AI Governance

How organisations can develop and enhance their governance and control frameworks to minimise the risks and capitalise on the opportunities of AI

Contents

1 Introduction

AI is often framed as a complicated problem with a simple solution. In truth (as is often the case), the situation is more the reverse: the 'problem' of AI is not a particularly difficult one for the Board to understand. However, the 'solution' will require a multi-layered and multi-faceted response.

The UK's financial services regulator, the FCA, gave one of the best summations of the approach to AI when it described itself as a "technology-agnostic, principles-based and outcomes-focused regulator" – and organisations will benefit from taking the same view:

- Technology-agnostic – AI is a 'thing', like the Web or cloud-based computing. It can either be an opportunity or a risk, depending on its application.
- Principles-based – organisations should respond to AI by staying true to their strategy and mission, and operating within the guardrails of their Risk Appetite.
- Outcomes-focused – the key to success will be in understanding what the organisation wants to happen, and what the organisation wants to prevent from happening.

This outcomes-focused approach will considerably simplify the Board's assessment of AI. Any organisation will have a series of outcomes it seeks to achieve. These may not always be fully articulated (in the strategy or the Enterprise Risk Management Framework), but they will follow common themes:

- Increase profits (within Risk Appetite)
- Ensure and maintain a sustainable business
- Satisfy customer needs
- Maintain an effective and productive workforce
- Operate within legal and regulatory requirements

The question then becomes - the extent to which AI (the 'thing') will help or harm the organisation's chances of achieving its objectives. The difficult part will be identifying what the positive and negative forces are, and what their impact will be.

As with any new 'thing' (AI, cybersecurity, operational resilience or a Collateralised Debt Obligation), the danger is that the organisation delegates the management of risks and opportunities to a 'priestly class' who profess to have a greater understanding of the issues – this can be internal 1st or 2nd Line, or a third-party provider.

If an incurious Board is not prepared (due to time constraints, over-confidence or fear of looking stupid) to:

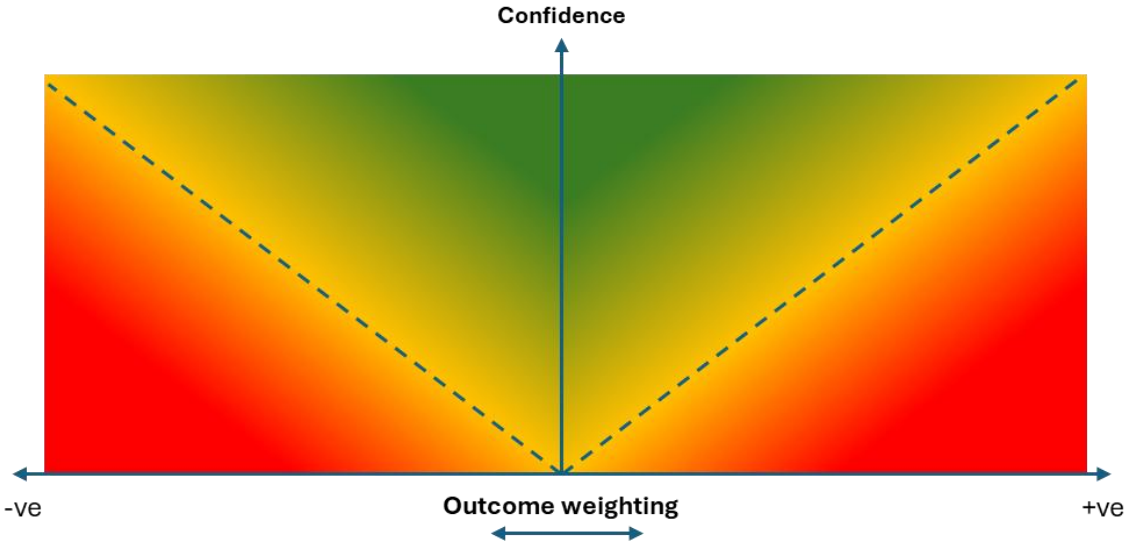
- Ask Management to "explain this to me as if I were a five-year-old"
- Get out of the Boardroom and into the business to talk to the experts on the ground
- Seek external assurance if they are still not feeling the level of confidence they require

Then the Board (and the shareholders…) should not be at all surprised if the organisation fails to manage the risks of AI, or capitalise on its opportunities.

1.1 AI uncertainty

For simplicity, this paper is framed in the traditional terminology of Enterprise Risk Management – risks, risk owners, risk management etc. However, when considering new technology, organisations will benefit from joining the FCA in taking an outcomes-focused approach, and thinking more in terms of Outcome Management (and the Outcome Managers that come with that concept). The impact of AI will not be a binary good/bad; it will be a continuum of positive and negative Outcomes, based on how the technology is used by the organisation, but importantly by the organisation's customers and competitors.

The role of the Board is to ensure that a suitable framework is in place to identify both the positive and negative forces that will influence these outcomes. And most importantly, it is for the Board to test the level of confidence Management has in their assumptions. And the Board should expect a level of certainty commensurate with the size of the impact.



Certainty (and how it can be tested) should be the key area of focus for Boards.

This paper discusses how organisations can approach AI Governance, and how they can expand and enhance the existing control framework to accommodate the new technology. And, most importantly, how they can reduce uncertainty.

Ensuring effective AI governance will require a structured approach:

- Assessing the opportunities and risks AI brings (Section 2)
- Undertaking a gap analysis to establish where improvements need to be made (Section 2.4)
- Updating and enhancing the Governance and Control Framework, to increase the likelihood of the desired outcomes (Section 3)
- Ensuring the organisation has people with the right capacity, capability and culture to make sure the objectives will be delivered (Section 3.1)

2 Assessment and gap analysis

As with the run-up to the tech bubble of the 90s, the AI sector is awash with promise. The expectations for cost-saving efficiencies and revenue-smashing solutions far outweigh their likely results. In situations like these, the Board should be the cooler heads that prevail. The Board is there to help Management differentiate between the goldrush and the shovels - to view risks through the glass half full and view opportunities through the glass half empty.

In considering the costs and benefits of AI, the Board should be guided by the 'North Star' of the organisation's business strategy and the guardrails of the Risk Appetite Statement (RAS). If either of these is out of alignment, the organisation will be making flawed decisions.

Sam Brannan was California's first millionaire. He started the 1848 gold rush by parading through San Fransico, holding a nugget of treasure and shouting "Gold! Gold! Gold! Gold from the American River!"

Prior to this, he had already cornered the market in shovels and mining equipment, which he then went on to sell to prospectors at hugely inflated prices.

For simplicity (and consistency with the traditional approach), the following sections split out Opportunities and Risk. But, as noted in the introduction, for an emerging technology like AI, it is recommended that organisations consider the continuum of positive and negative outcomes, and how they interact together.

2.1 Opportunities

AI has the potential to realise the Fourth Industrial Revolution we've been promised since the late 90s. But there will be plenty of fool's gold in this gold rush. The Board will be presented with a range of internally and externally generated propositions and use cases for AI, across a range of categories:

- Analysing customer behaviour, and making recommendations for what the organisation should do next
- Providing tailored customer offers and solutions, based on inputs and aggregated knowledge
- Answering customer queries and providing information, based on inputs and aggregated knowledge
- Identifying unusual activity and potential risks
- Analysing staff behaviour, and making recommendations for next actions
- Handling routine tasks and increasing productivity
- Cataloguing and sorting information, and enabling easier retrieval
- Improving process management and information flows
- Producing drafts of work product

When considering the benefits, the oft-quoted adage from Britain's Olympic-winning rowing team holds true "will it make the boat go faster". It is important for organisations to assess any propositions through the lens of their strategic objectives – controlling costs, increasing revenues, ensuring a sustainable business model, etc.

Whilst Boards (and senior management) should take a healthily sceptical view, they should also recognise the inevitability that AI will become a universal technology in the (not too distant)

future. Organisations that don't grasp the opportunities of AI will find that others in the marketplace will, and they will rapidly gain a competitive advantage to overtake them.

In the late 1980s High Streets were dotted with video rental shops. Consumers went to the store on a Friday night, selected the VHS cassette they wanted, and then went back to the store to return it the next day (or pay an extortionate late fee). The market was fragmented, with countless independents offering inconsistent service and little differentiation.

Then came Blockbuster - a clean, branded, standardised experience that won over consumers. Through rapid expansion and acquisition, they quickly became that most dangerous of things – a business increasing market share in a declining market.

Because Netflix arrived and changed the game. First, by leveraging the shift from bulky VHS to lightweight DVDs in the late '90s to launch a mail-order rental service. Then, in the late 2000s, Netflix capitalised on another technological advance (the rising home broadband speeds) to pioneer streaming.

Blockbuster went bankrupt, and closed its last physical store (in Bend, Oregon) in 2014. Netflix is now worth $529 billion. Blockbuster, and many other firms wiped out by technological change forgot (or never knew) the classic marketing aphorism - "People don't want quarter-inch drill bits; they want quarter-inch holes"

2.2 Risks

The outcome of technological change is difficult to predict, but organisations can mitigate the risk by taking a full 360-degree view of the threats, and by putting in place effective horizon scanning (both internal and external) to monitor it.

2.2.1 Commercial risk

For most organisations , it is likely to be Commercial Risk, rather than Operational Risk or Legal/Regulatory Risk, that will have the most significant impact. And Commercial/Business Risk is the area that most ERM Frameworks fail to scrutinise sufficiently.

2.2.1.1 Direct risk

AI will be a significant disruptor. Either through substitution (other services will supplant existing propositions), or through increased competition as existing or new entrants use AI to drive out organisations that are failing to keep up.

If organisations are not keeping up with the competition (see Red Queen Theory below) or not effectively monitoring the business landscape (competitor myopia), they place themselves at a significant disadvantage.

Kodak invented the digital camera in 1975. They chose not to exploit the technology because they feared it would harm their core business of selling analogue film.

Kodak continued to prioritise film, even while Sony and Canon were taking significant market share with their digital cameras.

Kodak filed for bankruptcy in 2012.

2.2.1.2 Indirect risk

Indirect commercial risks are difficult to monitor and quantify. But guidance from other sectors can help organisations assess the impact of AI. In particular, work done on ESG is helpful. In 2021, the EBA produced EBA/REP/2021/18 (ESG risks for credit institutions and investment organisations), which described how climate change and moves to transition away from carbon-intensive sectors could impact credit and investment risks, and how the weighting of these risks would differ across industries and geographical regions. AI changes will bring similar impacts, but they will often be harder to identify – there is no Mauritius to be washed away by rising AI uptake levels.

Equally, organisations involved in consumer lending (particularly longer-term consumer lending) should consider how AI will impact the sustainability of certain jobs, and the ability of customers holding those jobs to continue repaying their loans. For example, in the 70s, an airbrush artist was a skilled and well-rewarded profession; no prog-rock album or fantasy novel worth its name was not embellished with airbrushed imagery. The arrival of Photoshop wiped out that trade.

There is also an indirect risk from Supply Chain/Value Chain disruption, as organisations could find that their strategic partners may exit the market or change their business model.

2.2.1.3 Red Queen Theory

The Red Queen Theory is inspired by Lewis Carroll's Through the Looking-Glass, where the Red Queen tells Alice, "It takes all the running you can do to keep in the same place". It was initially applied to evolutionary biology, but it holds true for business too, and doubly so for business in the age of AI.

AI technologies are evolving rapidly. Organisations must invest in ongoing model upgrades, data governance, and ethical oversight to avoid falling behind competitors or regulatory expectations.

In areas like financial crime and cybersecurity, the Red Queen dynamic is stark: organisations must evolve their detection systems faster than criminals evolve their evasion tactics. This creates a perpetual cycle of innovation and counter-innovation.

Organisations that treat AI as a one-off transformation will be outpaced by rivals who are committed to continuous investment and improvement - they will rapidly gain competitive advantage to overtake them.

In the 19th century, Lancashire was the Silicon Valley of its day - the 'dark satanic mills' that crowded the landscape were filled with innovative technology like spinning mules and steam-driven ring frames. The competitive advantage they brought meant Lancashire manufactured and exported cotton yarn across the globe.

The mill owners faced high initial costs to purchase the technology, but the machines were very durable and easy to maintain. Given there was little global competition for their goods, the mill owners saw no incentive to further invest in upgrading their equipment.

But, while the mill owners were 'sweating their assets', the white heat of technology marched on. Platt Brothers, Howard & Bullough, Asa Lees and the host of other Northwestern inventors were building innovations like quick-change shuttles, jacquard attachments and patent shedding mechanisms to improve the speed and quality of output.

Finding no market for their products in the UK, the engineering firms set up sales offices in Bombay, Ahmedabad, Yokohama, New England and São Paulo. By 1900, Lancashire's share of global spindle-hours fell below 25 %

2.2.2 Regulatory risk

UK-based organisations should certainly focus on the developing domestic requirements. But there are also insights to be gained from frameworks being developed in other jurisdictions, particularly those more mature than the UK.

2.2.2.1 UK approach

The UK regulatory framework has been guided by the DSIT paper from 2024 (Implementing the UK's AI Regulatory Principles) and the government's 'pro-innovation' approach. DSIT laid out five principles, which provide a helpful framework for organisations to consider when developing their own approaches to AI Governance:

1) safety, security, robustness - AI systems should function in a robust, secure and safe way throughout the AI life cycle, and risks should be continually identified, addressed and managed
2) appropriate transparency and explainability - AI systems should be appropriately transparent and explainable
3) fairness - AI systems should not undermine the legal rights of individuals or organisations, discriminate unfairly against individuals or create unfair market outcomes. Actors involved in all stages of the AI life cycle should consider descriptions of fairness that are appropriate to a system's use, outcomes and the application of relevant law.
4) accountability and governance - governance measures should be put in place to ensure effective oversight of the supply and use of AI systems, with clear lines of accountability established across the AI life cycle
5) contestability and redress - where appropriate, users, impacted third parties, and actors in the AI life cycle should be able to contest an AI decision or outcome that is harmful or creates a material risk of harm

The individual regulators have laid out their approaches in response to the DSIT framework.

2.2.2.2 Other jurisdictions

In Europe, the AI Act is somewhat less 'pro-innovation' than the UK framework. It is also broad-reaching and will apply to public and private actors inside and outside the EU, if a qualifying AI system is placed on the EU market or its use affects people located in the EU.

Even if not directly applicable, the Act has a number of useful elements which UK organisations can use as guidance when developing their own AI Governance frameworks. In particular, the risk taxonomy, which categorises AI systems into four risk levels: Unacceptable - AI systems considered a clear threat to the safety, livelihoods and rights of people. High risk - AI use cases that can pose serious risks to health, safety or fundamental rights. Limited risk – systems that require specific transparency and disclosure obligations. Minimal risk – low impact systems like spam filters and basic recommendation engines.

2.2.3 Operational risk

The inclusion of AI should not alter the Operational Risk landscape, but certain areas within it will be particularly impacted.

2.2.4 Model risk

Organisations outside of financial services may not be familiar with the PRA's guidance for Model Risk Management (SS1/23). However, it provides a useful framework when considering how to put effective controls in place. AI is the classic 'rubbish in, rubbish out' technology, so inconsistencies, errors and biases in the model are key risks. Organisations should therefore consider the key principles of:

- Model identification and model risk classification - Organisations should have an established definition of a model that sets the scope for MRM, a model inventory and a risk-based tiering approach to categorise models to help identify and manage model risk.
- Governance - Organisations should have strong governance oversight with a board that promotes an MRM culture from the top through setting a clear model risk appetite.
- Model development, implementation and use - Organisations should have a robust model development process with standards for model design and implementation, model selection, and model performance measurement.
- Independent model validation - Organisations should have a validation process that provides ongoing, independent, and effective challenge to model development and use.
- Model risk mitigants - Organisations should have established policies and procedures for the use of model risk mitigants when models are under-performing, and should have procedures for the independent review of post-model adjustments.

2.2.5 Data security and verification

The use of AI, particularly from third-party vendors, opens organisations to increased avenues for data theft and leakage. AI also offers cyber criminals an increased range of tools and techniques to attack systems, particularly using social engineering.

AI content generation offers a host of opportunities for bad actors to create false documentation – to forge identities and/or to provide incorrect information (payslips, etc). Voice cloning and deepfake videos offer further tools for criminals to impersonate customers or staff.

Staff may also unwittingly leak the organisation's intellectual property if they upload documents to AI systems to generate revisions or expanded content.

2.2.6 Third-party risk and operational resilience

Relying on third-party AI platforms or data providers can obscure accountability. A vendor outage or unidentified model flaw can cascade into operational disruptions. Organisations should ensure their due diligence and oversight frameworks are expanded to cover AI requirements. Also, whilst organisations may gain cost savings by increasing the use of

chatbots and other AI solutions, this brings with it an expectation that the organisation will be assessing and controlling the Operational Resilience risks.

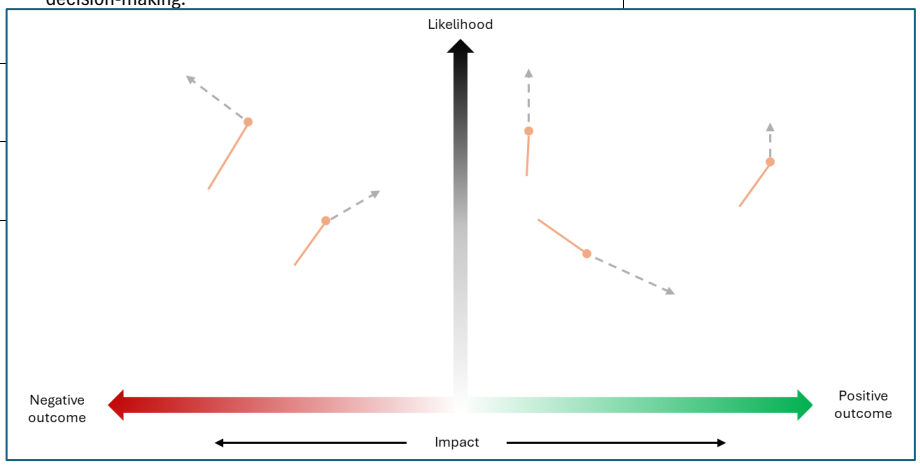
2.2.7 People risk

Organisations should be aware of the People Risks related to downsizing and reorganisation. Reducing headcount and replacing staff with chatbots may bring revenue savings, but it can come at a significant cost, including staff demotivation, poor industrial relations, and the unplanned exit of key personnel. Organisations should be very certain as to how they are going to approach any reorganisation, and how they are going to mitigate the impact on the culture and productivity of the remaining employees.

2.2.8 Gap analysis

Organisations should make an honest assessment of the risks and opportunities AI brings, and the impacts on the current assumptions for strategic plans and Risk Appetite. Organisations can support the process using a range of management tools.

Strategy	• Map how AI-driven products or services align with long-term goals. • Identify strategic gaps: Are we underinvesting in AI R&D?
Structure	• Assess if your organizational design supports data science hubs, cross-functional AI squads, or centralized AI governance. • Highlight bottlenecks between business units and AI teams.
Systems	• Review core processes—credit scoring, fraud detection, customer onboarding—to embed AI models and automation. • Check for missing controls: Are model-validation workflows integrated?
Shared Values	• Gauge cultural readiness for AI experimentation and data-driven decision-making.
Staff	
Skills	



Whilst assumptions should be evidenced and stress tested; it is important not to let the perfect be the enemy of good. This will be an iterative and ongoing process. The aim of the initial review is to identify significant gaps, and use this information to inform the development of the Governance and Control Framework.

3 Governance and control framework

Within a relatively short space of time, AI will be as pervasive as the Web in the day-to-day operations of organisations. Therefore, it is unlikely that organisations will be able to effectively oversee the risks and opportunities associated with this emerging technology with a dedicated AI Committee, a separate pillar of AI risk within the Risk Appetite Statement, or specific sections on AI in Board packs.

> The current Governance Framework for AI in most organisations - Half the team are playing with matches, the other half are rubbing two sticks together hoping for a spark (we don't know which is which)

A 'technology agnostic' approach is likely to be the most productive model for most organisations, with AI being a 'lens' that is applied to all areas of the Governance and Control Framework.

The Board should be seeking to use the Governance and Control Framework to reduce uncertainty. With the Board challenging Management on how their approach, controls and strategic plans are being helped or hindered by the developments in AI. Therefore, Boards should expect that in the MI and information packs they receive, management is able to justify the confidence that:

- The opportunities of AI can be successfully exploited
- The risks of AI can be effectively controlled

It is illustrative that Kodak would have been a very different company in the 80s and 90s if the Board had, had greater certainty on how emerging technology would impact their competitive environment.

3.1 The 3Cs – Capacity, Capability and Culture

As with any endeavour the organisation undertakes, the People element will be the key determinant of the success or failure of the Governance and Control Framework. Unfortunately, this is the area that is often overlooked - a organisation can have as many policies, procedures, and Board packs as it likes, but if staff don't have the will or skill to follow them, or act in accordance with their direction, any impression of control is entirely illusory.

For a complex, fast-moving, and highly significant issue like AI, People Risk will be key. And it will be essential to consider the 3Cs at all levels of the organisation. And in particular, the capacity, capability and culture of the Board:

- Capacity - At the Board (and in sub-committees), the Chair/CoSec should ensure there is sufficient time to consider and debate AI issues. This should include sufficient time for explanations and knowledge building.
- Capability - The inclusion of dedicated AI experience (e.g. an AI-literate NED) can be useful, but this should not remove the need for a baseline level of understanding across the Board/Exco.
- Culture - Given the significant information asymmetries, the Board should promote an open culture, and expect a Duty of Candour from managements (and management should expect the same from their subordinates)

3.2 Training and competency

Based on the 3Cs approach, training and competency should be a priority. It is essential that staff at all levels have a thorough understanding of the risks and opportunities of AI and how it aligns with the strategic objectives and risk appetite of the organisation. Staff should have the capability to:

- Use AI appropriately – understanding 'why' the rules and guidelines are in place, and how they align with the organisation's culture. There should be a key focus on bias mitigation techniques, explainability and transparency requirements, and incident escalation for issues and anomalies
- Use AI effectively – staff should be clear on how to identify opportunities and develop solutions to utilise AI in the business.

Training should be tailored to the business requirements and the needs of the groups within the organisation. Whilst periodic 'sheep dip' online training has its place, some teams will require a more tailored approach. As with all T&C frameworks, organisations will be expected to demonstrate that staff have the required level of competency to discharge their responsibilities:

- Board and senior management – high level knowledge and understanding of macro risks and opportunities
- Control Functions (2nd Line, 3rd Line, Legal and HR) – details on legal and regulatory requirements
- 1st Line – deep insights into the technical requirements and controls
- Whole organisation – safe and effective use of AI

3.3 Risk management

For organisations taking a 'technology agnostic' view, monitoring the risks of AI should be easily accommodated within the current Enterprise Risk Management Framework.

RCSA/Risk Register

When facilitating the RCSA process, 2nd Line Risk should be mindful of the potential information asymmetry between themselves and the risk owners (as with other technical areas like cybersecurity). 2nd Line should ensure there is sufficient time allocated to fully understand the risks and controls – with particular emphasis on preventative and detective controls.

Organisations should also be careful about the allocation of risk ownership. It is likely that the IT function will own some of the risks, but not all. Risk ownership should be with the Operational department that is using the AI tool. It is their responsibility to ensure it is being controlled effectively.

Risk Appetite Statement (RAS)

As with all areas of the organisation, the RAS should have an AI lens applied. Organisations could have a separate section of the RAS focused on AI, but this is less likely to be effective. In reality, AI will have a positive or negative effect across the existing elements within the RAS.

When applying the AI lens, organisations should seek to reduce uncertainty – with an emphasis on explainability, transparency and auditability. Situations where it is unclear how an outcome has been achieved should be outside of tolerance.

Risk myopia

As noted in Section 2.2.1, the key AI risk organisations are likely to face is commercial – external threats (new market entrants, increased competition) and internal failings (ineffective adoption of AI, increased costs). Organisations should consider how they are identifying, quantifying and controlling these risks within the ERMF. And, more importantly, how they are being reported to the Board.

The word luddite is often used to describe someone who is a technophobe. The term originates from Ned Ludd a (possibly mythical) textile worker who destroyed two cotton spinning machines in 1779.

His actions became a rallying cry for spinners and weavers – skilled artisans who feared that their jobs would be lost to machines. By 1811, factories across Yorkshire and Lancashire were suffering nighttime raids, with equipment smashed, and the calling card of Captain Ludd left in the wreckage.

3.4 The 3 Lines of Defence

There is no reason why the organisation's response to AI should change the structure of the 3LoD, or the risk-ownership/risk-monitoring split. However, it will be important to ensure there is clarity around who is doing what in relation to AI:

- Risk ownership – if organisations have not done so already, this may be an opportunity to restate and reinforce the risk ownership in the 1st Line.
- RACI – as with risk ownership, organisations should take the opportunity to refresh and clarify who is: Responsible and Accountable, and who will be Consulted and Informed. If necessary, there may also be changes required to SoRs, Role profiles and Job Descriptions.

3.5 Policies and procedures

There is always caution in developing a standalone policy to address an emerging issue – there are many Conduct Risk Policies and Consumer Duty Policies languishing in drawers. However, given the broad impact of AI and the nuanced requirements, it does make sense for organisations to document their high-level principles and then use these to inform operational policies and procedures.

This AI Policy should include the guardrails from the RAS and the cultural framework that reinforces the principles. Given the clear commercial risks organisations will face, it will be beneficial to document the expectations for AI use, as well as the prohibitions. Without this overarching framework, organisations may find there's a risk of divergence between business functions.

Other policies and procedures within the organisation can be amended, as needed, to include necessary guidance for staff in relation to AI use.

3.6 Horizon scanning

AI is a fast-changing environment, both in terms of technology and regulation. The near-term horizon is very crowded. Boards don't need a data dump, they need qualitative assessment and analysis – what are the most important issues, what do we need to do about them?

Organisations should scan the external horizon, but also the internal horizon too. As developments within the organisation will be equally fast-moving, the costs and benefits of projects and initiatives will be changing, and operational risks will be increasing or decreasing.

Organisations should be confident that they have visibility of:

- What are the priority regulatory changes we're going to see in the next 12-24 months?
- What competitor activity should we be concerned with?
- What competitive advantage can we gain?
- Which areas of the organisation are heading outside of risk appetite
- Which projects and strategic objectives are being put at risk
- What additional resourcing is required

3.7 Monitoring and assurance

2nd and 3rd Line will need to ensure they have the capacity and capability to undertake audit and monitoring on the organisation's AI systems. The Audit Universe, risk landscape and control repositories will need to be updated to encompass any new AI systems and approaches the organisation introduces.

2nd and 3rd Line should consider how they are going to assess AI and what standards they are going to use (IEEE P7000, ISO 42001 etc). This should be reflected in the Audit Plan and the Compliance Monitoring Plan. The approach should be assessed and agreed at RiskCo/AuditCo and approved by the Board.

1st Line should also be adapting their Quality Assurance framework to accommodate any additional elements required.

4 Next steps

AI will not be the rising tide that lifts all boats. Organisations can either ride the wave or be swept away by the current. It is the classic VUCA environment – in some sectors, the impact will hit like a tsunami, in others it will be a slow, inexorable erosion (of market share).

In situations like these, the Board has a clear role in delivering the appropriate push and pull momentum - to add value, whilst protecting sustainable growth:

- Pushing management to utilise AI to deliver benefits
- Pulling management back into alignment with the strategy and risk appetite

Written by Frank Brown

Director – GRR Consulting Ltd