FIDUCIARY ROLES IN AI GOVERNANCE

Abstract

AI governance has emerged as a core element of boardroom responsibility. Compliance is an evolving fiduciary obligation that requires real time monitoring of AI systems' ethical, legal, and operational impacts. Boards are central actors in risk mitigation and governance. Fiduciary care and protection have the potential to transforms AI governance from a compliance obligation into a strategic asset, empowering organizations to innovate responsibly while safeguarding their legitimacy in a rapidly evolving digital economy. In this backdrop an attempt has been made in this article to highlight the devolving additional AI related fiduciary duties and responsibilities in the midst of exponential pace of increase in uses and embedment with existing solutions of AI tools and devices by businesses to reap its inherent potential for more efficient value creation.

Introduction

AI has persistent, broad and global consequences that are transmuting societies, economic sectors and the ecosphere of work, and are poised to progressively do so in the future. AI has inherent potential to deliver greater welfare and well-being of mankind, to enhance and strengthen constructive enduring global economic activity, to intensify innovation and to increase productivity, and to help meeting key global challenges. Concomitantly, these evolutions might impact differently various segments of societies and economies particularly regarding economic shifts, competition, transitions in the labour market, inequalities, and implications for democracy and human rights, privacy and data protection, and digital security.

Failing to incorporate AI into decision-making could also constitute a breach of fiduciary duty.

Biplab Chakraborty
General Manager (Retd.)
Reserve Bank of India, Kolkata
biplabchakraborty@yahoo.com

AI's potential to create value will become an avenue for differentiation as management continues incorporating AI even deeper within their operations. But with this demand comes increased regulatory scrutiny of AI practices from regulatory agencies and international authorities.

Intrinsic features of AI as driver of fiduciary risks & responsibilities

AI is being increasingly used in applications for wide ranging products and processes viz., IOT, healthcare, automobiles, business and data analytics etc. Main objectives of using AI are to Improve productivity and efficiency, support regulatory compliance and risk management and enhance core business/revenue generating activities. Entities / institutions use AI to do things faster, economising on costs and, to do things which humans cannot do with the accuracy and high speed that AI can deliver. AI by virtue of its proficiencies to carry out complex analyses and computations at a speed beyond the capacities of humans generates quicker insights. Significant benefits out of the AI abilities to generate and process voluminous data at high speed have accrued to Businesses and insurance.

While use of AI has the potential to accrue transformative benefits to the user entities it may also exacerbate existing risks. While AI might mitigate some elements of risks, it also brings along with it some new species of exposures and risks. This may be partly attributable to the fact that AI depends

largely on human feed huge volume of data.

Misuse of AI can also impair the very basics of any organization's business model: its brand popularity and reputation. There is no dearth of evidence of biases in AI-driven outcomes.

AI systems are prone to malfunctioning and failures arising due to improper maintenance, design defects or human error. These defects may trigger financial loss, property damage or bodily injuries (viz., effects of malfunctioning of AI in autonomous vehicles, health care & medical devices, Industrial robots &manufacturing, Aviation & transport system, every day consumer devices etc.)

Generative AI has exhibited proclivity to respond to questions with "hallucinations"—plausible-sounding answers that are factually incorrect or misleading. AI can describe a non-existent product or issue harmful product instructions. Such inappropriate information may make organisations liable for deceptive marketing or for injuries caused by defects in the AI components built into their products. They generally fall in systematic hard to predict ways due to data, modelling or other blind spots.

AI designers keep secret their algorithms rendering it difficult to identify the root causes of errors. Insureds, in turn, may not fully understand risks when purchasing AI products. Thus, more often than not insureds while buying AI products cannot see through the associated risks and the insurers cannot differentiate between unintended errors to be covered and intentional acts that have to be excluded from coverage. This misperception vitiates precise risk assessment and accurate pricing.

Using AI Often requires setting up of new interfaces (like API or web service) that are accessible from outside the organisation. These end points let the people or the systems to query AI. From business point of view these new public end points expand the opportunity horizon. However, they also increase the attack surface since each new public end points could be potential source of security or privacy risk if not managed properly and thus entail greater risks.

Failing to take appropriate preventive safeguards against such attacks can result in breach of fiduciary duties and devolvement of civil liability. Further, in speedier data processing (e.g., high frequency trading, real time risk calculation, AI inferences at scale) there would be less time available for validation, error checking and quality control and consequently the chances of errors slipping through increase. Faster the data are pushed through a system the greater would be the need for error detection and correction mechanism, because natural tendency in such scenario is for errors to multiply. All these give rise to more complex vulnerabilities and more potential legal violations.

Boards and top management have fiduciary responsibilities and duties to all stakeholders of the organisation. The directors and senior officers must exercise due diligence and prudential cautions in good faith while making hard choices based on reasonable information. Adherence to these principles encourages innovation and promotes higher risk taking to reap opportunities of high returns

Dependence of the Board on faulty AI tools while making business decisions, might breach its fiduciary duties. This is more likely when the goals and values of AI, the corporate, the shareholders and the data subjects are not aligned.

When a consultant is appointed to advise a company on AI-related matters, the fiduciaries still hold ultimate responsibility. They can delegate tasks, but not accountability. The fiduciary should ensure that the consultant's advice is critically reviewed, not blindly accepted. They should ensure exercising due diligence that consultant recommended AI adoption serves the interests of all stakeholders and prevent conflicts of interest (e.g., consultant also selling AI tools for profit). The fiduciary should verify compliance of consultant's recommendations with AI regulations (e.g., EU AI Act, GDPR, India's DPDP Act). Fiduciaries must monitor how the consultant's AI recommendations are implemented. While doing so it must over see whether ethical guidelines, risk controls, and impact assessments are followed and data subjects' rights (privacy, fairness, transparency) are protected.

AI can turn out to be a two-edged sword for cyber resilience. While AI can meaningfully fortify cyber security by proactive sensing of threats and finding vulnerabilities at the same time, cyber criminals can use similar AI tools to conduct more sophisticated cyber-attacks.

LLMs are evolving from powerful text models into complex multimodal systems that unify language, vision, and audio, unlocking new capabilities

while introducing new risks. This transformation is reshaping AI's role—from a text-based assistant into a context-aware partner in human and machine ecosystems. Deepfake technology will appear more resounding, and more manual functions will be automated.

As deepfakes proliferate identification management will pose governance issue for the Boards and the investors.

AI's rapid integration is reshaping the foundations of both legal and insurance frameworks. Legal systems must redefine doctrines of liability, authorship, and privacy, while insurers must develop new models and products to capture novel risks. The convergence of technology, law, and risk transfer mechanisms demands a proactive approach: embedding explainability, accountability, and fairness into AI systems.

AI's rapid integration is reshaping the foundations of both legal and insurance frameworks. Legal systems must redefine doctrines of liability, authorship, and privacy, while insurers must develop new models and products to capture novel risks. The convergence of technology, law, and risk transfer mechanisms demands a proactive approach: embedding explainability, accountability, and fairness into AI systems. Ultimately, those who adapt legal codes and insurance practices to responsibly manage AI risks will foster trust, innovation, and resilience in the digital economy.

Regulatory change is also inevitable, whether that means more jurisdictions with comprehensive AI regulations or allocating liability for AI harms to its human controllers.

Board-Level AI Risks

i. Strategic Risks:

a. Overreliance on AI without human judgment in core decisions (credit approvals, hiring).

b. Missing competitive threats from AI-driven disruptors. The competitive threat from AI driven disruptor is less about technology and more about the pace of business model innovations they enable. Missing it can turn strong incumbent into legacy players almost overnight. For example, disruptors using AI offer hyper personalisation ,predictive services, and frictionless experiences. Missing these shifts may render incumbent appear outdated even if their core product is strong. AI system might optimise operations, personalise offerings and automate decision making far faster than legacy firms. Incumbents that depend on slow product cycle or outdated infrastructure may fail to keep pace. Boards and executives may underestimate AI's transformative impacts treating it as a tool rather than core driver of new business model. This leaves scopes for disruptors redefine value chains.

ii. Operational Risks

a. AI hallucinations, errors, or poor data governance leading to financial losses. A hallucination transforms from a technical glitch into a systemic risk when embedded in a critical decision-making systems without proper oversight.

b. Vendor concentration risk (few big providers dominate). Vendor concentration risk in AI is not just a supply chain problem. It is a systemic vulnerability that affects resilience, cost stability, and strategic independence.

iii. Compliance & Regulatory Risks:

Compliance and regulatory risk arise from the gap between rapid technological adoption and slower, fragmented framework. Organisations must treat compliance as moving target, not a one-time check. Specific attention need be bestowed to the following:

a. AI-specific regulations (EU AI Act, U.S. AI executive orders, India's evolving digital rules).

b. Cross-border data transfer and privacy laws (GDPR, DPDP Act in India)

c. Algorithmic accountability requirements

iv. Ethical & Reputational Risks:

Ethical risks are about doing harm; reputational risks are about being seen as harmful. Together , they can undermine AI adoption, customer trust and long term business sustainability. Ethical lapses trigger reputational fallout. Reputational damages may persist longer than fines or compliance costs eroding competitive advantages. The following aspects would require close attention.

a. AI bias or discrimination (in recruitment, lending, insurance pricing).
b. Deepfakes and misinformation tied to corporate brand: Deepfakes and misinformation tied to corporate brand can hijack a corporate brand's identity, distort public perception and inflict lasting reputational and financial harms. The best defence would be a mix of tech safeguards, governance protocol, legal readiness and stakeholders' trust building.
c. Stakeholder backlash against job losses or misuse of AI.

v. Cyber security& Input-Output Risk

Cybersecurity protects AI from external threats while I/O risk addresses vulnerability in what AI consumes or produces. Together they determine the trustworthiness and safety of the system. Models may leak sensitive data (prompt injection, model inversion). Infringement risks arises if AI outputs violate copyright/IP. Prompt injection refers to malicious user manipulating the input(the prompt) in order to override its instructions , bypass safeguards or make it reveal unintended information. Attackers insert hidden instructions inside prompt. The AI model interprets these as higher priority than its original safety rules and intended tasks. Model inversion is an attack where adversaries exploit a train AI model to reconstruct or infer sensitive information about its training data. Attackers repeatedly query the model and analyse output to reverse engineer private data it was trained on.

vi. ESG & Stakeholder Risk

a. AI's environmental footprint (energy-heavy models).

   AI environmental foot print spans energy, emission, water and materials from training to deployment. Its sustainability depends heavily on greener data centres, renewable energy adoption and responsible hardware life cycle. AI use can harm environment but it can also be used to reduce environmental harms.

b. Responsible AI demanded by investors, regulators, and society

   Responsible AI is no longer optional. Investors demand it for value protection, regulator mandate it for compliance and the society expects it for trust. Entities ignoring these demands would face financial , legal and reputational risks.

Fiduciary Duties

- **Duty of Care**

  Boards are expected to make informed decisions. If AI strategies are adopted in an organisation without sufficient understanding and comprehensive appreciation of the entailed risks (bias, explainability, regulatory compliance), it would amount to breach of duties on the part of the Board of Directors. It has to be ensured that AI systems are robust, accurate, and tested before deployment. Regular audit of AI models should be carried out for errors, bias, and cybersecurity vulnerabilities. Effective human oversight must be exercised in high-risk decision-making (finance, healthcare, HR)

- **Duty of Loyalty (Avoiding Conflicts of Interest)**

  Conflicts of interest must be scrupulously avoided by the Directors. AI must be deployed in ways that benefit stakeholders, not just for profits. Customer privacy and data rights need be protected eschewing exploitation of personal data unfairly. Avoidance of use of manipulative AI (e.g., dark patterns in e-commerce) would be a prudent strategy.

- **Duty of Oversight**

  Boards must exercise effective governance and oversee risk management and compliance systems. Failure to be agile and receptive to AI's impact on privacy, discrimination, cyber risk, or consumer protection could lead to devolvement of avoidable liability.

- **Duty of Accountability**

  Explicit disclosure should be made about when and how AI would be used in decision-making (e.g., loan approvals, hiring) need be ensured. Limitations of AI and potential risks must be clearly enumerated and disclosed . Implementation of clear governance structures as to who would be responsible if AI causes harm is advisable.

- **Duty of Compliance:**

  Compliance with and scrupulous adherence to related laws and regulations (GDPR, DPDP Act, EU AI Act, sector-specific rules) must be ensured. Emerging future possible AI regulations may be anticipated for proactive policy alignment.

- **Duty of Ethical Stewardship**

  Fairness and non-discrimination in AI-driven decisions need be encouraged and patronised. The fiduciaries should ensure that AI aligns with corporate values and social responsibility commitments. Evaluation of environmental impact of AI (e.g., energy-hungry large models) is the need of the hour.

What Boards Should Do?

a. Governance & Oversight
   - Establish AI risk committees or integrate AI into risk/audit committees.
   - Require management to maintain AI inventories (where and how AI is used).

b. Policy & control
   - Adopt Responsible AI principles (fairness, transparency, accountability).
   - Mandate explainability for high-stakes AI use cases.
   - Ensure third-party vendor audits.

c. Training &Expertise
   - Appoint or consult with a Chief AI Officer / AI Ethics Officer.
   - Bring AI literacy to the board (just as with cybersecurity).

d. Disclosures & Reporting
   - Transparent reporting on AI use, risks, and governance
   - Proactive stakeholder communication to manage reputational risk.

Mitigation of AI-Related Risks by the Board

We focus here specifically on how the Board can mitigate AI-related risks. The emphasis is on oversight, governance, and proactive frameworks. A structured summary is given below:

i. Governance & Oversight

AI Oversight Structures may be created. Towards this end a dedicated Technology/ AI Committee, may be constituted or alternatively responsibility may be assigned to the Risk/Audit Committee. Board-level accountability for AI decisions may be defined. The governance framework might prescribe and define the role of human intervention to mitigate detrimental outcomes from AI systems. Management should maintain a registry of all AI systems in use, with their purpose, risk category, and regulatory exposures.

ii. Policy & Frameworks

Unwavering Organisational commitments to fairness, transparency, accountability, explainability, and data protection should be made visible in the whole of the organisation. Human oversight in critical areas (credit scoring, recruitment, healthcare, compliance monitoring) should be mandated. Third-Party Risk Management assume greater important needing pointed focus. Intensive due diligence, on contracts entered into may be undertaken. Periodic audits for AI vendors and cloud providers may be mandated. Board must scrupulously monitor compliance on audit findings.

iii. Risk & Compliance Controls

AI may be integrated into the Enterprise Risk Management (ERM) framework. Regular review of emerging risks(bias, privacy breaches, cybersecurity, intellectual property, and ethical misuse) must be carried out. Alignment with global AI regulations (EU AI Act, U.S. AI directives, India DPDP Act) need be ensured. Anticipate cross-border data and algorithmic accountability obligations.

iv. Board Competence & Training

For AI Literacy of Directors board-level periodic trainings on AI basics, risks, and governance obligations may be organised. For Independent Expertise external experts may be engaged or an AI Ethics Advisory Council to provide guidance , oversight and accountability in responsible deployment and use of AI may be put in place.

v. Monitoring & Assurance

Regular independent audits of high-risk AI

systems may be commissioned. Explainability reports and bias testing results need be obtained and perused for corrective and strategic action.

Crisis management and disclosure protocols need be in place for AI failures, cyberattacks, or reputational crises.

vi. Disclosure & Stakeholder Communication

Transparent disclosure of AI usage, governance measures, and risk mitigations in annual reports, ESG disclosures, or issuance of sustainability statements may be ensured.

vii. Continuous Review

Regulatory Horizon Scanning need be carried out on continuous real time basis to Keep abreast of evolving global AI regulation. Treat AI governance as a recurring agenda item, with periodic updates on progress and risks.

The Board mitigates AI risks by embedding them into the core governance framework—not as a technology issue, but as a fiduciary and strategic risk domain. Proactive oversight, policies, audits, and transparent communication protect both shareholder value and corporate reputation.

Conclusion

AI creates a new fiduciary landscape. The fiduciary role in AI governance extends the traditional duties of care, loyalty, and prudence into the sphere of algorithmic decision-making wherein instead of a person weighing the facts, an algorithm process data and outputs a decision or recommendation with enhanced efficiency and consistency and scalability. The responsibilities of Directors and trustees extend besides financial oversight, also to ensuring that AI systems are transparent, auditable, and aligned with organizational and stakeholder interests. Deficiencies in governance of AI attributable to negligence in monitoring bias, failure to protect data subjects, or overreliance on opaque models have the potential to expose boards to direct legal liability under corporate and fiduciary law. This additional duty of oversight renders AI governance a core element of boardroom responsibility.

Beyond liability management, fiduciary AI governance provides a way forward to a state of sustainable competitive advantage. By embedding transparency, fairness, and accountability into AI strategies, boards can strengthen stakeholder trust, protect reputational capital, and differentiate themselves in markets increasingly sensitive to ethical and regulatory standards. Fiduciary care and protection thus have the potential to transforms AI governance from a compliance obligation into a strategic asset, empowering organizations to innovate responsibly while safeguarding their legitimacy in a rapidly evolving digital economy.

Boards treating AI as just an IT issue might face litigation, regulatory penalties, or reputational harm. Strong oversight, ethical alignment, and informed governance are the board's defence.

Considering the rapid development and implementation of AI, there is a need for a stable policy environment that promotes a human-centric approach to trustworthy AI, that fosters research, preserves economic incentives to innovate, and that applies to all stakeholders according to their role and the context.

Stephen Hawking said in 2016 at the launch of the Centre for the Future of Intelligence (CFI), "the rise of powerful AI will either be the best or the worst thing ever to happen to humanity. We do not yet know which." It all depends on how we harness it! MA

References

1. *AI Governance Handbook :A Practical Guide for Enterprise AI Adoption: By Sunil Gregory , Anindya Sircar*
2. *AI Governance Ethics: AI with Shared Values and Rules :Christoph Stuckelberger, Maria MerchanRocamora, Diviya Singh, Pavan Duggal(Editors)*
3. *Human Compatible: Artificial Intelligence and the Problem of Control: Stuart Russell*
4. *The Balancing problem in Governance of Artificial Intelligence: Tshilidzi Marwala*
5. *AI Governance Professional (AIGP) Exam Guide :Hemang Doshi*
6. *Regulating AI in the financial sector: recent developments and main challenges by Juan Carlos Crisanto, Cris Benson Leuterio, Jermy Prenio and Jeffery Yong(BIS)*
7. *Mitigating Board and Corporate Fiduciary Risks of AI :Richik Sarkar , Jarman J. Smith*