IBM **Services**

IBM and Promontory Controlling AI—How to regulate and safely exploit the proliferation of AI

By Michael Conway,
Sebastian Weir
and Promontory

IBM

Summary

As the use of Artificial Intelligence (AI) permeates the Financial Services sector, it has become increasingly important to trace and challenge its decisions, whilst regulating both internally and externally. It is not the popularity of AI that is of concern; it is whether the AI is fair and whether the management that owns the AI can be held to account through their actions and areas of conduct, albeit in a smart, lightweight and focussed manner.

Whilst the discussion is underpinned by IBM's Principles of Trust and Transparency[1], we see two distinct ways of looking at AI from a regulatory point of view. Firstly, it is the responsibility of Firms to understand their AI and so to ensure that they treat their customers fairly and transparently and for it be regulated as such.

Secondly, is how these Firms can best leverage AI to aid industry regulations: to spot anomalies and early warning signs that things need to be investigated. This latter point then leads us to the more positive question of how regulation can best embrace the AI revolution and enforce future legislation. We will now view each of these two lenses in turn.

Regulating the AI

Never has the application of Martec's Law[2] been more pervasive than with the application of AI to the enterprise. As companies fight to compete with disruption and reduce the ongoing cost of operation, attention is turning to how AI, complemented with automation, can be fully utilised to answer these challenges. With this comes the conflicting pace of change in organisational process and upskilling.

There is an ever-growing delta between the capabilities of the technology and the cultural approach that Firms are taking to understand, control and evidence an eco-system that takes close nurturing to perform with the requisite accuracy and efficiency to deliver long-term sustainable value.

An AI ecosystem requires daily monitoring, regular training and intervention alongside integration with a growing suite of capabilities to complement the expectation of customers and executives. Whilst the technology is capable of performing this capability consistently today, what is becoming increasingly apparent is that the balancing act of evidence and risk is not keeping up. Too often, traditional technology and software approaches are being applied to this new capability. In most cases, it is not as simple as applying an existing approach, controls and agile methodology to this new technology and expecting this to work. Some re-engineering of our control frameworks is required in order to appropriately manage AI technology and give it the best chance to flourish.

Given the unprecedented pace of change, rules of engagement for Firms are quickly shifting. The most notable of these is how senior managers, specifically under the remit of the Senior Managers regime are increasingly being expected to address emerging and innovative technologies that include AI. Incomplete or faulty decision making resides with the person responsible for that area. To bring this to life, imagine a scenario in which the AI is responsible for assessing AML risk. It is the person that is responsible for that area's incorrect decisions that takes the blame for errors and not someone with a broader functional or oversight role, say an SMF 24, whose responsibility is more related to operational resilience.

At board level, the core mandate remains to enable and encourage management to deliver superior, long-term shareholder value through the pursuit of a growth agenda, assisted by measurable KPIs and operational resilience. As all forms of technology become an increasingly important risk area for boards, much of this remit falls under cognitive technology.

With this and our collective experience in mind, AI needs to be "better than the baseline human being" to be considered a success within an organisation. Put more simply, the AI engine must be able to perform at an equivalent or higher standard than the typical human employee. This is true from both a business outcome point of view, as well as a regulatory one. The former, because if you can do the same tasks with technology that doesn't require breaks, pay or sleep then your cost of operation goes down whilst your speed to execute goes up. For the latter, if AI is able to do at least as well as the human in servicing customers then this is the baseline that should be measured against.

This should not be construed as a recommendation to replace Humans with AI, this is simply an opportunity to augment the human role with the technology, taking the mundane tasks away allows Agents to dedicate time to customer service in more sensitive or complicated areas.

Too often we are seeing clients expecting AI to be the panacea and be correct 100% of the time. Both clients and the regulator need to keep in mind the "human baseline" to assess whether the technology is performing adequately or not. To expect any more is unreasonable.

Additionally and especially in the customer servicing arena, the fact that with AI, you have "one brain to train" that answers customer's queries, means that you are reducing the level of risk in your operation compared with the often tens of thousands of employees that have their own free will and desires.

With that being said, and as more powerful AIs proliferate in society, the ability to trace their decisions, challenge them and ensure that they work in an ethical way are becoming a pressing concern for Firms. In order to explain the AI's behaviour to regulators, covering a legal or ethical dimension, we need to take an AI's decision and work backwards through the control points of the program's neural network. We are therefore able to show in a transparent manner how the technology came to a decision, and evidence how this decision is in line with the Firm's policies, procedures and approaches. What is more, as this capability begins to permeate across the enterprise, taking a greater control of central decisioning, then the importance of regulatory considerations grows exponentially.

Today most FTSE250 companies are exploring the application of AI, with the overwhelming majority doing so in a risk-averse manner that will not impact their Firm's performance or reputation in any meaningful sense. However, as the market and capability for AI grows, the dial of risk appetite will change dramatically over the course of the next 18 months whereby the attention of regulatory functions will pay increased scrutiny to the application of AI and the conduct-related customer outcomes it is determining.

Currently, training AI systems is best undertaken in an assisted state, meaning humans are required to educate AI about the rights and wrongs of the decisions it is making. This provides auditable training records to support the reverse engineering of the decisioning and so also increase transparency. This does however come with the inherent risk of introducing bias into the system based on the training undertaken and the individuals that implemented it.

Both conscious and unconscious bias is something that requires a watchful eye when we are considering the appropriate controls and evidence necessary to demonstrate that the eco-system is safe. By evidencing the control points throughout the AI decisioning, using increased dependency on core data science capabilities we can reactively evidence how bias may have entered a system and take account of it in future iterations of the technology. Not only that, there are also active steps teams can take to mitigate this bias. Diversity in teams, both in terms of background and thinking help minimise the risk as individuals will bring very different points of view to bear. Often mixing the teams up and having independent checkers also helps to manage the risk of "group think" where every group tends to end up agreeing with each other if they work together for long enough.

To that end, we come back to culture being one of the key risk mitigants for the proliferation of AI technology. As Firms embrace cognitive technology more, the concept of the "Cognitive Enterprise" helps ground this into something tangible. This enterprise is composed of multiple business platforms, where each underlying business unit is itself made up of capability layers, supporting major digital transformation.

The key layers include:

- A culture of agile innovation that embraces new skills, workforces and ways of working, and humanising the enterprise.
- Cognitively enabled workflows for front and back-office processes and decision making.
- Applied Science in technology—including deep machine learning and advanced analytics.
- Data that is curated to support key workflows and platforms.
- Next-generation applications that span new and legacy solutions.
- Open, hybrid and secure multi-cloud infrastructures.

The financial disruption that has taken place over the course of the past five years, has driven Firms to turn their attention to exposing legacy platforms and data in a manner that makes it far easier to integrate. This creates the opportunity for AI to digest, process and present this information more efficiently to reduce the cost of operation or better inform Next Best Actions whilst giving customers a far better experience.

There are a number of things executives and boards can do now. These include making sure there is strong governance around AI's usage. This includes but is not limited to:

- Management should ensure the bank's policies, procedures and practices don't create compliance, conduct or non financial risk problems.
- There must be an audit trail surrounding the use of AI and its decisioning that can be thoroughly explained to regulators by the senior person overseeing this technology.
- This audit trail must be monitored to make sure that AI is producing understandable outcomes and isn't being used where there isn't sufficient reason or experience to rely on it.
- The risk function should upskill itself in these new areas of technology so they are able to ask the pertinent questions to the programmes delivering the change.

Management should also carefully scrutinise AI vendors. Unreliable or unproven AI companies may not have given regulators' expectations as much attention as more established technology providers. It is most cost effective to select technology and other innovative approaches that take regulatory concerns into account on the front end. To patch and fix after a system is partially or fully installed—even if possible—can be highly disruptive and costly.

The application of AI is increasingly relevant to the way in which all companies will do business over the course of the next two years. Appropriate attention needs to be given today to ensure that this application is implemented in a safe, controlled and compliant manner. This will in turn ensure that any growth can be predicated on an ecosystem of trust and transparency.

If achieved, the industry will enjoy unparalleled growth and success; if ignored, the next wave of rectifications could be looming for many.

Regulating with AI

Now to our second lens; how do we use this technology to actually aid us in spotting issues before we could even imagine them previously. Given AI's potential, the industry can expect regulators to have an increasing number of questions about how Firms are using this technology. That will include how well it works, how the decisions are made, the quality of the data it uses, whether it demonstrates fair customer outcomes, and whether decisions that the AI makes have an auditable trail and how it is governed.

At this time, regulators' rules and regulations aren't specific around the use of AI, as the technology is still new and still evolving. Even still, specific rules about the use of AI are not required, whether it is new or not. But regulators have latitude to determine whether certain practices are unsafe and unsound—which could include misuses of AI.

Over time, regulators will want to understand AI applications and be able to examine matters for themselves. Eventually, legislatures' interest in how quantitative models are designed, tested and validated will extend to AI—Firms will need to be able to demonstrate that the technology isn't harming consumers or creating undue risk to the system. Regulators will also attempt to use AI, but to supplement and enforce regulatory compliance.

The use AI and machine learning can be applied to better and more efficiently understand the Regulator's rules, whilst saving the regulated time and money that could be invested elsewhere.

Already we are seeing usage of AI in spotting AML patterns and stopping large cases of fraud. The more data there is available, and the more sophisticated Firms get in utilising that data, the more apparent it will become that whilst the risk of AI needs careful consideration, so does the opportunity in keeping us all safe.

In closing, at the time of writing, the EU are drafting and publishing Guidelines on how it plans to approach Regulating AI[3]. Whilst IBM have been working closely with the EU on how best to implement Trustworthy AI[4]. Subsequent points of view will consider the arguments and perspectives raised in this Paper in light of their approach .

Possible themes in further iterations of this code could include: standards for system design and development of cognitive technologies; guidelines for the reduction of bias, conduct and ethics; and ways of monitoring and mitigating the non-financial risks that AI could introduce.

What is for sure, is that we stand at a point in time where AI is going to start having a fundamental impact on how the regulator and the regulated go about their business. How they arm themselves in mitigating the risks that the opportunity affords them will separate the winners from the losers.

Authors

Michael Conway (IBM Services AI Leader)
Sebastian Weir (IBM Services, AI Strategist)
With the support of Promontory

IBM Corporation
Route 100
Somers, NY 10589

1 IBM’s Principles for Trust and Transparency: ibm.com/blogs/policy/wp-content/uploads/2018/05/IBM_Principles_OnePage.pdf.

2 **Martec’s law** states that technology changes exponentially, but organisations change much more slowly (logarithmically).

3 Ethics guidelines for trustworthy AI:ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

4 IBM Statement on EU Ethics Guidelines for Trustworthy AI: ibm.com/blogs/policy/ai-ethics-eu/.

Please Recycle

79025479GBEN-00