Deloitte.

AI and risk management
Innovating with confidence

CENTRE *for*
REGULATORY
STRATEGY
EMEA

Contents

1. Executive summary

Artificial Intelligence (AI) is not a new concept, but it is only in recent years that financial services (FS) firms have started to learn about and understand its full potential.

AI can drive operational and cost efficiencies, as well as strategic business transformation programmes, including better and more tailored customer engagement. However, limited availability of the right quality and quantity of data, insufficient understanding of AI inherent risks, a firm's culture, and regulation can all act as real, and in some cases, perceived barriers to widespread adoption of AI in FS firms.

EU and international regulators have also taken an active interest in AI, and while they recognise the benefits that AI can bring to financial markets, consumers, and their own work, they are also increasingly mindful of the potential risks and unintended consequences that the use of AI by regulated firms may have.

This is particularly relevant given that, over recent years, the FS sector has been hit by a significant number of financial and other penalties in relation to the mistreatment of customers and market misconduct. The resulting focus on the fair treatment of customers and market integrity, together with the relatively untried and untested nature of AI in a regulatory context, have meant that FS firms have been understandably cautious about the adoption of AI solutions.

Effective risk management, far from being an inhibitor of innovation, is in fact pivotal to a firm's successful adoption of AI.

To overcome these obstacles, reap the full benefits of AI adoption, and avoid problems further down the road, it is imperative for boards and senior management to develop a meaningful understanding of the technology, including its existing and potential uses within their organisations, and take a firm grip on the implications of AI from a risk perspective. In this context, effective risk management, far from being an inhibitor of innovation, is in fact pivotal to a firm's successful adoption of AI.

"The biggest challenge for firms is less about dealing with completely new types of risk and more about existing risks either being harder to identify in an effective and timely manner, or manifesting themselves in unfamiliar ways".

In relation to the latter point, which is the focus of this paper, we believe the biggest challenge for firms is less about dealing with completely new types of risks and more about existing risks either being harder to identify in an effective and timely manner, or manifesting themselves in unfamiliar ways. In this paper, we discuss how firms should review and adapt their existing Risk Management Framework (RMF) to reflect some of the important differences which need to be taken into account when deploying complex AI use cases.

For example, the ability of AI to learn continuously from new data, and to make decisions which are driven by complex statistical methods, rather than clear and predefined rules, can make it challenging for firms to understand the decision drivers which underpin the final output. In many ways, this is not dissimilar from the challenges that organisations face in managing human resources. However, evolving AI solutions can make auditability and traceability challenging, and the speed at which they evolve can result in errors manifesting on a large scale, in a very short space of time.

Firms will need to review and update their risk practices to manage risks through the various stages in the RMF lifecycle (identify-assess-control-monitor). The continuously evolving nature of AI solutions will require some of these activities to happen at shorter and more frequent intervals. Existing risk appetite statements will also need to be reviewed, and a number of new components, such as a fairness policy for example, may need to be developed to inform the various phases of the RMF.

Evolving AI solutions can make auditability and traceability challenging, and the speed at which they evolve can result in errors manifesting on a large scale, in a very short space of time.

In this paper, we use a simple conceptual RMF to draw out some of the challenges that AI introduces and bring the framework to life by working through a practical example of how a firm may manage model risk within an AI solution for policy pricing in the insurance sector. We also discuss how regulators are responding to AI solutions, and highlight their key areas of focus and expectations. We conclude by reflecting on the challenges and options available to regulators to regulate AI.

This paper is intended to be a starting point for understanding the implications of AI for existing risk management practices, as well as the broader regulatory context. By highlighting these areas for consideration, we hope to empower firms to provide more effective challenge and oversight in the development of an AI strategy more generally, and in the development of an AI RMF more specifically.

2. A brief overview of AI

The concept of AI dates back to the 1950s, when researchers first started contemplating the possibility of using machines to simulate human intelligence. However, AI only really "took off" in the late 2000s, when several enabling factors reached tipping point: the unprecedented availability of affordable computer power; the rise in the volume and variety of data, and the speed of access to it; and the emergence of new and advanced algorithms[1] able to analyse data in a more "intelligent" way.[2]

There is no single definition of AI, but broadly speaking, AI is the theory and development of computer systems able to perform tasks that normally require human intelligence.[3] Examples of such tasks include visual perception, speech recognition, and decision-making and learning under uncertainty.

The lack of a consensus on a definition may be explained, at least in part, by the fact that AI is not a technology per se, but rather a collection of techniques that mimic human behaviour. Some of the key techniques, which are currently relevant for FS and to which we refer in this paper, are:

Machine Learning - the ability to improve computer systems' performance by exposure to data without the need to follow explicitly programmed instructions. At its core, machine learning is the process of automatically discovering patterns in data and using them to make predictions.

Deep Learning - deep learning algorithms are a class of machine learning algorithms that are becoming more and more popular because of their effectiveness in tasks related to speech and computer vision. They are complex techniques where it is hard to decipher exactly how each input drives model outcomes, often resulting in them being characterised as "black boxes".

Speech recognition and Natural Language Processing (NLP) - the ability to understand and generate human speech the way humans do by, for instance, extracting meaning from text or even generating text that is readable, stylistically natural, and grammatically correct.

Visual recognition - the ability to identify objects, scenes, and activities in images. Computer vision technology uses sequences of imaging-processing operations and techniques to decompose the task of analysing images into manageable pieces.

Improving the customer experience in financial services

Behaviour and Emotion Analytics Tool (BEAT) is Deloitte's voice analytics platform that uses deep learning and various machine learning algorithms to monitor and analyse voice interactions. It has three core functions:

1. Monitoring customers' voice interactions
2. Identifying high-risk interactions through NLP
3. Mapping interactions to potential negative outcomes (such as a complaint or conduct issue) and providing details as to why they arise

BEAT analyses both the words that were spoken by the customer and the tone with which they were spoken, and utilises machine learning technology to constantly develop and enhance the algorithms that analyse the interactions – the greater the volume the more accurate the risk assessment will be. BEAT gives firms a significant uplift in accuracy rates against traditional solutions.

BEAT has been developed to analyse over 30 different languages and 30 different behavioural indicators. It can be tailored to meet specific risk requirements and user needs.

1 A process or set of rules to be followed by a computer in calculations or other problem-solving operations.
2 Demystifying artificial intelligence - https://www2.deloitte.com/insights/us/en/focus/cognitive-technologies/what-is-cognitive-technology.html
3 https://en.oxforddictionaries.com/definition/artificial_intelligence

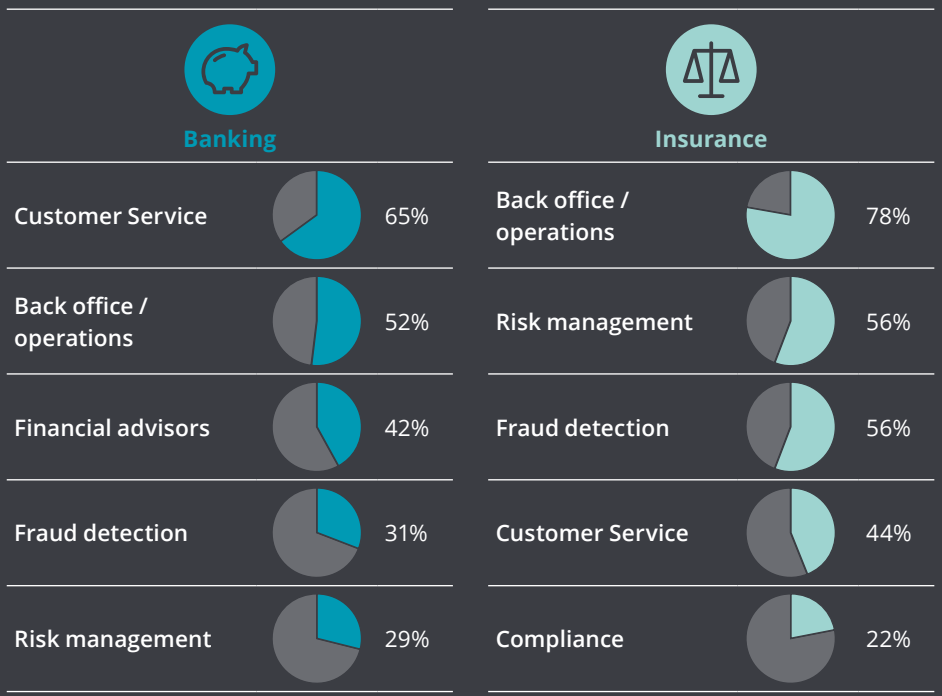
3. Challenges to widespread AI adoption in Financial Services

Since the financial crisis of 2008, FS firms have been striving to drive cost efficiencies and maintain competitiveness to offset margin pressures. To achieve this, one area they have looked to is technology, with greater use of AI being explored in the last few years. However there has not been a one size fits all approach to AI adoption, and there are a number of reasons why this is the case.

Differing views on where AI should be applied

A recent Deloitte survey[4] of more than 3,000 C-Suite executives, conducted in partnership with the European Financial Management Association (EFMA), shows that the activities and functions in which firms believe AI could have the biggest impact for their organisation vary considerably by sector (Figure 1).

Figure 1: "On which part of the value chain do you see the Artificial Intelligence use case you have developed having the greatest impact?"



The survey also concluded that, overall, the adoption of AI in FS is still in its infancy. Of the firms surveyed, 40% were still learning how AI could be deployed in their organisations, and 11% had not started any activities. Only 32% were actively developing AI solutions.

4 AI and you | Perceptions of Artificial Intelligence from the EMEA financial services industry - https://www2.deloitte.com/nl/nl/pages/financial-services/articles/ai-and-you.html

Data availability and quality

One of the key differences between AI and other traditional technological solutions is that, while the latter are limited to tasks that can be performed by following clear rules defined in advance, AI applications are able to analyse data to identify patterns and make decisions based on them. Additionally, AI applications are programmed to *learn* from the data they are supplied with, either as a one-off at the time of their design, or on a continuous basis, to refine the way decisions are made over time.

This means that the quality of any decision made by an AI solution has a significant dependence on the quality and quantity of the data used. An absence of large sets of high quality data is, in general, one of the major obstacles to the application of AI solutions. For many FS firms this is exacerbated by the prevalence of legacy systems and organisational silos which can prevent the seamless flow of data and/or affect its quality.

Transparency, accountability and compliance

Some AI solutions have multiple hidden layers of decision making which influence the final outcome. In the case of complex AI applications, such as those using deep learning, it can be challenging for FS firms to maintain, and evidence, the necessary level of understanding and control over AI-based decisions, including their appropriateness, fairness, and alignment with the organisation's values and risk appetite.

AI solutions which learn and evolve over time, and contain many hidden decision processing layers, can make auditability and traceability challenging.

This is not dissimilar from the challenges that organisations face in managing human resources. However, AI solutions which learn and evolve over time, and contain many hidden decision processing layers, can make auditability and traceability challenging. In addition, the speed at which AI solutions learn and evolve can result in errors manifesting on a large scale, very quickly.

The opacity of some AI solutions also poses practical challenges in relation to certain regulations, such as the new General Data Protection Regulation (GDPR) in the EU, which in certain circumstances will require firms to be able to explain to customers how their personal data is used, and give them a meaningful explanation of the assumptions and drivers behind a fully automated decision with a significant impact on the customer.

Understanding AI and its implications

AI is a complex and fast-evolving field that, in the eyes of the non-expert, could justifiably be seen as difficult to control. In addition, the use of AI can heighten existing enterprise risks, change the way they manifest themselves, or even introduce new risks to the organisation.

FS is a highly regulated industry, comprising a wide and complex variety of business lines and products, and firms must always apply an adequate level of prudence in conducting their business. The history of regulatory penalties for non-compliance or misconduct experienced by the FS industry, however, introduces an additional level of conservatism in the adoption of relatively unknown technologies in regulated activities, which may act as a barrier to innovation.

An excess of caution may derive from relatively low levels of familiarity with, or understanding of, the technology and its inherent risks. Key stakeholders, such as risk, compliance, heads of business, board members and executives may rightly be hesitant to approve and be held accountable for the use of AI for regulated activities in their organisation, unless they feel they have a meaningful understanding of the technology. Such understanding would need to extend beyond the risks that the technology introduces, to how these can be minimised, managed, and monitored.[5]

Addressing the individual and collective AI understanding of key stakeholders is a challenge for firms. Using real, applicable use cases, and understanding the related customer journey could help stakeholders familiarise themselves with the potential benefits of AI but also with what can go wrong, and how, practically, risks can be mitigated or managed.

The human impact

For organisations adopting AI, especially on a large scale, it will be essential to understand fully the impact that such transformation will have on their culture and talent strategy, and to put in place the necessary measures to address any adverse effects.

In all likelihood, firms will need additional skilled technical resources to help design, test and manage AI applications. The current scarcity of such talent, as well as the perception that FS firms struggle with innovation, may make this challenging. Recruitment practices and channels will need to be updated, and career paths and retention/integration/succession strategies for technical staff developed.

Some of the implications for existing staff may be even more profound. Developments in AI are expected to reduce aggregate demand for labour input into tasks that can be automated by means of pattern recognition.[5] Significant changes to employment practices, such as reduced staffing needs, or re-assignment of existing staff to different activities (with the associated re-training considerations), are likely to affect staff morale and, if not addressed promptly, may lead to an increase in unwanted staff turnover.

Excessive loss of personnel may jeopardise firms' ability to retain the necessary expertise and enough skilled staff able to perform processes manually if AI applications fail or must be retired at short notice. It may also have implications for the development of firms' next leadership generation.

5 https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

4. Embedding AI in your Risk Management Framework

The adoption of AI, and innovation in general, require firms to go through a learning journey. Such a journey, however, is not about avoiding all AI-related risks, but about developing processes and tools to give businesses the confidence that such risks can be effectively identified and managed within the limits set by the firm's risk culture and appetite. Therefore, and perhaps despite common misconceptions, effective risk management plays a pivotal role in firms' ability to innovate.

Nature of risks inherent to AI applications

We believe that the challenge in governing AI is less about dealing with completely new types of risk and more about existing risks either being harder to identify in an effective and timely manner, given the complexity and speed of AI solutions, or manifesting themselves in unfamiliar ways. As such, firms do not require completely new processes for dealing with AI, but they will need to enhance existing ones to take into account AI and fill the necessary gaps. The likely impact on the level of resources required, as well as on roles and responsibilities, will also need to be addressed.

Firms do not require completely new processes for dealing with AI, but they will need to enhance existing ones to take into account AI and fill the necessary gaps.

The Deloitte AI Risk Management Framework provides a mechanism for identifying and managing AI-related risks and controls. In the table presented on the next page, and the following sections, we set out some of the key considerations from the overall population of over 60 AI risks covered in the framework. These considerations are expressed in general terms – in reality the level of risk and the necessary controls will vary, in some cases significantly, from use case to use case, and organisation to organisation.

A scientific mind-set

Adopting and advancing AI require an organisation and the people who work in it to embrace a more scientific mind-set. This means being comfortable with a trial and error journey to the final product, accepting risks and tests that fail; and continuously testing the feasibility of the product by introducing external shocks or data and observing outcomes. Essentially, it means creating a "sandbox" (a controlled, isolated environment representative of the business environment) across the organisation. This mental shift is not just solely for Heads of business or functions, but is relevant to all areas of the organisation, including the board and other functions such as risk and compliance, HR and IT.

It is particularly important to involve all three lines of defence (business lines, risk/compliance and internal audit). As the guardians of compliance and controls oversight, full participation in the sandbox would allow them to understand some of the critical technical aspects, and help shape, from the start, the appropriate AI governance and risk management policies.

Enterprise risk category	Example sub categories	Examples of key considerations for AI solutions relative to other technologies
Model	Algorithm risk – bias	• Dependence on a continuously evolving dataset that drives AI decisions makes it harder to identify inherent bias in the model. • Inherent bias in input data may result in inefficient and/or unfair outcomes. • Lack of consideration, by data scientists, of bias as an issue makes it more likely that bias risk will not be addressed adequately from the start.
	Algorithm risk – inaccuracy	• Incorrect type of algorithm(s) applied to a problem, poor data quality or suboptimal choice of algorithm parameters.
	Algorithm risk – feedback	• Increased risk of inappropriate feedback going undetected (in those AI solutions allowing for continuous feedback and learning) may compromise the solution's ability to produce accurate results.
	Algorithm risk – misuse	• Increased probability that business users may lack adequate understanding of complex AI model limitations and incorrectly interpret AI outputs leading to poor outcomes.
Technology	Information and cyber security	• Increased dependency on open source components (software packages, programming language, API, etc.) which are no longer supported or updated or freely available by the creator may introduce additional security vulnerabilities. • Complex algorithms make it harder to understand how the solution reached a decision and therefore may be subject to malicious manipulation, both by humans or other machines.
	Change management	• Difficulty in identifying the impact of changes to upstream systems that feed the AI solutions. This may result in unforeseen consequences for how AI interacts with its environment.
	IT Operations	• Significant dependence, in some instances, of AI applications on big data increases the risk posed by existing legacy IT infrastructure, as the latter may not be compatible with AI (e.g. legacy systems unable to process big data).
Regulatory & Compliance	Data protection	• Increased risk of breaches in relation to data protection legislation (e.g. GDPR), including data subject rights around automated decision making, due to the continuously evolving and opaque nature of some AI solutions.
	Regulatory compliance	• Difficulty for management to understand, and justify to regulators, how decisions are made in complex AI applications, such as those employing neural networks, which consist of a number of hidden decision-making layers.
Conduct	Culture	• Cultural challenge for large scale AI adoption due to actual or perceived regulatory and ethical concerns. • Negative impact of fear of change or concerns about the changing profile of jobs within the organisation.
	Product innovation	• Risk of products being developed which do not meet customer needs (i.e. use of AI for the sake of using AI), and related risk of widespread mis-selling.
People	Roles and responsibilities	• Increased risk that roles, responsibilities and accountabilities may not be clearly defined across the AI lifecycle. Lack of continuous engagement, and oversight from stakeholders (compliance, business, IT, coders, etc.) may increase the risk of things going wrong.
	Recruitment and skills	• Increased risk of lack of/insufficient in-house skills to understand, use and appropriately supervise AI solutions being adopted. • New risks arising from the lack of cultural integration of AI-savvy resources within the organisation. • Over-reliance on a small number of resources with AI knowledge and expertise.
Market		• Over-reliance in the market on a relatively small number of large third-party AI vendors increases concentration risk and may have network effects in the event that one of these entities becomes insolvent or suffers a significant operational loss. • Increased systemic risk resulting from herding behaviour (i.e. organisations acting identically to other market participants), if algorithms are overly sensitive to certain variables (e.g. stock market prices).
Supplier		• Use of "black box" algorithms may result in a lack of clarity around allocation of liability between vendors, operators and users of AI in the event of damages. • Increased risk of AI third-party providers' failure, especially in the case of new and smaller players, which may not have sufficient governance structures and internal controls in place.

Risk appetite

A firm's risk appetite is the amount of risk that the organisation is prepared to accept, at any point in time, in the pursuit of its objectives. In order to establish effective risk management processes and controls, any AI adoption strategy needs to align with the overall risk appetite from the start.

Equally, a firm's risk appetite may need to be revisited to incorporate AI-specific considerations. Although the overall risk appetite for the firm is unlikely to change as a result of the introduction of AI solutions, the relative balance of its components may, and the tools and measures to manage them definitely will.

AI solutions can inherently increase or decrease certain types of risks (e.g. model risk), and change both the current and future risk profile of an organisation. This means that risk appetite needs to be reconsidered at the level of each risk type. This includes reviewing not only target risk levels, but also the policies and management information that support effective management and monitoring of that risk.

For firms to assess the impact of AI use cases on their risk appetite, they should first develop a set of clear and consistent assessment criteria to apply to all such cases – for example "Is this AI solution external facing?" is one of the questions that would help determine what type of conduct risk implications an AI use case may have. Developing a set of standard questions may help firms understand which risk areas require more or less focus, both at the level of individual AI use case and in aggregate.

Risk Management Framework lifecycle

Details and language will vary from firm to firm, but conceptually a RMF lifecycle comprises four key stages:

1. **Identify**
   Understanding the risk universe by identifying which risks could have a material adverse impact on the organisation's business strategy or operations. This stage also involves monitoring the internal/external operating and regulatory environments to identify changes to the inherent risk landscape and ensure the framework remains fit for purpose.

2. **Assess**
   Defining and embedding a risk assessment process to assess the level of risk exposure.

3. **Control**
   Embedding a control framework to mitigate inherent risks to a residual level that is in line with risk appetite.

4. **Monitor and report**
   Designing a methodology for assessing the effectiveness of the control environment, including relevant metrics for measuring effectiveness, tolerance thresholds, and controls testing.

   Reporting the status of the residual risk profile, the control environment and remediation programmes to the relevant governance fora.

In the following sections we draw out some of the key AI considerations for each stage of the RMF and bring these to life by working through a practical example of how a firm may manage model risk within an AI solution for policy pricing in the insurance sector.

1. Identify

The complex nature and relative immaturity of AI in FS mean that the ways some risks manifest themselves and their magnitude may evolve over time, in some cases very rapidly. This could have important ramifications for firms, both from a conduct and a stability perspective (e.g. widespread mis-selling).

Firms will therefore need to perform periodic reassessments to determine whether the risk profile of an AI use case has changed since its introduction, as the model has learned and evolved.

Likewise, an AI solution developed as a proof of concept or for internal use only will require a reassessment if its use is expanded – e.g. if a firm plans to extend the use an AI solution initially developed for the sole purpose of providing internal advice to providing advice to external customers, the resulting risks arising throughout the new customer journey will need to be understood.

The complex nature and relative maturity of AI in FS mean that the ways some risks manifest themselves and their magnitude may evolve over time, in some cases very rapidly.

It is worth noting that the "definition" of AI will evolve too, and so will the risks – e.g. risks associated with mobile phones have changed drastically over time, as the "definition" and functionality of the mobile phone have expanded.

Firms need to determine how AI risk considerations can be integrated into their existing RMF, and to what extent it needs to change. Such considerations include regulatory and ethical implications such as algorithmic bias and the ability of AI models to make inferences from data sets without establishing a causal link – we illustrate this in the example.

However, in general, for complex and evolving AI use cases, firms will need to review their governance and methodology for identifying risks, and adopt a more comprehensive and continuous approach to risk identification. The identification of AI risks should include risks associated with the adoption of specific AI use cases (e.g. a risk profiling application) as well as risks that are introduced across the organisation through the adoption of AI more generally (e.g. impact on employee relations and corporate culture).

For identifying risks arising from AI solutions it is equally important to think about the broader organisational impact and what it means for the human capital of an organisation in the short and longer term.

1. Identify – Example

- As noted above, one of the key risks arising from risk profiling AI models is *algorithmic bias* and the ability of AI models to make inferences from data sets without establishing a causal link.
- An AI property insurance pricing model, for example, may use a variety of unstructured data for assessing a property. Such data may capture one-off local events (such as a carnival or a demonstration) into the risk profile for the area. This poses a number of risks – the primary risk is the lack of certainty around the features, i.e. decision drivers, which the algorithm will use for pricing. The secondary risk is that any one-off event happening in the location may be priced as a permanent risk for that location.
- In addition, the same data may be used in different AI models, at a future date, and inadvertently inform the risk profile of others of individuals. For example, a different AI model may use the same assessment data to risk profile an individual motor or holiday insurance policy for one of the participants or bystanders in the aforementioned event, by tagging their photos and triangulating them with their social media presence, without their consent.
- The risks generated are diverse in this example, ranging from data protection, customer consent and mispricing, not to mention the ethical considerations. While bias, model, reputational, and regulatory risks are not new enterprise risks, in AI use cases they can manifest themselves in new and unfamiliar ways, making them more difficult to identify.

2. Assess

A risk assessment process should be designed, and agreed by the firm's management, before the development of each AI use case. The process should give careful consideration to the key factors that may make a specific use case more or less risky (e.g. regulatory, customer, financial or reputational implications). For example, the level of inherent risk in, and therefore scrutiny to be applied to, an AI solution providing customers with financial advice will be different to that for a solution providing IT troubleshooting support to internal staff.

Existing risk appetite and assessment frameworks may not be sufficiently comprehensive to cover some of the qualitative considerations that AI solutions pose – for example to assess bias in AI models, firms will first need to define concepts such as "fairness" and how this should be measured. In this context therefore, firms' values, such as fairness, play a fundamental role in assessing the nature of certain risks, especially from a conduct and reputational perspective.

In addition, since AI models can evolve over time, firms may find that the original definitions and assessment metrics may not adequately address the model's decision drivers. Therefore, the assessment process will need to be more frequent and dynamic and reviewed both "bottom up" (for each individual use case) but also "top down" (overall risk appetite).

It will also require increased engagement, and sign off from a wider set of stakeholders including AI subject matter experts, risk and control functions such as technology risk and regulatory compliance, as well as representatives from the business.

Finally, AI use cases typically use an agile development approach, while many technology risk management frameworks are designed to cater to traditional waterfall models. As such, processes, policies and governance designed to assess a traditional technology development framework will need to change, and become more dynamic. In practice this may mean that, for high-risk use cases at least, risk functions may need to become much more involved on a day-to-day basis throughout the development stages. This is likely to put pressure on existing resources.

2. Assess – Example

- Following through the insurance pricing example, where the AI solution uses a number of data sources, both structured and unstructured, to provide risk weights for pricing, it is important to assess whether the results are in line with the results produced by non-AI pricing models which use static and identifiable decision drivers, and to understand the rationale for any variance. For example, for a commercial property pricing portfolio, a non-AI model may consider only the physical features of the property and its immediate surroundings, whereas an AI model may include a much larger set of drivers.

- Equally, where the pricing is done in a modular fashion by a number of discrete AI solutions, with the outcome of one AI solution feeding into another, the results from each module should be assessed by a wide set of stakeholders, to review and challenge the validity of inferred decision drivers, especially if they do not have a causal link to the risk weights which inform the pricing.

- Assessment should include both the technical parameters of the model (such as bias and classification error) but also the business (e.g. the number of policies being written by customer segment) and operational parameters (e.g. the speed of policy written from initiation to issue).

3. Control

Control and testing processes will also need to be more dynamic. In practice, this is likely to require regular and frequent testing and monitoring of AI solutions, far beyond the development stage of the solution and training on the initial data set.

This may increase the amount of testing required relative to traditional technology solutions. A risk-based approach should be used to determine the appropriate level of control for each use case, proportionate to, and consistent with, the organisation's risk assessment framework.

In addition, because of the widespread impact the adoption of AI will have on an organisation, relevant controls are likely to span multiple areas (e.g. HR, technology, operations etc.). This further emphasises the need for a wide range of stakeholders to be engaged throughout the risk management lifecycle.

Business continuity plans may need to be redefined to provide contingency for organisations to roll back to current processes in the event of system unavailability or in response to control failures in the use of AI (e.g. the deployment of a "kill switch"). The algorithm should be stress tested regularly to analyse how it responds to severe scenarios and whether atypical behaviour generates the right flags.

The control process should consider how AI interacts with the stakeholders (customers, underwriters) and what the touchpoints are. It is particularly important for firms to test the customer journey, from the initial engagement to the outcome that is produced by AI solutions, and to do so with sufficient frequency as to identify and, if need be, rectify anomalies and outliers at an early stage.

Equally, firms should have a well governed "hand to human" process in place – i.e. the point at which the AI solution hands over to a human – for when the algorithm cannot produce an output within the predefined risk tolerances (e.g. if it cannot price a case with sufficient confidence it should pass it on to a human underwriter).

This is likely to require regular and frequent testing and monitoring of AI solutions, far beyond the development stage of the solution and training on the initial data set.

Key performance metrics should be designed using an out of sample test – i.e. running AI models using completely new data for which the tester knows the correct outcome. Frequent and/or continuous testing and statistical analysis of the algorithm (including model drivers) should be conducted to gain assurance that the AI solution is performing in line with expectations, and the firm's risk appetite, when new and updated data sets are used in a production environment.

Finally, one way to manage model risk and increase transparency is to build a modular solution where a number of smaller algorithms, narrower in scope, are used to determine the final output, rather than one single and more complex one. This makes the identification of inferences and decision drivers easier to understand and control.

3. Control – Example

- The algorithm should be trained to understand different outcomes of a decision driver – for example, for a property insurance pricing algorithm trained for surveying cracks in a building through satellite feeds, pictures of buildings with cracks, and without cracks, should be used for training the solution.
- Once the assessment process identifies any variances (positive or negative) with a non-AI pricing system, firms should put in place the relevant manual review requirements, or other model constraints.
- Controls over an insurance pricing model should span the validity, relevance and accuracy of the algorithm as well as the data:
  - Accuracy of the algorithm – as mentioned above the results of the algorithm should be checked against the results of a non-AI pricing system to check for accuracy of model performance. In addition, the algorithm should be tested on different data sources, to analyse the consistency of the risk weights generated for pricing.
  - The algorithm should be trained and tested on different data sets to ensure the outputs remain valid when the model is confronted with new data. Different methodologies exist, but one way to do this is to partition the available data set (e.g. past insurance pricing data) and train the algorithm on only, for example, 80% of the data. The remaining 20% of data can then be used to test the results, and confirm the accuracy and fairness of outcomes.
  - Controls should ensure that the algorithm has a good level of accuracy in the training data, and maintains a relatively stable level of accuracy once fed continuously with live data, i.e. the algorithm indicates the best possible insurance policy price considering the pricing criteria it was designed to identify.
  - Data representation – different data sources over different time periods are tested on the same algorithm to check for data bias. Additionally, biased data is fed into the algorithm to see if the outcomes reflect the bias.

4. Monitor and report
Due to the continuously evolving nature of learning algorithms, a more dynamic monitoring approach will be required to ensure a model is still performing as intended for its specific use case.

Moreover, the limits and targets associated with AI solutions (e.g. Key Performance Indicators - KPIs) need to be more regularly monitored for appropriateness, relevance and accuracy.

Monitoring and reporting should cover both the technical performance of the model and also the business and operational outcomes achieved by it.

Monitoring should also include all legal and regulatory developments that require a change in the design of the model as well as external events that would indirectly feed into the data consumed by the model and influence the outcomes. Static technology solutions are affected by these developments too, but in their case the impact on decision drivers and the outcome can be relatively easily identified. The evolving decision drivers in AI solutions make it harder to isolate, assess and monitor the impact of external events on the decision drivers.

Due to the continuously evolving nature of learning algorithms, a more dynamic monitoring approach will be required to ensure a model is still performing as intended for its specific use case.

4. Monitor and report – Example

- Firms should define clear and precise success metrics/KPIs for the monitoring of their algorithms. The suggested metrics should encompass the firm's fairness and anti-discrimination policy. A simple metric, for example, is the number of times someone is rejected for a policy in the out-of-sample test - if a certain group of people is consistently rejected during testing, it could indicate some degree of bias.
- The algorithm should be assessed against the predefined success metrics: the firm should assess if the algorithm used has produced discriminatory results, and if measures have been taken to counteract any discriminatory effect.
- Relevant staff should monitor potential market or regulatory changes that could have an impact on the design of pricing models for insurance policies. For example, any regulation that would strengthen the definition of fair treatment of vulnerable customers would require a revision of the algorithm to make sure that it does not lead to discriminatory output.
- Complaints from customers who consider themselves having been discriminated against are included in the review process of the algorithm, and changes to the algorithm are made accordingly, if necessary.
- Ongoing analysis of model performance should be performed by humans:
  - edge case analysis (e.g. a comparison between someone who was only "just" rejected for a policy and someone who was only "just" approved for a policy); and
  - feedback corrections to the model after validation.
- Incoming data distribution should be analysed to ensure no underlying change in the datasets feeding the model.
- Business KPIs should include metrics such as a comparison of the value of premiums, loss ratios, cost of sales and overall profit generated by the AI model with those generated by non-AI pricing models.
- Profit and portfolio mix should be monitored to ensure that there is no material drop-off of certain customer groups due to increased pricing or, equally, profits from a certain customer subset have not sharply increased due to customers not being treated fairly.
- Operational monitoring should include capturing and comparing metrics such as the volume of transactions being pushed to a human underwriter by the AI sytem and the speed with which policies are written when the AI solution is deployed relative to non-AI systems.

5. What are regulators likely to look for?

Understanding the implications of the use of AI by regulated firms is already high on the agenda of regulators and supervisors, as the number of speeches and discussion papers issued by international, EU and UK authorities demonstrates.

In general, firms planning to adopt, or already using, AI can reasonably expect that the level of scrutiny from their supervisors will only increase in future.

Although there is no prescribed set of rules for AI, existing rules and supervisory statements relating to the use of algorithmic trading, supervision of internal models, the Senior Managers and Certification Regime (SM&CR) in the UK, and the wider requirements around systems and controls give a good indication of what regulators and supervisors are likely to expect in relation to governance and risk management around AI.

Based on these sources, as well as our own experience with our clients, we have set out a summary of some of the key regulatory-related principles and measures firms should consider when adopting AI. These principles have been derived, in large part, from the well-developed use case of algorithmic trading. The extent to which these considerations apply to other AI use cases will depend on their nature and complexity.

Firms planning to adopt, or already using, AI can reasonably expect that the level of scrutiny from their supervisors will only increase in future.

Governance, oversight and accountability

- Supervisors will expect firms to have in place robust and effective governance, including a RMF, to identify, reduce and control any of the risks associated with the development and ongoing use of each AI application across the business. The RMF should be approved by the board and firms should be able to explain to their supervisor how each AI application works and how it complies with the applicable regulatory requirements and the firm's risk appetite.

- Due to the fast evolving nature of AI and/or the increasing levels of adoption of AI solutions within an organisation, risk exposures and associated controls should be reviewed regularly to confirm that they remain in line with the firm's risk appetite. This should consider factors such as the extent of use of AI within the organisation, the organisation's internal AI capabilities and external threats and events.

- In line with accountability regimes, particularly the SM&CR in the UK, supervisors will expect firms to have clear lines of responsibility and accountability, including a clearly identified owner for each AI application. The owner will be responsible for reviewing and approving AI algorithms, in line with clearly defined testing and approval processes. The owner should also be responsible for initiating the review and updating of AI applications, whenever there are relevant factors (e.g. market or regulatory changes) that may affect their accuracy, fairness or regulatory compliance.

- Governance committees, whose members should be trained to understand the risks associated with AI applications, should establish the testing and approval processes, including Quality Assurance metrics, and regularly review AI applications' performance to identify any emerging issues.

- Reflecting the potentially far reaching implications of AI use on the entity as a whole, effective AI governance should include a wider set of stakeholders from across the organisation. In particular, key development and testing stages should include technical AI specialists, as well as relevant representatives of the first, second and third lines of defence.

Reflecting the potentially far reaching implications of AI use on the entity as a whole, effective AI governance should include a wider set of stakeholders from across the organisation.

- Firms should also document procedures and controls in relation to manual "kill-switches" or "exit chutes", to stop an algorithm from operating as soon as an error or abnormal behaviour is detected. Firms should put in place a governance process around the use of such controls, which should include business continuity and remediation protocols.

- All AI algorithms should be subject to periodic re-validation. The frequency of such reviews will vary depending on the amount of risk firms, their customers or other market participants could be exposed to if the algorithms were to malfunction. The frequency should also take into account the degree to which an algorithm is allowed to evolve/learn over time and the volatility in the key drivers of decisions, e.g. macro-economic indicators.

Capability and engagement of control functions

- Firms need to ensure that staff in their risk, compliance and internal audit teams have adequate expertise to understand properly the risks associated with each adopted AI solution. In addition, they should be given enough authority to challenge the business owner and to impose, if necessary, additional controls to ensure effective risk management.

- The risk and compliance functions in particular should be meaningfully involved at each key stage of the development and implementation process of a new AI application, to be able to provide input into establishing suitable risk controls, determine if it fits within the risk appetite of the firm, and act as an independent check in relation to any potential conduct and regulatory risks.

- Internal audit functions should ensure that reviews of AI applications and models are part of their audit planning process, and should consider whether more continuous monitoring is required.

Documentation and audit trails

- Firms should have a clear and full overview of all AI applications deployed through their organisation, as well as the relevant owners, and the key compliance and risk controls in place.

- Testing and approval processes should be documented, including a clear explanation of the conditions that AI models need to meet before they can be implemented into a live environment.

- Similarly, supervisors will expect firms to have a process in place for tracking and managing any identified issues to an auditable standard.

- Finally, any variation to existing algorithms should also be clearly documented. Firms should define what amounts to a significant change and ensure all criteria are consistently applied across the business. Any significant change should be subject to rigorous and documented testing, the extent of which should be commensurate with the risks to which the change may expose the firm.

Regulated firms cannot, under any circumstance, outsource responsibility for meeting their regulatory obligations to a third party.

Third-party risk and outsourcing

- Regulated firms cannot, under any circumstance, outsource responsibility for meeting their regulatory obligations to a third party. Consistent with this, any AI models and the associated risk controls which have been designed and provided by external vendors should undergo the same rigorous testing and monitoring as those developed in-house prior to deployment.

- Firms should design effective business continuity arrangements to maintain their operations in case the AI solutions developed by third-party providers stop functioning, or the provider is unable to provide its services, for example, as a result of a cyber-attack. This is particularly relevant due to the current relatively small number of enterprise AI third-party providers in the market, including a prevalence of small start-ups.

AI and GDPR

Firms are increasingly using AI solutions to design tailored services and products that better suit customers' needs, and also to determine customers' individual risk profiles more effectively.

Being able to leverage these technologies is predicated on the availability of large sets of relevant customer data. As GDPR goes live, firms' ability to use customers' data while remaining compliant with data protection requirements will be tested.

GDPR will give consumers additional rights to understand and take control of how firms are using their personal data. Firms whose business models rely on wholesale processing of customers' personal data – regardless of whether or not they use AI solutions - will need to prepare appropriately before May 2018. This means being able to satisfy supervisors once supervisory programmes start and, importantly, also to respond to customers' enquiries in a meaningful, transparent and understandable manner.

"[...] where a decision has been made by a machine that has significant impact on an individual, the GDPR requires that they have the right to challenge the decision and a right to have it explained to them. [...]"

Elizabeth Denham, UK Information Commissioner, oral evidence to the House of Commons Science and Technology Committee, January 2018[6]

To be in a defensible position by GDPR's May 2018 implementation deadline, firms will need to have a plan to complete Data Privacy Impact Assessments for AI applications processing customers' personal data as they evolve, and if necessary, put in place a remediation plan to ensure ongoing compliance.

More generally, firms should adopt the principles of algorithmic accountability and auditability – these require firms to have organisational and technical processes in place to demonstrate, and for third parties to be able to check and review, that an algorithm is compliant with data protection requirements. Last, but not least, firms will also need to ensure the data used for the processing meets the test of being lawful to use and free of bias.

"[...] We may need, as a regulator, to look under the hood or behind the curtain to see what data were used, what training data were used, what factors were programmed into the system and what question the AI system was trained to answer."

Elizabeth Denham, UK Information Commissioner, oral evidence to the House of Commons Science and Technology Committee, January 2018

GDPR will require a gear shift in relationships with data protection supervisory authorities, both at firm and industry level. This means that firms will need to establish more structured and appropriately funded regulatory affairs teams to conduct regular briefings with the data protection supervisory authorities to discuss their data privacy strategy and any high-risk automated data processing being planned.

6 http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-decisionmaking/oral/77536.html

6. Regulating AI – some reflections

Understanding the implications and risks of the increasing use of AI is not only a challenge for FS firms, but also for their regulators and supervisors. The latter recognise that AI could bring efficiency gains to financial markets and benefits to consumers, in the form of better service and tailored offerings. In fact, regulators and supervisors themselves have been exploring ways to use AI in their own work.

However, as we mentioned earlier, regulators are also increasingly mindful of the potential risks and unintended consequences that the use of AI by regulated firms may have. From a financial stability perspective, potential network and herding effects and cybersecurity are some of the major areas of concern; from a conduct perspective, regulators are mindful of the potential for unfair treatment of customers and mis-selling resulting from inaccurate AI models, the lack of customer understanding about how their data is processed, any increase in financial exclusion, and negative outcomes for vulnerable customers.

As is the case for firms, most of these risks are not new to regulators. The challenge that regulators face with respect to AI, and innovative technologies more generally, is finding the right balance between supporting beneficial innovation and competition and protecting customers, market integrity, and financial stability.

Finding such a balance is made particularly difficult by the mismatch between the pace at which new technologies evolve and are adopted, and the speed at which new regulations can be developed and implemented. For example, the second Markets in Financial Instruments Directive (MiFID II) was first proposed in 2011[7] to address, amongst other things, the increasing use of algorithmic trading in financial markets. However, MiFID II only became applicable seven years later, in January 2018.

Regulators are conscious of this lag, and historically have addressed this issue by adopting the principle of "technological neutrality", i.e. that the same regulatory principles apply to firms regardless of the technology they use to perform a regulated activity. Technologically neutral regulation does help reduce the risk of rules becoming obsolete quickly, but it also may hinder regulators' ability to address risks specific to individual technologies and use cases.

However, we see some signs of regulators being prepared to move away from their technology neutral position, if and when the use of a specific technology becomes, or has the potential to become, systemically important. The MiFID II rules on algorithmic trading[8] are a case in point.

We are also increasingly seeing regulators issuing detailed and technology-specific guidance to clarify their expectations for firms in a number of areas including robo-advice[9], outsourcing to the "cloud"[10][11], and, again recently, algorithmic trading[12].

Technologically neutral regulation does help reduce the risk of rules becoming obsolete quickly, but it also may hinder regulators' ability to address risks specific to individual technologies and use cases.

7 https://www.fca.org.uk/mifid-ii/1-overview
8 http://ec.europa.eu/finance/docs/level-2-measures/mifid-rts-06_en.pd
9 https://www.fca.org.uk/publication/consultation/cp17-28.pdf
10 https://www.eba.europa.eu/-/eba-issues-guidance-for-the-use-of-cloud-service-providers-by-financial-institutions
11 https://www.fca.org.uk/publications/finalised-guidance/fg16-5-guidance-firms-outsourcing-%E2%80%98cloud%E2%80%99-and-other-third-party-it
12 https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/consultation-paper/2018/cp518.pdf?la=en&hash=89AB31B883DF430E36387BACCC93F15FC7A75A4A

In relation to AI, we believe regulatory guidance is a powerful tool to help firms understand the supervisor's expectations with respect to their risk management approaches. This in turn could give governing bodies and senior executives either additional confidence to progress their innovation plans, or help them identify issues that need to be addressed. More AI-specific guidance would also facilitate the work of the supervisors themselves, as it would make oversight activities more consistent, increasing their ability to identify industry-wide compliance gaps and residual risk.

The challenge is that, to be truly effective, i.e. give firms enough information about how supervisors expect firms to comply with existing regulatory regimes, any AI guidance will need to be use case specific, rather than general. Although some of the principles set out in the recent Prudential Regulation Authority draft supervisory statement on algorithmic trading[13] are relevant to other AI applications, their real power resides in their specificity to algorithmic trading activities. Given the breadth and complexity of AI use cases, regulators will need to use a risk-based approach to select carefully where to focus their, limited, resources. Regulatory sandboxes, TechSprints[14], and industry roundtables will continue to be essential fora for the regulators to be able to do so effectively.

Another tool in the regulators' box is to define the issues to be addressed, but call upon industry to develop the relevant AI standards and codes of conducts. This is similar to what the UK's Competition Market Authority did following its retail banking investigation, when it required the nine largest UK banks to develop Open Application Programme Interfaces standards[15]. Such an approach seems to be supported by the UK Information Commissioner, who recently explained[16] that, in the context of AI and data protection, sector specific codes of conduct developed by the industry, but certified by the relevant regulators, is a likely way forward.

Finally, the regulation of AI is not solely a challenge for the FS sector, nor can it be contained by geographical boundaries. Regulators and supervisors will increasingly need to overcome national and sectoral borders, and work with a broad range of counterparties not only to develop policies that address emerging risks effectively, but also to address broader public policy and ethical concerns.

Another tool in the regulators' box is to define the issues to be addressed, but call upon industry to develop the relevant AI standards and codes of conducts.

13 https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/consultation-paper/2018/cp518.pdf?la=en&hash=89AB31B883DF430E36387BACCC93F15FC7A75A4A

14 https://www.fca.org.uk/firms/regtech/techsprints

15 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/600842/retail-banking-market-investigation-order-2017.pdf

16 https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2017/algorithms-in-decision-making-17-19/

7. Conclusion

AI will increasingly become a core component of many FS firms' strategies to deliver better customer service, improve operational efficiency and effectiveness and gain a competitive advantage.

Overall, however, adoption of AI in FS is still in its early stages. Firms are still learning about the technology and which use cases could deliver the most value for them, given their individual business models, products and services.

An essential part of this learning process involves understanding the implications of AI from a risk perspective – this is not only a business imperative, but also a regulatory one, given how extensively regulated the FS sector is.

It is important for firms to recognise that this is a two-way learning process – the board, senior management teams and business and control functions will need to increase their understanding of AI, while AI specialists will benefit from an understanding of the risk and regulatory perspectives, to the extent that they do not already have them. FS firms that identify such cross-functional teams and incentivise them to collaborate in this way will be better able to exploit the benefits of AI.

This "partnership" will enable firms to recognise that AI will introduce some important differences in the way familiar risks (e.g. bias) may manifest themselves, or the speed and intensity with which they will materialise. This means that, when adopting AI, firms will need to consider carefully how AI-specific considerations can be integrated into their existing RMFs to ensure they remain fit-for-purpose, and able to give businesses the confidence that AI-related risks can be effectively identified and managed, within the limits set by the firm's culture and risk appetite.

Regulators are also increasingly mindful of the potential risks and unintended consequences of AI adoption in FS, and the challenge of finding the right balance between supporting beneficial innovation and competition and safeguarding customers, market integrity, and financial stability. Possible responses may include departing, in some cases, from their current position of technological neutrality, and/or calling on industry to work with them to develop AI standards and codes of conduct for specific applications.

We should also recognise that the challenges of regulating AI are not unique to FS, and it is important for both the industry and regulators to work together and contribute to the cross-border and cross-sectoral debate about the long-term societal and ethical implications arising from widespread adoption of AI, and what the appropriate policy response should be.

It is important for firms to recognise that this is a two-way learning process – the board, senior management teams and business and control functions will need to increase their understanding of AI, while AI specialists will benefit from an understanding of the risk and regulatory perspectives, to the extent that they do not already have them.

Contacts

Paul Garel-Jones
Partner, Risk Advisory
Lead Partner, Artificial Intelligence
pgareljones@deloitte.co.uk

Jack Pilkington
Partner, Risk Advisory
Head of FS Technology Risk and Controls and Lead Risk Advisory Innovation Partner
jpilkington@deloitte.co.uk

Gurpreet Johal
Partner, Consulting
Lead Partner, Artificial Intelligence
gjohal@deloitte.co.uk

David Strachan
Partner, Risk Advisory
Head of EMEA Centre for Regulatory Strategy
dastrachan@deloitte.co.uk

Authors

Tom Bigham
Director, Risk Advisory
Technology and Digital Risk Management Lead
tbigham@deloitte.co.uk

Suchitra Nair
Director, Risk Advisory
EMEA Centre for Regulatory Strategy
snair@deloitte.co.uk

Sulabh Soral
Director, Consulting
Artificial Intelligence
ssoral@deloitte.co.uk

Alan Tua
Director, Risk Advisory
Artificial Intelligence Lead
altua@deloitte.co.uk

Valeria Gallo
Manager, Risk Advisory
EMEA Centre for Regulatory Strategy
vgallo@deloitte.co.uk

Michelle Lee
Manager, Risk Advisory
Artificial Intelligence
michellealee@deloitte.co.uk

Tom Mews
Manager, Risk Advisory
Technology and Digital Risk Management
michellealee@deloitte.co.uk

Morgane Fouché
Senior Associate, Risk Advisory
EMEA Centre for Regulatory Strategy
mfouche@deloitte.co.uk

CENTRE *for*
REGULATORY STRATEGY
EMEA

The Deloitte Centre for Regulatory Strategy is a powerful resource of information and insight, designed to assist nancial institutions manage the complexity and convergence of rapidly increasing new regulation.

With regional hubs in the Americas, Asia Paci c and EMEA, the Centre combines the strength of Deloitte's regional and international network of experienced risk, regulatory, and industry professionals – including a deep roster of former regulators, industry specialists, and business advisers – with a rich understanding of the impact of regulations on business models and strategy.

Deloitte.

Designed and produced by The Creative Studio at Deloitte, London. J15117