industry Philippines: https://www.dti.gov.ph/archives/news-archives/dti-drives-ai-innovation-national-roadmap-2-0-cair-launch/

European Parliamentary Research Service. (2022, July). Review of Auditing the Quality ofDatasets Used in Algorithmic Decision-Making Systems. European Parliamentary Research Service. Retrieved from European parliament: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729541/EPRS_STU(2022)729541_EN.pdf

Financial Stability and Oversight Council. (2023). Annual Report 2023. Retrieved from US Department of Treasury: https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf

Financial Stability Board. (2017, November). Artificial Intelligence and Machine Learning in Financial Services. Retrieved from Financial Stability Board: https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

House Bill No. 10457, 18th Congress of the Republic. (2021, November 9). Retrieved from Senate of The Philippines: https://issuances-library.senate.gov.ph/bills/house-bill-no-10457-18th-congress-republic

House Bill No. 7396, 19th Congress. (2023, March). Retrieved from Senate of the Philippines: https://issuances-library.senate.gov.ph/bills/house-bill-no-7396-19th-congress

Kumar, U. A. (2005). Comparison of neural networks and regression analysis: A new insight. Expert Systems with Applications, 424-430. doi:https://doi.org/10.1016/j.eswa.2005.04.034

Lakhno, M. (2023). What Is a Policy Framework? An Attempt at Conceptualization. Sociální studia/Social Studies, 89–111. doi:10.5817/SOC2023-35675

Monetary Authority Of Singapore. (2018, November 12). Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector. Retrieved from Monetary Authority Of Singapore: https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat

National Institute of Standards and Technology. (2023, January). Artificial Intelligence Risk Management Framework (AI RMF 1.0). Retrieved from National Institute of Standards and Technology: https://doi.org/10.6028/nist.ai.100-1

Organisation for Economic Co-operation and Development. (2024, September). Regulatory Approaches to Artificial Intelligence in Finance. Retrieved from OECD iLibrary: https://doi.org/10.1787/f1498c02-en

Organization for Standardization (ISO). (2006). ISO 3534-1:2006.

Personal Data Protection Commission (PDPC) & Info-communications Media Development Authority (IMDA). (2020). Model Artificial Intelligence Governance Framework Model Second Edition. Retrieved from https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation

Quismorio, E. (2023, March 12). Barbers Bats for Creation of AI Agency. Retrieved from Manila Bulletin: https://mb.com.ph/2023/3/12/barbers-bats-for-creation-of-ai-agency

Republic Act 10173 Data Privacy Act of 2012. (2012). Retrieved from National Privacy Commission: https://privacy.gov.ph/data-privacy-act/

Schwab, K. (2016, January 14). The Fourth Industrial Revolution: What It Means, How To Respond. Retrieved from World Economic Forum: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

United Nations Industrial Development Organization. (2021). Empowering SMES through 4IR Technologies: Artificial Intelligence. Retrieved from UNIDO Knowledge Hub: https://hub.unido.org/sites/default/files/publications/Empowering%20SMEs%20through%204IR%20Technologies_0.pdf

Xu, X., Lu, Y., Vogel-Heuser, B., & Wang, L. (2021). Industry 4.0 and Industry 5.0—Inception, conception and perception. Journal of Manufacturing Systems, 530-535. doi:https://doi.org/10.1016/j.jmsy.2021.10.006