BANK OF ENGLAND

Speech

Managing Machines: the governance of artificial intelligence

Speech given by

James Proudman, Executive Director of UK Deposit Takers Supervision

FCA Conference on Governance in Banking

4 June 2019

I am very grateful to Philip Sellar for preparing these remarks, and to Sadia Arif, Jamie Barber, Jas Ellis, Magnus Falk, Orlando Fernandez Ruiz, Anna Jernova, Carsten Jung, Tom Mutton, Lyndon Nelson, Jennifer Nemeth and Oliver Thew for very helpful advice and comments.

Introduction

Consider the well-known story of one Big Tech company's attempt to use artificial intelligence to improve the efficiency of its staff recruitment. The machine learning system reviewed job applicants' CVs with the aim of automating the search for top talent. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars. Within a year, the company realised its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way. That is because the computer models were trained to vet applicants by observing patterns in CVs submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the technology industry. In effect, the company's system taught itself that male candidates were preferable. It penalised CVs that included the word "women's" as in "women's chess club captain". And it downgraded graduates of two all-women's colleges.

The story is a clear example of how artificial intelligence can produce bad outcomes for all concerned. It also offers a case study for exploring the root causes that lead to bad outcomes - and so in turn offers insights for boards on how to govern the introduction of artificial intelligence. The art of managing technology is an increasingly important strategic issue facing boards, financial services companies included. And since it is a mantra amongst banking regulators that governance failings are the root cause of almost all prudential failures, this is also a topic of increased concern to prudential regulators.

In my comments, I will provide a quick overview of the scale of introduction of artificial intelligence in UK financial services; then make three observations about it, and suggest three principles for governance derived from them.

Artificial Intelligence

Technological innovation is inevitable and welcome. As the Governor noted during this year's UK Fintech Week,[1] we are shifting towards a new economy, which is powered by big data, advanced analytics, smartphone technology and more distributed peer-to-peer connections. This new economy will go hand in hand with fundamental changes to the structure and nature of the financial system supporting it. Indeed some of the largest international investment banks are now declaring that they are technology companies with banking licences.

Some of the most important recent developments in technology are associated with the introduction of automation – by which I mean the replacement of humans by machines for conducting repetitive tasks. Within the broad concept of automation, two areas pose unique challenges for governance. These are artificial intelligence (AI) – by which I mean the use of a machine to perform tasks normally requiring human intelligence – and by machine learning (ML) – by which I mean the subset of AI where a machine teaches

[1] Carney, M., (2019) 'A Platform for Innovation'

itself to perform tasks without being explicitly programmed. The focus of my remarks is therefore on AI and ML, but the principles I discuss could be applied to automation more broadly.

It is certainly not the role of the regulator to stand in the way of progress. Indeed, AI/ML in financial services could herald an era of leaner, faster and more tailored operations, reducing costs and ultimately improving outcomes for customers. It could also help to make financial services more bespoke, accessible and inclusive. In capital markets, there is some evidence from market participants to suggest that automation is leading to increased effectiveness. According to the IMF, two thirds of cash equities trading by volume is now associated with automated trading, about half of FX spot trading.[2] And the Governor recently explained how AI provides the opportunity to reduce bias.[3] On balance, it is probable that increased automation will enhance net resilience of institutions, and support the PRA's statutory objectives.

For example, AI and ML are helping firms in anti-money laundering (AML) and fraud detection. Until recently, most firms were using a rules-based approach to AML monitoring. But this is changing and firms are introducing ML software that produces more accurate results, more efficiently, by bringing together customer data with publicly available information on customers from the internet to detect anomalous flows of funds. About two thirds of banks and insurers are either already using AI in this process or actively experimenting with it, according to a 2018 IIF survey.[4] These firms are discovering more cases while reducing the number of false alerts. This is crucial in an area where rates of so-called "false-positives" of 85 per cent or higher are common across the industry.

ML may also improve the quality of credit risk assessments, particularly for high-volume retail lending, for which an increasing volume and variety of data are available and can be used for training machine learning models.

But it is a prudential regulator's job to be gloomy and to focus on the risks. We need to understand how the application of AI and ML within financial services is evolving, and how that affects the risks to firms' safety and soundness. And in turn, we need to understand how those risks can best be mitigated through banks' internal governance, and through systems and controls.

Firms' rates of adoption of AI and ML

So what do we know about how – and how fast – the application of AI and ML is evolving within UK financial services? In 2018, an FT survey of banks around the world provided evidence of a cautious approach being taken by firms.[5] And according to a McKinsey survey of financial and non-financial firms, most barriers to the more rapid adoption of AI were internal to the firms themselves: poor data accessibility; lack of suitable

[2] IMF (2019) 'Global Financial Stability Report April 2019'
[3] Carney, M., (2019) 'AI and the Global Economy'
[4] IIF (2018) 'Machine Learning in Anti-Money Laundering'
[5] FT (2018) 'AI in banking reality behind the hype'

technology infrastructure; and a lack of trust in the insights of AI, for example.[6] Despite the plethora of anecdotal evidence on the adoption of AI/ML, there is little structured evidence about UK financial services.

To gather more evidence, the Bank of England and the FCA sent a survey in March to more than 200 firms, including the most significant banks, building societies, insurance companies and financial market infrastructure firms in the UK. This is the first systematic survey of AI/ML adoption in financial services.

The survey is focused on building our understanding of key themes. First, the extent to which firms have adopted, or are intending to adopt, AI/ML within their businesses and what they regard as the most promising use cases. Second, the extent to which firms have clearly articulated strategies towards the adoption of AI/ML. Third, the extent of barriers - regulatory or otherwise - to adoption and what techniques and tools can enable safe use of this technology. Fourth, an assessment of firms' perceptions of the risks, to both their own safety and soundness as well as to their conduct towards customers and clients, arising from AI/ML. And fifth, the extent to which the appreciation of these risks has given rise to changes in risk management, governance and compliance frameworks.

The full results of the survey will be published by the Bank and FCA in Q3 2019, and are likely to prove insightful. Responses were returned to the Bank in late April, so some early indicative results are emerging.

Overall, the mood around AI implementation amongst firms regulated by the Bank of England is strategic but cautious. Four fifths of the firms surveyed returned a response; many reported that they are currently in the process of building the infrastructure necessary for larger scale AI deployment, and 80 per cent reported using ML applications in some form. The median firm reported deploying six distinct such applications currently, and expected three further applications to go live over the next year, with ten more over the following three years.

Consistent with the McKinsey survey, barriers to AI deployment currently seem to be mostly internal to firms, rather than stemming from regulation. Some of the main reasons include: (i) legacy systems and unsuitable IT infrastructure; (ii) lack of access to sufficient data; and (iii) challenges integrating ML into existing business processes.

Large established firms seem to be most advanced in deployment. There is some reliance on external providers at various levels, ranging from providing infrastructure, the programming environment, up to specific solutions.

Approaches to testing and explaining AI are being developed and, perhaps unsurprisingly, there is some heterogeneity in techniques and tools. Firms said that ML applications are embedded in their existing risk

[6] McKinsey (2018) 'Adoption of AI advances, but foundational barriers remain'

frameworks. But many say that new approaches to model validation (which include AI explainability techniques) are needed in the future.

Of the firms regulated by the Bank of England that responded to the survey, 57 per cent reported that they are using AI applications in risk management and compliance areas, including anti-fraud and anti-money laundering applications. In customer engagement, 39 per cent of firms are using AI applications, 25 per cent in sales and trading, 23 per cent in investment banking, and 20 per cent in non-life insurance.

By and large, firms reported that, properly used, AI and ML would lower risks - most notably, for example, in anti-money laundering, KYC and retail credit risk assessment. But some firms acknowledged that, incorrectly used, AI and ML techniques could give rise to new, complex risk types - and that could imply new challenges for boards and management.

Challenges of AI and ML for boards

Let me suggest that there are three challenges for boards and management.

The first challenge is posed by data. As any statistician knows, the output of a model is only as good as the quality of data fed into it – the so-called “rubbish in, rubbish out” problem. Of course, there are various techniques for dealing with this problem, but fundamentally, if there are problems with the data used – be they incomplete, inaccurate or mislabelled – there will almost certainly be problems with the outcomes of the model. AI/ML is underpinned by the huge expansion in the availability and sources of data: as the amount of data used grows, so the scale of managing this problem will increase.

Further, there are complex ethical, legal, conduct and reputational issues associated with the use of personal data. For example, are data being used unfairly to exclude individuals or groups, or to promote unjustifiably privileged access for others? Recent examples amongst retailers suggest that overly-personalised marketing can seem plain ‘creepy’. These questions require complex answers that are beyond my philosophical pay-grade. From a regulatory perspective, they are perhaps more directly a primary concern to the FCA given its statutory objectives of consumer protection, but are also potentially relevant to safety and soundness, not least through their impact on reputation and, in turn, confidence.

The data challenge is not limited simply to its selection and processing – but also to its analysis, and how inferences are drawn. AI/ML is driven by what seems to be objective historical data – but that itself may reflect longstanding and pervasive bias, as the example I used in the introduction showed. So there is a need to understand carefully the assumptions built into underlying algorithms, and how they will behave in different circumstances, including by amplifying potential and/or unintended human prejudice. This implies the need for a strong focus on understanding and explaining the outcomes generated by AI/ML. In my introductory example, the hidden flaws were only revealed over time by the outcomes. The focus of

governance, therefore, should not only be on the role of testing in the design stage, before AI/ML is approved for use, but also on testing during the deployment stage, as well as the oversight needed to evaluate outcomes and address issues when they go wrong. This includes in certain cases the ability for a human or other override – the so-called 'human in the loop', for example – and to provide feedback to minimise gradually the risk of adverse unintended consequences.

The second challenge posed to boards by AI/ML concerns the role of people – in particular, the role played by incentives. This may seem somewhat paradoxical, because the role of AI is often thought of as automating tasks formerly done by people.

Machines do not have human characteristics. But they do what they are told by humans. Humans design and control machines, and the algorithms that let those machines learn, whether that is automating the recruitment process or providing financial advice. As with any member of staff, coders, programmers and managers can be subject to the myriad of human biases, and the outputs of machines may likely reflect those biases. It follows that the regulatory reforms over recent years were developed to overcome the very 'human' problems embodied in people-centric workplaces – be they cultural failings and lack of diversity of thought; poorly aligned incentives, responsibilities and remuneration; or short-termism and other biases – remain equally relevant to an AI/ML-centric workplace.

In fact, it may even become harder and take longer to identify root causes of problems, and hence attribute accountability to individuals. For example, how would you know which issues are a function of poor design – the manufacturer's fault if you have bought an 'off the shelf' technology product – or poor implementation – which could demonstrate incompetence or a lack of clear understanding from the firm's management. In the context of decisions made by machines which themselves learn and change over time, how do you define what it means for the humans in the firm to act with "reasonable steps" and "due skill, care and diligence"? In a more automated, fast-moving world of AI/ML, boards – not just regulators – will need to consider and be on top of these issues. Firms will need to consider how to allocate individual responsibilities, including under the Senior Managers Regime.

Machines lack morals. If I tell you to shoplift, then I am committing an unethical act - and so are you, if you follow my instruction. "I was only following orders" is not a legitimate defence. There is, if you like, a double-lock on unethical instructions within a wholly human environment - on the part of the instructor and the instructed. This is one reason why firms and regulators are so determined to promote 'good' cultures, including, for example, 'speak up' cultures, and robust whistle-blowing. But there is no such double-lock for AI/ML. You cannot tell a machine to "do the right thing" without somehow first telling it what "right" is - nor can a machine be a whistle-blower of its own learning algorithm. In a world of machines, the burden of correct corporate and ethical behaviour is shifted further in the direction of the board, but also potentially further towards more junior, technical staff. In the round this could mean less weight being placed on the judgements of front-office middle management.

There have been some initial steps to promote the ethical use of big data and AI/ML in financial services. Notably, for example, in Singapore,[7] and – more broadly – within the EU.[8] In the UK, the Centre for Data Ethics and Innovation is looking at maximising the benefits of AI,[9] and many consider them leaders in this field. Principles-based expectations have focused on areas such as fairness, ethics, accountability and transparency. Nevertheless, promoting the right outcomes, even if framed as principle-based expectations, will require appropriate, up-to-date systems and controls across the three lines of defence to ensure an appropriate control environment throughout the firm. Further thought is needed.

The third challenge posed by greater use of AI/ML to boards is around change. As the rate of introduction of AI/ML in financial services looks set to increase, so too does the extent of execution risk that boards will need to oversee and mitigate.

It appears to supervisors, and consistent with the early results from the Bank of England/FCA survey, that some firms are approaching the introduction of AI/ML piecemeal, project by project; others appear to be following a more integrated, strategic approach. Either way, the transition to greater AI/ML-centric ways of working is a significant undertaking with major risks and costs arising from changes in processes, systems, technology, data handling/management, third-party outsourcing and skills. The transition also creates demand for new skill sets on boards and in senior management, and changes in control functions and risk structures.

Transition may also create complex interdependencies between the parts of firms that are often thought of, and treated as, largely separate. As the use of technology changes, the impact on staff roles, skills and evaluation may be equally profound. Many of these interdependencies can only be brought together at, or near, the top of the organisation.

Conclusion

I noted at the beginning that I would conclude by trying to extract three principles for governance from the observations I had made.

First, the observation that the introduction of AI/ML poses significant challenges around the proper use of data, suggests that boards should attach priority to the governance of data – what data should be used; how should it be modelled and tested; and whether the outcomes derived from the data are correct.
Second, the observation that the introduction of AI/ML does not eliminate the role of human incentives in delivering good or bad outcomes, but transforms them, implies that boards should continue to focus on the oversight of human incentives and accountabilities within AI/ML-centric systems.

[7]Monetary Authority of Singapore (2019) 'Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector'
[8] European Commission (2019) 'Ethics Guidelines for Trustworthy AI'
[9]Centre for Data Ethics and Innovation (2019) 'Introduction to the Centre for Data Ethics and Innovation'

And third, the acceleration in the rate of introduction of AI/ML will create increased execution risks during the transition that need to be overseen. Boards should reflect on the range of skill sets and controls that are required to mitigate these risks both at senior level and throughout the organisation.