ITJ | The International Tax Journal

An integrated framework for AI governance in economic institutions: Managing ethical, operational, and regulatory compliance risks

Dr. Hammoudi Dalel [1], and **Dr. Abdelhak Djenane** [2]

[1] Mohamed Khider University of Biskra
d.hammoudi@univ-biskra.dz
[2] Mohamed Khider University of Biskra
abdelhak_djenane@univ-biskra.dz

Abstract---This article proposes an integrated governance framework for artificial intelligence (AI) in economic institutions, addressing ethical, operational, and regulatory compliance risks. As AI systems increasingly shape decision-making in financial institutions, the paper argues that fragmented governance approaches are insufficient to manage emerging risks such as algorithmic bias, opacity, cybersecurity threats, and regulatory uncertainty. Drawing on international principles from organizations such as the OECD and the European Commission, the framework emphasizes responsible AI grounded in fairness, accountability, transparency, and data stewardship. The study introduces a structured risk taxonomy aligned with institutional objectives and fiduciary responsibilities, highlighting the need for coordinated oversight across internal governance bodies, regulators, and market participants. By integrating risk assessment, monitoring, auditability, and continuous improvement mechanisms across the AI lifecycle, the framework supports both innovation and compliance. The paper contributes to policy and practice by offering a scalable, evidence-based approach for aligning AI deployment with ethical standards, operational resilience, and evolving regulatory expectations in economic institutions.

Keywords---AI governance; financial institutions; ethical AI; regulatory compliance.
JEL Classification: G21, G28, O33, K22

How to Cite:
Dalel, H., & Djenane, A. (2025). An integrated framework for AI governance in economic institutions: Managing ethical, operational, and regulatory compliance risks. *The International Tax Journal*, *52*(6), 5152–5167. Retrieved from https://internationaltaxjournal.online/index.php/itj/article/view/503

The International tax journal ISSN: 0097-7314 E-ISSN: 3066-2370 © 2025

Submitted: 12 Jun 2025 | Revised: 18 Nov 2025 | Accepted: 15 Sept 2025

1. Introduction

AI governance in economic institutions should be framed as an integrated, evidence-based framework that analyzes ethical, operational, and regulatory compliance risks within institutional objectives and stakeholder interests.

The unprecedented artificial intelligence (AI) capabilities of large language models and generative algorithms raise concerns about unintended consequences and misuse, challenging the intimidation of responsible and ethical AI by institutions (Choung et al., 2023). Economic institutions play a critical role in society by providing essential services that contribute to economic growth. These institutions, which include commercial banks, investment banks, asset managers, insurance companies, sovereign wealth funds, and development finance organizations, are under increasing pressure to develop and deploy AI systems responsibly in ways consistent with the analog era. Collaboration across technical, legal, and operational domains is required to develop an integrated Risk Governance framework that considers an institution's mandate, business model, and operating environment (Kurshan et al., 2020).

2. Conceptual Foundations of AI Governance

AI governance in economic institutions should be framed as an integrated, evidence-based framework that analyzes ethical, operational, and regulatory compliance risks within institutional objectives and stakeholder interests. The assessment of such risks requires a clear understanding of key concepts and security principles, informed by national and financial authorities and leading organisations.

The pursuit of AI deployment raises ethical concerns and calls for responsible governance. Definitions of responsible AI vary, but prominent guidelines converge around common principles. The OECD Principles on Artificial Intelligence advocate for AI that respects human rights and democratic values, promoting well-being, sustainability, and equality, while avoiding undue discrimination (Choung et al., 2023). The United Nations High-Level Expert Group on AI emphasizes that people should remain at the centre, encouraging beneficial welfare and freedom of choice (Gill et al., 2022). The European Commission recommends that AI should be lawful, ethical, and robust, and the G7 identifies trustworthy and reliable features as preconditions (Schneider et al., 2020). The unease surrounding the potential for model-generated disinformation further impels economic institutions to define a coherent position and organise related governance. The governance framework distinguishes internally developed AI systems relying solely on proprietary data from externally acquired, pre-trained models and application programming interfaces (APIs), which may require different handling based on varying degrees of autonomy and control over AI outputs.

2.1. Definitions and scope

AI governance in economic institutions should be framed as an integrated, evidence-based framework that analyzes ethical, operational, and regulatory compliance risks within institutional objectives and stakeholder interests.

Conceptual Foundations of AI Governance

Definitions and scope

AI governance encompasses the processes, policies, practices, and decision-making structures that determine how institutions develop, deploy, manage, monitor, and communicate about AI systems. Governance frameworks are typically designed to promote compliance with relevant regulations, to mitigate adverse consequences arising from AI systems, and to facilitate the alignment of AI strategies with institutional objectives (Schneider et al., 2020). AI governance hence forms part of the broader institutional structures needed to address potential risks associated with AI (Kurshan et al., 2020). A consolidated understanding of the governance landscape is essential to developing an integrated risk framework that satisfactorily addresses those risks from an institutional perspective.

At a higher level of abstraction, AI governance can be conceived as a special case of governance for algorithmic decision-making. Algorithms can be defined as "well-defined recipes for computation or inference that take one or more items of data as input and produce one or more items of data as output". Algorithmic decision-making itself is a further special case of machine-mediated decision-making, which can be defined as any situation in which human decision-making is aided, augmented, or replaced by input from machines. Algorithmic decision-making, then, occurs when the machines involved in a decision-making process make choices among options rather than simply providing auxiliary data or recommendations to supplement or narrow down the choices that remain for a human decision-maker. AI governance, ultimately, must be rooted in the legal foundations underpinning algorithmic decision-making, which differ between jurisdictions. (Suksi, 2023)

2.2. Principles of responsible AI

It is essential for AI governance to aim for beneficial outcomes while considering long-term impacts on society and the environment, to be accountable for all outputs and associated harms throughout the lifecycle of AI systems. An extension of the principle of beneficence is to consider not only the immediate benefits of AI but also to ensure that AI adoption does not hijack the institutional agenda by masking deeper social issues. Trust in AI systems requires a focus on competence, which encompasses safety, robustness, explainability, and integrity; integrity in turn includes fairness, nondiscrimination, privacy, and transparency. Competence alone is necessary but insufficient for socially beneficial AI, since future demands will include ensuring that AI and data use are aligned with institutional and societal objectives. Governance encompasses institutions' social responsibilities to minimise negative and ensure positive impacts through AI, government regulation and the responsibility of firms and individuals to meet regulatory expectations, corporate responsibility in adopting ethical AI beyond basic compliance, and societal pressures for greater governance of AI. Oversight can be provided by internal review bodies within institutions and external groups analogous to the FDA that monitor the acceptability of AI data, technology, products, services, and applications (Choung et al., 2023).

Fiduciary responsibilities arise from the social contract governing institutions that have been granted authority and legitimacy to manage resources on behalf of stakeholders and society. Such responsibilities go beyond compliance to encompass a wider set of stakeholder concerns regarding fairness, integrity, environmental sustainability, workforce skilling, and other ethical considerations. Adopting a stakeholder-oriented approach extends the rationale for governance to encompass the responsibility to uphold stakeholder interests and societal commitments throughout the AI lifecycle (Kurshan et al., 2020).

2.3. Risk taxonomy for economic institutions

The integration of Artificial Intelligence (AI) poses significant ethical and compliance risks to financial institutions and their stakeholders, necessitating the implementation of a comprehensive governance framework. Some of these risks are shareable through AI applications and models developed in-house, purchased from third-party vendors, or accessed through cloud-based services. To address this situation, a systematic approach to classifying the ethical and compliance risks stemming from the use of AI across financial institutions is essential.

The proposed taxonomy classifies risks across three key dimensions pertinent to economic institutions: operational risk, regulatory compliance, and consideration for ethical guidelines. These three areas of risk align with the central objectives of any economic institution, and adherence to ethical norms and regulatory compliance is embedded in acceptable operational practice for any organisation active in these environments (Kurshan et al., 2020) ; (Piorkowski et al., 2022).

3. Stakeholder Landscape and Governance Actors

The automobile sector in South Korea underwent thorough reform between 2019 and 2021. This process involved structural, administrative, regulatory, and safety dimensions concerning vehicles. A

comprehensive and authoritative multilevel integrated framework for post-reform governance of the automobile sector is proposed, and examples of applied governance are illustrated. Avoiding narrow and fragmented governance design and exploring diverse governance beyond a single sector—both of which are required—indicate significant research opportunities.

A multilevel integrated AI-governance framework with tailored content for national, regional, and institutional governance is proposed. Governance structures worldwide are becoming increasingly interconnected and interdependent, making it essential to consider regulatory design and principles throughout multi-scaling governance. The technical, economic, and policy development surrounding AI technology is often captured in a single layer. The analysis begins from this layer and identifies three supplementary dimensions: security safeguarding, extended administrative governance, and outcome-oriented monitoring for widespread economic sectors, systems, and administrative areas. These dimensions correspond to effective dissemination and deployment, compliance with existing or emerging norms, and assurance of conducive social benefit. Cyber deception, misinformation, and misinformation have surfaced as intermediate layers in the transportation sector, in parallel with the widely reported emergence of safety and stewardship in drone management. Current governance at such micro and separate levels rather than integrated macro governance hinders optimized overall governance. (Kahler, 2024)

AI-governance demand is explored based on an automobile-sector case in South Korea. Primarily viewed as a threat requiring preemptive regulation, AI technology is increasingly recognized as a substantial economic opportunity amid unprecedented global challenges. Notwithstanding the sector's vulnerability to protectionist measures, commercializing the regulatory sandbox represents an early opportunity before cross-border constraints—such as data jurisdiction, mobility, and chips—arise on a wider scale. Illustrating early AI governance within a transport-case frame assists in sharing South Korea's experience while engaging broader economic horizons such as money laundering surveillance, credit rating, and insurance. Therefore, the governance framework incorporates the pre- and post-reform national- and executive-level regulation, the regional dimension, and timely likelihood identification of AI yet without AI-layer description. The framework serves as a guide for future design and remains open to adoption elsewhere. (Gong & Zhang)

3.1. Internal governance bodies

Economic institutions are under pressure to develop the necessary governance frameworks for artificial intelligence and data analytics capabilities to balance operational risks and regulatory compliance with active innovation. Internal governance focused on risk, ethics, and data facilitates a holistic view of institutional operations and bounding values. The European Union's 2020 White Paper also addresses the delicate balance between promoting innovation and managing risks when discussing regulatory oversight of machine-learning systems (Kurshan et al., 2020). An integrated framework for AI governance is essential to achieve that balance.

AI governance should be construed broadly to encompass three aspects: risk management, ethical oversight, and data stewardship. An AI risk management framework outlines an institution's capacity to deploy AI systems effectively, an essential precondition for the widespread development of innovative applications (Choung et al., 2023). Supporting data strategy should address risk, ethics, and compliance in accordance with applicable regulations and institutional requirements (Gill et al., 2022).

Governance systems, including dedicated committees, cross-functional teams, and internal professional networks, provide additional support by ensuring institutional priorities align with individual initiatives and that AI deployment adheres to prescribed operational guidelines. Formal internal review bodies offer further scrutiny and can help build trust to foster adoption at scale by clarifying the boundaries of social acceptability and limiting valuation to trusted data.

3.2. External regulators and industry bodies

Participation in external AI governance initiatives has been growing across the financial services sector (Kurshan et al., 2020). Aulon defines the full universe of AI governance economic institutions must consider and indicates where the existing, major collective initiatives lie, clarifying scope and resulting targeted participation. Aulon's External Governance framework categorizes relevant organisations combined with institution activities and is complemented by supplementary insights. AI governance throughout outer tier centres on the need for trustworthy AI, broadly interpreted. The aspiration, encapsulated in the pronounced commitment, is to become a fully trusted financial institution globally serving the entire clientele spectrum; thus, the institution's AI governance strategy remains grounded in trustworthy principles. The framework, informed by other initiatives, also includes distinctive elements uniquely arising in external governance, such as compliance, sovereignty, enablement, and upholding fiduciary responsibility. Regulations and obligations from governing authorities dictate safeguards and resilience measures essential to protecting stability and maintaining society's well-being. Multi-regional-markets presence raises further jurisdictional complexities impacting data and technology. Enabling access to technology advances clients' interests, potentially accelerating growth while bolstering resilience post-pandemic. AI's transformative potential offers unprecedented opportunity, especially for underserved individuals and enterprises. Governance must ensure a secure approach capable of protecting end-users; several globally recognised initiatives are proactively pursuing such assurance. Economic institutions bear duty of care toward the general populace.

3.3. Market participants and fiduciary responsibilities

The identification and pursuit of unsecured lending and insurance opportunities are fundamental to the profitability of certain market participants. The relevant fiduciary duty across financial institutions is to assess risk-reward propositions tailored to the profiles and needs of individual clients. Obligations to comply with regulatory frameworks and promptly disclose risk-related information to clients constitute further responsibilities to the public that shape the governance of AI systems. In addition, provisions that block the pursuit of personal gain at the expense of the institution's reputation are relevant to the deployment of internally developed AI technologies.

The governance of AI systems must also align with a contingent value proposition that is relevant where intelligent agents have been designated as agents of market participants. A preliminary fiduciary duty concerns the reliability of the agents as representatives of the institutional principles. Institutions must take positive measures that ensure agent operations do not compromise the interests of the participating principals. Institutions remain accountable for ensuring that agent actions are consistent with the broader compliance framework of the institution and relevant regulatory stipulations (Benthall & Shekman, 2023).

4. Ethical Considerations in AI Deployment

AI systems have become integral in economic institutions. They are utilized in a multitude of applications, such as predicting market trends, detecting fraudulent transactions, and automatically excuting trades. These applications provide tremendous upside opportunities for the institutional objectives and efficiency, but come with serious ethical considerations. Ethical considerations in deploying AI systems are reflected by the traditionally known as FAT criteria: Fairness, Accountability, and Transparency. For a financial institution of certain class, fairness means that the model does not discriminate against any specific gender, race, or sensitive class of customers; accountability throughout the automation process needs to be reachable; and transparency indicates that stakeholders could have access to the underlying principles, workings, and rationales of the system (Choung et al., 2023). Every stage of the AI deployment process is susceptible to bias due to the intricate design of the algorithms, insufficient and unbalanced data, and software engineers' subjective choice of variables. Consequently, techniques for bias detection and mitigation must be integrated in order to continuously monitor official model characteristics, conduct fairness checks throughout data preparation phase, and subsequently look for and rectify clouds where information may leak (Gill et al., 2022). Lastly, the data

stewardship principle emphasizes that any data applied in AI must follow proper protocols regarding consent. Moreover, systems aiding in customer due diligence, transaction monitoring, and financial crime risk assessment must abide by data regulations (Giralt Hernández, 2024).

4.1. Fairness, accountability, and transparency

An AI system must be fair: its deployment requires consideration of fairness in the training data, algorithm, and model selection; and fairness also relates to its access and usage. It must be accountable: organizations must ensure their AI systems are developed, deployed, and utilized in a manner that aligns with the institution's guidelines and policy objectives. It must be transparent: organizations must make efforts to add transparency to their AI systems and modelling processes, enhancing the understanding and comprehension of these systems and helping to intuitively understand their risks, limitations, and impacts.

Fairness in AI encompasses: fairness of outcomes; group fairness, individual fairness; equality–equity distinction; disparate impact; are the trained models fair, and; fairness in data preparation (Choung et al., 2023). To support fairness, environment variables may be analyzed; available frameworks may apply such as Fairness-Aware Machine Learning (FAM), and a rules-driven approach may operate. Rules may cover the features criteria of all variables, the preservation of fairness across pipelines, and the consistency of decisions throughout the system life cycle. In addition to internal impacts, robustness against availability and security risks, and the sustainability of embedded materials within the pipeline—

4.2. Bias detection and mitigation

Bias detection and mitigation focus on identifying ethical risk factors such as data risk and technology risk that cause bias in AI decision making. Uncertainty in technology and incomplete data can lead to bias and ethical problems. Management failures can result in social risks like unemployment. Incorporating risk management elements into the risk feedback model can significantly reduce algorithm, technology, and data risk rates, thereby effectively decreasing social risks (Guan et al., 2022).

4.3. Privacy, consent, and data stewardship

AI/ML models and systems often require vast amounts of data that contain personal information, thereby creating risks related to privacy and data protection (Choung et al., 2023). Privacy risks can arise from combining datasets with publicly available datasets, and active surveillance and monitoring can lead to violations of the data subjects' right to privacy, creating ethical challenges.

Establishing effective data governance means implementing standards for the management of sensitive and personal information, such as the collection, use, access, dissemination, retention, and confidentiality of data. Recognizing and identifying data privacy and usage risks is crucial for setting appropriate controls and safeguards. Preparing documents that promote transparency about personal data collection, use, and retention can help mitigate risks surrounding the misuse of personal information.

5. Operational Risk Management for AI Systems

Advancements in artificial intelligence (AI) are reshaping various business processes within financial institutions, including audit and investigation, data monitoring, risk and controls, execution, and fraud analysis. Successful integration of AI systems necessitates the establishment of a governance framework to manage risks and ensure compliance with internal policies and external regulations. The framework, designed for both traditional and crypto-financial services and aimed at financial companies, addresses ethical, operational, regulatory compliance, and environmental, social, and governance issues. It specifies technical controls, cybersecurity mandates, and incident protocols in line with the Operating Guidelines on Digital Financial Technology and Cybersecurity issued by the Hong Kong Monetary Authority in December 2020. The Australian Prudential Regulation Authority's Information Security Prudential Standard (CPS 234) is also taken into account. Effective and responsible AI governance is

complemented by augmented digital capabilities and infrastructure, creating a robust model to meet customer objectives, comply with regulations, manage risk, and realise improvements through deployment (Kurshan et al., 2020).

Operational risk management (ORM) is a systematic approach that assists institutions in assessing operational risks, enhancing the resilience of existing infrastructure, managing the potential severity of incidents, and ensuring timely detection of early warning signals (Gill et al., 2022). ORM encompasses identification and assessment of the operational risk landscape, including inherent risks, existing controls, and vulnerability to emerging threats, as well as continuous monitoring of operating level compliance to comply with regulations and internal standards (Piorkowski et al., 2022). AI systems are generally classified into different categories according to the nature of their application. The challenges that financial institutions face in applying ORM principles to AI systems and the fundamental principles and concepts, risk management activities across the AI lifecycle, control considerations, monitoring design principles, and the establishment of AI risk appetite in the banking sector are discussed.

5.1. System reliability and safety engineering

Compliance with delivery, reliability, safety, and security, including cybersecurity requirements, is crucial for financial institutions. Operational risks arise from human error, failed internal processes, system failures, or external events. Consequently, operational risk assessment considers the likelihood of failure and its potential impact. Economic institutions aim to mitigate financial losses linked to such operational risk events, so the potential impact must reflect those financial losses. Economic institutions must evaluate whether a failure adversely affects their financial position, reputation, voluntary compliance with regulations, or adherence to ethical commitments. Regulators expect that economic institutions with a high market share possess reliable and safe AI systems. Institutions thus assess the operational risk and impact of such systems, irrespective of whether the system directly involves the regulator.

System reliability and safety should ensure that an AI system behaves according to specified objectives during its service life. Reliability according to high-level specifications facilitates system safety in compliance with organizations' ethical commitments. AI models must at least satisfy institutional requirements on the period before correcting noncompliances and on the maximum severity or class of avoided errors. Operating conditions significantly vary with time and locations, increasing the potential for unanticipated failures. Without adequate verification or validation, an AI system may remain in operation for a prolonged period, compounding the situation. An organization may require the ability to declare a system self-compliant and never issue related overriding instructions. Even with such declaration, and in order to ensure safe use of the system, verification comes necessary to clarify the corresponding high-level specifications, applicable environments, and elapsed compliance duration. The climate crisis prompted economic institutions to reinforce their ethical commitments in risk and pricing decisions. AI solutions can become central for implementing proper decisions, while numerous AI systems governing such decisions have been found misaligned with these ambitions (Kurshan et al., 2020). Yet, large data dependency complicates monitoring of compliance. Safety must thus cover ethical trajectories alongside functional objectives.

AI systems engage numerous cyber-threats. Efforts to obtain blocked information or modify a system's parameters persist; hence a broad understanding on vulnerabilities, threats, and security solutions remains vital. Robust augmentation offers a proactive alternative to handle various forms of data poisoning, yet such augmentation ultimately dilutes the information carried by the data. Performing some action on the input data and analysing whether the action earns the expected reward allows screening out a significant portion of competing strategies, without compromising stability and generalizability (Guan et al., 2022). Cyber-attacks and societal disruptions introduce additional operational risks where direct regulatory control remains unfeasible. Economic institutions take measures to detect such events.

5.2. Cybersecurity and incident response

Rapid digitalization characterized by advanced internet connectivity and novel Artificial Intelligence tools has led to a marked increase in cyber-attacks on Artificial Intelligence systems (R. McIntosh et al., 2024). Cyber attacks on AI systems are classified into four phases: (i) data poisoning on the training phase, (ii) adversarial attack at the inference stage, (iii) exploitation of a model backdoor, and (iv) extraction attack targeting model parameters and architecture (Polemi et al., 2024). Data poisoning on the training phase can be countered by good MLSecOps practices and risk management methods, while adversarial attacks can be mitigated by ensemble methods, defensive distillation, differential privacy, homomorphic encryption, and model watermaking. Model backdoors can often be revealed through rigorous testing, and extraction attacks can be thwarted by low-cost model access, restricting query limits, and implementing access control policies. General AI controls are divided into three categories: measures against security threats during development, security controls during runtime, and social mitigation combining technical and human oversight.

5.3. Monitoring, auditability, and explainability

Patterns of AI and machine learning (ML) are already underpinning many essential economic systems. Risk assessments are, therefore, the only course of action, highly contested though the desired deployment criteria and assurance barriers are. Tracking implementation details, meaningful output specifications, and residual shifts as the underlying world evolves are all steps taken in many operational models.

Active auditing has been proposed as central to reasonable AI governance. Models learn from the environment, yet these systems are held accountable to a broader mission that must be treated as a controllable variable in the absence of complete autonomous verification. Models operate in parallel to economic physical actions of physical damage. The effects of AI-greenhouse gas emissions are, therefore, as critical to monitor as years prior to and those following deployment.

Monitoring can specialize in certain common areas. The nature of possible critiques overestimates the scope of AI and plausibly misconstrues ignorance about its nature. Axes of AI critique may include system displacement, technological accumulations, warning frequencies, or utility validation. Monitoring specific activity on these axes promotes wider awareness of necessary scrutiny and leads to credible specialization.

Monitoring enhances the quality of corporate digital infrastructure and external accessibility, especially for lifecycle conversations regarding supplies or space frames and unlikely risks overestimation. Governance cannot replace uncertainty about the desired metric for a deployed AI system. Monitoring compatibility with its governance across diverse world architectures is nonetheless feasible and directs further inquiry toward a more satisfactory definition of the accepted metric.

Monitoring of AI model behaviour throughout crucial frame deployments still requires additional defined oversight in design, and both outside institutional deployment of first-generation models and launches of second-generation models into the global model pool continue to witness an absence of sufficient widely accepted monitoring focuses.

The governance surrounding the global atmosphere of AI interaction clearly stands outside a company contacting reinforcement-learning-based experimental models for beneficiary allocation to support early-phase pool understanding and encourages other platforms to provide exchange alternatives for those wishing to limit general monitoring.

The AI operational model established a modular framework combining oversight and post-deployment operations. Alignment with corporate institutional objectives became the centre of concern, viewed to integrate all framing instead of conflicting with it. Compliance with model-governance expectations

managed by experienced firms accessing, reworking and publicly releasing policies with and without AI figures across financial systems remained uncertain throughout the experiment.

The significant class of first-generation models capable of monitoring behaviour became actively swapped into even very early framed exploration. All direct corporate permitting, even to strategic competitors with ongoing surveillance, was overtly requested to avoid misunderstanding of model capability or pursuit of self-routing figures beyond the corporate goal framework (Kurshan et al., 2020). Model behaviour spans allocations and synthetic capture. Governance, however, remains the mediated goal widely visible throughout the first-generation exploration. Sufficient cross-competitive circulation of sufficiently diverse contributor-origin meta correspondences outside real-time model contact and the expanded release of singular alternative synthetics within that framework had not represented the level of openness deemed conditionally appropriate. Multiple concurrent contributor organisations, including corporations with strictly aligned modelling objectives, encountered extensive reciprocal distribution without necessarily incurring clarity erosion.

Even generative modelling remains possible through pooled observability. World-class and widely unique framing perspectives on populations across multiple ideological axes remain available from platforms holding substantial historic fluency, including leading competitors pursuing fully aligned grand agendas (Choung et al., 2023).

6. Regulatory and Compliance Landscape

The fast-paced evolution of artificial intelligence (AI) technologies has led to a situation in which many jurisdictions are actively exploring new AI regulations to mitigate the risks associated with their deployment across a wide range of applications. In Canada, guidelines issued by the Office of the Privacy Commissioner, known as the Proposals for a Framework for the Regulation of Automated Decision-Making, aim to govern the use of AI technologies in the financial services sector and secure the protection of personal data (Kurshan et al., 2020). The Ontario Securities Commission has also published guidelines around AI use in the investment management industry. Economic institutions therefore face a pressing need to implement AI governance frameworks that address the specific guidelines enacted by supervisory bodies and maintain a clear alignment between business objectives and ethical principles.

When electronic systems manage the movement of data or the generation of outputs in potentially biased or unfair ways, the requirements set forth by domestic and international supervisory agencies to ensure AI fairness, ethics, and safety become critically important. These requirements apply to internal systems as well as any third-party systems that are used on behalf of stakeholders, clients, or staff. Instances of unfair bias in AI systems can generate allegations of untrustworthiness against economic institutions and render them liable for financial loss or reputational harm owing to the adverse operational risk that accompanies unaddressed fairness requirements. (Varona & Suárez, 2022)

6.1. Supervisory expectations and compliance requirements

Institutions are subject to compliance with a comprehensive and rapidly evolving framework of regulations and supervisory guidance pertaining to AI and machine learning (ML). Regulatory expectations are set forth by international standard-setting bodies as well as national and international regulators and supervisory authorities. Guidelines extend beyond the financial sector and cover the broader technology and digital economy realms. Key expectations include adhering to principles of fairness, transparency, accountability, and responsible data treatment. Required assessments encompass risk not only to data but also to affected individuals, society, business sustainability, and fair competition (Kurshan et al., 2020).

6.2. Cross-border data flows and sovereignty

Data access and sharing across institutional borders is a prerequisite for a flourishing digital economy. At the same time, national and regional governments are taking data sovereignty and security measures to protect strategic national assets. Empirical findings indicate that countries with stricter data sharing regulations have lower cross-border data flows (Schneider et al., 2020). Such controls can impede international trade in ideas and research and hamper the advancement of technologies essential for improving productivity, competitiveness, and societal wealth (Choung et al., 2023). The formulation of legislation is often accompanied by sweeping cross-border data transfer restrictions. The aftermath of conflicts, however, has induced nations to safeguard local economic development and fundamental human rights, forcing synthetic data and commercial data to undergo such transformations that they cannot risk transpiring data on individuals, sensitive matters, or trade secrets. Data control and sharing become subject to bilateral treaties among such host and domicile nations to prevail over terminologies or secure raw input/output without disclosing algorithm or model structure.

6.3. Reporting, disclosure, and forensic readiness

Governance systems must also ensure proper procedures for recording, archiving, reporting, and disclosing AI activities, and that institutions are consequently ready for forensic investigatons. These activities must also address protection granted under attorney-client privilege and legal professional privilege (Kurshan et al., 2020) and any accelerate relevant steps prepared before the publication of an AI-generated report (Gill et al., 2022).

Institutional governance systems must also guarantee that AI-governed processes and records show engagement with, and impacts on, the institution's overall mission; business strategy; and organization of governance, risk and management across all systems.

7. Integrated Risk Management Framework

AI systems entail new risks that differ in nature from traditional technologies, requiring tailored management solutions. An integrated risk management approach, applicable at both enterprise and functional levels, focuses on ethical, operational, and regulatory compliance risks associated with AI deployment (Kurshan et al., 2020). A strategic institutional understanding of overall risk governance enables the establishment of context-specific boundaries for these interrelated concerns. The identification of common underlying causes across ethical, operational, and regulatory dimensions facilitates effective coverage of AI-related risks.

Occupying the forefront of regulatory concern, ethical risk manifests in direct or indirect harm to clients or society through the implementation of AI models. Operational risk is engendered by the functioning of AI systems, which may deviate from intended initialization or exhibit unpredictable behaviour. As AI models evolve and the operational landscape changes, continuous re-evaluation of ethical and operational risk exposure is necessary to maintain regulatory compliance.

7.1. Risk assessment methodologies

Assessing AI risk stems from analyzing operational characteristics, ethical considerations, and regulatory compliance. Contacts with appropriate stakeholders help ascertain specific concerns across these categories. Workshops and reviews of prior assessments often assist in the process (Piorkowski et al., 2022). Furthermore, an extended consideration of domain-specific scenarios based on an extensive set of risk categories may broadens understanding of residual risks and mitigation strategies (Kurshan et al., 2020).

7.2. Control design and governance mapping

To demonstrate effective integration between institution-wide objectives, stakeholder interests, operational practices, and risk management frameworks, a framework for the design and mapping of controls is proposed. Institutional AI control frameworks should define a governance structure

comprising the specific governance bodies and stakeholders expected to oversee and directly engage with AI development and deployment activities, and outline the opposite mapping from the various risk categories identified in the risk taxonomy to the mandatory governance attendance requirements set out by these stakeholder groups.

Development of such an integrated, multi-risk-governor framework would permit significant acceleration of the AI governance strategy and deepening of its embedding within the institution. The governance structure mapping exercise also enables institutions to employ the growing array of operational component design specifications and assessment templates explicitly generated to target anticipated ethical, operational, and regulatory AI deployment requisites and thereby to minimise bespoke development. The broad framework design and mapping examples subsequently proposed as part of an adjunct governance accelerant would therefore allow institutions to implement a comprehensive operational risk framework addressing AI operational integrity, data privacy and safeguarding, system security, cyber-resilience, and traceability and auditability standards within the pre-established analytical categories and directionally aligned with regulatory repository risk classifications (Kurshan et al., 2020).

7.3. Metrics, monitoring, and continuous improvement

AI governance in economic institutions should be framed as an integrated, evidence-based framework that analyzes ethical, operational, and regulatory compliance risks within institutional objectives and stakeholder interests.

A growing body of literature recognizes the importance of establishing metrics and monitoring mechanisms to facilitate the evaluation of AI governance/control frameworks (Choung et al., 2023). An effective governance framework should not only enable informed decisions on risk appetite and design controls, but also allow for ongoing assessment of compliance with such requirements. Continuous improvement and refinement of the control design must reflect evolving assessment outcomes and changing operational requirements, data characteristics, business priorities, and regulatory domains. Monitoring and review processes should be regularly scheduled to ensure that control measures remain applicable over time.

AI governance/control frameworks at financial institutions must also foster an authentic culture of accountability. Promoting clear articulation of AI governance/control objectives, stakeholder ownership, accountability for overseeing the implementation of governance/control solutions, and executive buy-in to define and cascade the risk appetite heightens the potential for successful establishment and sustained operation (Kurshan et al., 2020).

8. Policy Design and Strategic Alignment

AI strategies in economic institutions must continuously evolve to remain relevant to society's changing expectations and adhere to the institutions' broad objectives and risk appetite (Kurshan et al., 2020). As stakeholder interests diversify and technical options proliferate, the design of AI policies calls for both a systematic approach and coordination across institutional silos. Many institutions operate AI systems in silos and have implemented policies ranging from responsible AI principles to bias-mitigation techniques for data, algorithms, models, and outcomes. Consequently, institutions often have multiple responsible AI frameworks, but the frameworks are rarely aggregated into a single coherent strategy or operationalized through automated policies (Gill et al., 2022). Institutions and technology ecosystems are developing rapidly in parallel with AI policy, and internal, external, and collective oversight mechanisms for AI governance are being integrated into broader governance strategies overseen by other organizational agendas (Choung et al., 2023).

8.1. AI strategy alignment with institutional objectives

A soft copy of the "Integrated Framework for AI Governance in Economic Institutions: Managing Ethical, Operational, and Regulatory Compliance Risks" is available at [Link]

AI strategy alignment with institutional objectives

AI is integrated into economic and financial institutions with increasing intensity, fueled on the supply side by technology diffusion and on the demand side by the unyielding quest for improved efficiency, lower costs, higher profits, and enhanced customer experience. The widespread urgency to adopt AI at scale is perceived as a key threat to value generation as AI services deployed by competitors threaten to disrupt existing service offerings, render prior investments virtually valueless, or—most significantly—spark a loss of competitiveness, customer attrition, and a shriveling business. Institutions are thus pressured to deploy AI strategies and align policies against the crystal-clear backdrop of institutional objectives (Gill et al., 2022). Institutions undertake heavy investments in technology, infrastructure, capability, and talented personnel not to improve AI systems or services in some abstract sense but to align AI-made decisions, operating procedures, and service offerings with institutional objectives.

Considerable rhetoric surrounds "Rubin" governance and "Renaissance" governance. An open question looms regarding whether the rhetorical examples constitute examples of real governance—and whether extant frameworks indeed provide operationalized counterparts for economic and financial value generation. Institutions are yet to receive formal reckoning regarding how AI might augment the pursuit of economic and financial value across the full range of portfolio management domains, only outside the routine confines of exhausted institutional programming (Kurshan et al., 2020).

8.2. Ethics-by-design and policy automation

Ethics should be taken into account at the earliest possible stages of every artificial intelligence (AI) project and policy cycle. Such considerations should not be treated as an afterthought or as mere add-ons to ongoing technological developments (Choung et al., 2023). Ethics-by-design combines the rigour of governance with a trust-based approach, aiming to balance robust oversight and stakeholder participation in AI deployment. This perspective contrasts with frameworks based purely on regulatory compliance or algorithmic fairness, which may inadvertently promote complacency and lead to a "check-the-box" approach that undermines effectiveness. AI governance is best understood as a process of negotiation and practical deliberation within institutions pursuing substantive objectives; it is not a commodity, a simple recipe, or a set of universal principles (Morley et al., 2021). Ethical assessments can be streamlined and scaled through operative systems within the economic institution, resulting in preallocated guidance and feedback. Policy automation involves the systematic translation of institutional objectives, stakeholder interests, regulatory obligations, and ethical principles into operational directives for relevant processes, procedures, and activities. The content of these policies should be established on the basis of comprehensive and independent assessments of AI projects prior to deployment. Further iterations of policy design are feasible to enshrine complementary principles—such as considering the adaptation of service offerings to better align with institutional, stakeholder, and regulatory expectations.

8.3. Vendor and ecosystem risk management

The acceleration of digital transformation over the past decade has led to the convergence of data analytics and machine learning technologies, collectively known as artificial intelligence (AI). Technology leaders have embraced AI due to its capacity to analyze massive datasets, automate complex decision-making processes, and enhance productivity. AI applications have the potential to create substantial value for companies and society at large. These benefits range from replacing mundane manual tasks to fundamentally changing the characteristics of financial services, for instance, by accelerating decision-making processes, reducing operational costs and risks, and creating a more personalized customer experience;. Whilst organizations and entities have dramatically increased their adoption of AI technologies, AI also raises serious ethical, operational, and compliance-related risks that must be carefully assessed and independently monitored throughout the entire AI life cycle;;;.

Restrictions can take different approaches such as the design of AI strategy, or the definition of an interdisciplinary governance board (Guan et al., 2022) ; (Kurshan et al., 2020) ; (Piorkowski et al., 2022). An organization is part of a broader economic ecosystem often comprising clients, suppliers, vendors, strategic partners and/or third-party agents. Exposure to risks unrelated to the organization's internal environment, but linked to these associated parties, is therefore often described either as ecosystem risk or vendor risk. Organizations that deploy AI systems may also partner with vendors to develop effective monitoring, compliance and governance controls that can support the broader adoption of AI;.

9. Implementation Pathways and Capability Building

Providing a pathway to operationalising AI governance in economic institutions requires a methodology for assessing the maturity of AI governance capabilities, identifying existing gaps, and formulating roadmaps and milestones for institutional implementation. Multiple maturity models have been proposed to achieve this, and once the roadmap is established, an updated inventory of roles, skills, and responsibilities can facilitate necessary organisational change (Choung et al., 2023).

Complementing the evaluation of governance capability and maturity is the establishment of a comprehensive programme for building AI risk governance and compliance skills and awareness across an institution. Such a programme comprises training curricula tailored to specific roles and tailored community forums and events spanning the institution's AI ecosystem. These efforts contribute to nurturing an institutional culture inclusive of AI governance and risk management, a process that can drive awareness of associated opportunities and challenges and further accelerate the hunt for AI-enabled value (Kurshan et al., 2020).

9.1. Maturity models and roadmaps

Maturity models became popular in the 1990s when software engineering applied the Capability Maturity Model to quality management. Widely used today by organizations evaluating information technology, project management, business maturity, and more, maturity models help identify strengths and weaknesses, pinpoint areas for improvement, establish priorities, and measure progress (Kurshan et al., 2020). Criticized by some researchers as oversimplifying the complex process of organizational change and fostering complacency, maturity models nevertheless present advantages: they stimulate debate and encourage goal setting and proactivity (Vakkuri et al., 2021). Several organizations have proposed AI-level maturity models, and an industry consensus has emerged on factors necessary for AI-level maturity (Dotan et al., 2024).

To progress toward responsible AI, economic institutions need a maturity model that captures the organization's overall AI-risk management capability, completion of which becomes the goal. Existing maturity models focus primarily on research and development or portfolio processes, limiting their applicability to broader enterprise use. An organization-wide and operation-oriented maturity model can articulate responsible AI for economic institutions—articulating the institution's capabilities and designated responsible AI classification and compliance for each AI system across the AI-life cycle.

9.2. Roles, skills, and organizational change

As institutions transform into AI-driven organizations, the implications for governance roles, skills, and structures are significant and wide-ranging, affecting operational, regulatory, and ethical dimensions (Choung et al., 2023). Changes should be scalable according to the AI maturity of the institution and systemic adjustments may be necessary in the longer-term.

Specific roles and skills provide a foundation for systematic capability building over time. For operational risks, the existing roles of system and operational risk may add AI-specific skills that cover all risks during development and deployment. For regulatory risks, subject matter expertise can assist in identifying applicable frameworks, and operational modeling skills can support interpretation of needs.

For ethical risks, many of the teams supporting existing frameworks may extend their focus to cover AI topics. Additional staff may be needed depending on AI maturity and coverage.

Since these governance changes may impact traditional power dynamics, senior management and board backing is essential, ideally via collective resource capacity models that maintain existing coverage while adapting to AI developments. Institutions may also consider introducing AI-specific committees focusing solely on the relevant risks.

9.3. Training, literacy, and culture

Training initiatives should be customized to address the specialized needs and responsibilities of each participant group, incorporating pertinent scenarios to demonstrate the relevance of an integrated AIG framework. A foundational curriculum targeting the entire workforce can introduce organizational AIG principles and foundational concepts, further strengthened by supplementary resources and learning opportunities tailored to UX designers and development engineers involved in AI system design, AI-enabled products and services, or data assembly and processing. Such a strategy fosters an understanding of ethical implications based on institutional objectives and stakeholder expectations, ensuring consideration of AIG objectives from the earliest phases of the AI system development lifecycle.

Designing AI systems and managing data collection and processing in line with institutional values, stakeholder preferences, and applicable regulations requires ongoing reinforcement of knowledge and skills among specialists operating in these domains (Giralt Hernández, 2024). Promoting organizational AIG knowledge and expertise contributes to a shared understanding of objectives and challenges, guiding decision-making on investment and resource allocation, and shaping internal discussions with peers and external partners. An integrated AIG framework enjoys maximum efficacy when practitioners throughout the organization recognize its purpose, appreciate associated dilemmas, and commit to procedures that support compliance. Continuous reinforcement of teachings and principles fosters a culture of responsible AI practice, motivating personnel to embrace AI's power for positive societal advancement while monitoring the accompanying risks to the organization and its customers.

10. Evaluation and Accountability Mechanisms

The AI Provision Explainer stipulates that the AI Provisioning component should include an evaluation of AI provisions established by universities and research institutes, together with continuous evaluations for AI models. It also specifies that institutions must disclose when they have developed AI systems that automate or augment substantial decision-making processes. Regulatory goals, target dates, and updates on AI-related projects must be communicated with stakeholders. The operationalisation of Ethical AI fosters accountability and engagement with external stakeholders by delivering precise project reports designed according to the AI Provision Explainer.

Regulatory expectations on reporting, disclosing, and seeking prior consent are projected for automated or augmented scheduling and recommender systems. External stakeholders are thereby informed about properties of models and their intended uses. Timely and continuous status updates, aligned with operational capabilities and model risk classification, remain a general programming provision. Plans for release, including target date or known delays, receive special attention. Definitions for these systems and criteria to prioritise further clarification are under development.

Requirements about reporting, auditing, and documentation are anticipated on supervisory matters in the context of AI models. These obligations are relevant both prior to deployment and at general release in a complex environment. An approach is being crafted to allow distinct updates on deployment, usage, and verification, while also catering to (Kurshan et al., 2020) specific statutory and business consideration.

10.1. Independent reviews and audits

Many economic institutions have adopted AI technologies to enhance the quality and speed of decision-making processes, thereby improving efficiency and competitiveness. Nevertheless, concerns about compliance, risk management, and ethical considerations in deploying and operating such technologies have arisen. For instance, a set of principles has been developed that involve concerning fair, responsible, and transparent Technology to integrate 'human' factors into the 'machine' learning process. Ethical AI use by Banks is either a legal obligation or a voluntary commitment determined by an organization's values and accepted by their stakeholders. The utilization of AI systems that could have a significant impact on the economy, financial stability, and customer protection is associated with ethics. Such measures account for the potential rise in human-like risks, such as a growing conflict of interests. Central Banks have analyzed external developments and techniques to address these issues (Kurshan et al., 2020).

10.2. External reporting and stakeholder engagement

AI governance in economic institutions should be framed as an integrated, evidence-based framework that analyzes ethical, operational, and regulatory compliance risks within institutional objectives and stakeholder interests.

An increasing number of jurisdictions are enacting laws and issuing guidelines that mandate disclosure of information related to the use of AI (Choung et al., 2023). An integrated approach to AI governance must therefore include consideration of what to report externally, to whom, and at what frequency. Guidance on external reporting is rapidly evolving, and stakeholders frequently desire more detail than what is otherwise communicated. To navigate this landscape, institutions should start by engaging with their stakeholders and determining what AI practices and impacts most concern them. Operational and ethical risk governance then provides a structure to identify, assess, and describe relevant AI practices and the potential impacts of both the systems deployed and the institution's commitment to managing associated risks throughout the AI lifecycle (Kurshan et al., 2020).

10.3. Remedies, redress, and escalation procedures

Well-designed and operated AI systems can still produce undesirable outcomes due to uncertainties, limitations in the modeling and data, or ill-specified objectives, requiring a robust approach to monitor and mitigate possible adverse or unintended consequences. Risk management capabilities remain incomplete for AI systems despite efforts to promote a principled approach to responsible AI. Both formal and practical considerations call for organizations to pursue minimal to moderate risk profiles in implementing AI initiatives. In addition to ongoing developments in model, dataset, and performance monitoring in line with regulatory objectives, and in light of workforce capacity expansion for building and operating AI systems, the promotion of oversight and reviews of AI initiatives driven by governing bodies constitutes an integrated approach to addressing these ethical and operational risks.

Collaboration between governing bodies and functional areas also supports and aligns independent evaluations with a comprehensive understanding of operations and coordinated responses to regulatory engagement or actor influence. Governance participants—including internal review boards, ethics boards, and other committees with relevant domain and societal expertise— warrant inclusion in independent scrutiny of AI systems. Additional safeguards contribute to responsible, effective, and sustainable AI initiatives (Kurshan et al., 2020) within the industry while establishing principles for transparent and equitable treatment of market participants and vulnerable populations (Choung et al., 2023).

11. Conclusion

While artificial intelligence is a rapidly growing form of technology with unbounded potential, its capability is constrained by the ethical, operational, and compliance risks posed by the individual financial institutions implementing its use. The purpose of this document is to discuss the approach

taken to design an integrated framework of analysis tasked with assessing these risks associated with AI-related activities across a financial institution, and discuss the recommendations that arise from the investigation into these risks. Financial institutions have access to powerful AI tools that must be deployed with extensive risk analysis, designed within an integrated framework that encompasses AI-related ethics, operations, model governance, and business compliance to avoid power abuses and damaging effects on stakeholder communities (Gill et al., 2022).

References:

[1] Choung, H., David, P., & S. Seberger, J. (2023). A multilevel framework for AI governance.

[2] Kurshan, E., Shen, H., & Chen, J. (2020). Towards Self-Regulating AI: Challenges and Opportunities of AI Model Governance in Financial Services.

[3] Gill, N., Mathur, A., & V. Conde, M. (2022). A Brief Overview of AI Governance for Responsible Machine Learning Systems.

[4] Schneider, J., Abraham, R., Meske, C., & vom Brocke, J. (2020). AI Governance for Businesses.

[5] Suksi, M. (2023). The Rule of Law and automated decision-making: Exploring fundamentals of algorithmic governance.

[6] Piorkowski, D., Hind, M., & Richards, J. (2022). Quantitative AI Risk Assessments: Opportunities and Challenges.

[7] Kahler, M. (2024). From complex interdependence to complex governance. Informal governance in world politics.

[8] Gong, J. J. & Zhang, V. Y. (). China's Evolution in International Standardization: From Follower to Global Leader. connaissancedesenergies.org.

[9] Benthall, S. & Shekman, D. (2023). Designing Fiduciary Artificial Intelligence.

[10] Giralt Hernández, E. (2024). Towards an Ethical and Inclusive Implementation of Artificial Intelligence in Organizations: A Multidimensional Framework.

[11] Guan, H., Dong, L., & Zhao, A. (2022). Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making.

[12] R. McIntosh, T., Susnjak, T., Liu, T., Watters, P., Nowrozy, R., & N. Halgamuge, M. (2024). From COBIT to ISO 42001: Evaluating Cybersecurity Frameworks for Opportunities, Risks, and Regulatory Compliance in Commercializing Large Language Models.

[13] Polemi, N., Praça, I., Kioskli, K., & Bécue, A. (2024). Challenges and efforts in managing AI trustworthiness risks: a state of knowledge. ncbi.nlm.nih.gov

[14] Varona, D. & Suárez, J. L. (2022). Discrimination, bias, fairness, and trustworthy AI. Applied Sciences.

[15] Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mokander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI Ethics.

[16] Vakkuri, V., Jantunen, M., Halme, E., Kemell, K. K., Nguyen-Duc, A., Mikkonen, T., & Abrahamsson, P. (2021). Time for AI (Ethics) Maturity Model Is Now.

[17] Dotan, R., Blili-Hamelin, B., Madhavan, R., Matthews, J., & Scarpino, J. (2024). Evolving AI Risk Management: A Maturity Model based on the NIST AI Risk Management Framework.