It's not magic:
Weighing the risks of AI in financial services

Keyur Patel and Marshall Lincoln

endava
Swiss Re
CSFI
Centre for the Study of Financial Innovation

The Centre for the Study of Financial Innovation is a non-profit think-tank, established in 1993 to look at future developments in the international financial field – particularly from the point of view of practitioners. Its goals include identifying new areas of business, flagging areas of danger and provoking a debate about key financial issues. The Centre has no ideological brief, beyond a belief in open markets.

This CSFI publication is available through our website www.csfi.org
or by calling the Centre on +44 (0) 20 7621 1056

Published by Centre for the Study of Financial Innovation (CSFI)

Cover Illustration: David Bromley

Email: info@csfi.org
Web: www.csfi.org

ISBN: 978-1-9997174-7-6

Printed in the United Kingdom by Heron Dawson & Sawyer

Contents

Preface

There are no certainties in financial services. But it is at least odds-on that the future for the financial services sector is going to be shaped by Big Data – and that means Artificial Intelligence and Machine Learning, since they are the technologies (the *only* technologies) that can handle the massive quantities of information that are now being produced.

We hear a great deal about the opportunities opened up by AI and ML – and no one (sane) doubts their potential. But there is another side to the discussion, and it is one that is just as important. AI and ML are *difficult* – complex, opaque, hard for non-specialists to grasp. They may also lend themselves to anti-competitive market concentration – and they will certainly throw up a range of new challenges for incumbent and regulators. This report focusses on the skills gap, on ethical issues that are raised, on consumer protection and on systematic threats. It is a timely counterpoint to much of the AI boosterism about which we read so much.

I am very grateful to the two authors – Keyur Patel, who has worked closely with the CSFI on our *Banana Skins* surveys, and Marshall Lincoln, whose hands-on experience of life in Silicon Valley is a rare commodity. They have pulled together what I believe is the first comprehensive review of the problems that AI and ML may pose – while not ignoring the very considerable potential upside. It is a timely report – and an extremely readable one.

I am grateful to the two corporate sponsors who supported their work. Swiss Re has long been at the forefront of thinking about insurance – not least, through its own Swiss Re Institute. Endava is, equally, at the forefront of thinking about how human beings interact with technology. The CSFI is grateful to both of them for their support – and for the fact that they gave us complete editorial freedom. The conclusions of the report are entirely those of the authors – and of the CSFI, which kept a beady eye on what they were up to. Many thanks to our friends at Swiss Re and Endava – and to both the authors.

Andrew Hilton
Director
CSFI

Foreword

It is commonly agreed that the use of Big Data and Artificial Intelligence (AI) brings a number of benefits for the financial services industry. Machine Learning techniques can process far larger amounts of data more quickly and accurately than individuals. From this vantage point, algorithms have enormous potential to significantly improve the efficiency of operational and decision-making processes while providing tailored solutions to customers. In the insurance sector specifically, Big Data and AI may also help reduce the large protection gap that currently exists in Europe and globally by making more risks insurable and insurance protection more affordable to consumers.

Digitisation has the potential to redefine the relationship between consumers and financial institutions. In this new era, regulators should ensure consumer protection and financial stability while giving the industry the scope to adjust and innovate in response to the tech revolution taking place. It is, therefore, critical for the industry to identify not only the benefits associated with the use of Big Data and AI but above all the potential risks of these new technologies. In this context, the CSFI report is making an important contribution in introducing a framework which defines the main risk drivers (i.e. opacity and complexity, distancing of humans from decision making and changing incentive structures) and the key risks (i.e. new ethical challenges, skills gap and market dynamics) that might arise from the increasing use of AI in financial services. Finally, the report explores the consequences that an increased reliance on AI might have for consumers, institutions and the stability of the financial system.

Recently, ethical challenges associated with the use of AI by financial services firms have received a lot of public attention. The report points out that the use of AI creates strong incentives for financial institutions to collect, aggregate and centralise data – as well as to share these data with third parties, increasing concerns about data security and privacy. Financial services firms are information-based businesses that rely on access to data and the ability to process it. For example, insurers have always used data to fairly assess, price and manage risk, even though historically they have had less data and technology available than today. The increasing availability and use of data in this process requires a robust governance and risk management framework to ensure data security and privacy. It is also important that firms explain to consumers what data is used, how and why, as well as how that data is protected. This gives consumers confidence in the benefits of data use and builds trust that their data is managed responsibly with clear accountability for any breach of privacy.

With respect to the risks of *undesired behaviour* and *insufficient transparency of AI models*, the report highlights that the difficulty of understanding and explaining decisions made or augmented by AI could damage trust in financial services. Modelling techniques currently used in in the industry, including Machine Learning and AI-based ones, should allow firms to identify the major factors that influence an individual decision. AI tools are typically used for operational purposes and the classification of information, however, major factors influencing the decision-making process can still be determined. In this regard, it is crucial a human being remains involved in the process to minimise the risk of underlying errors or biases in the data and the algorithms used.

The report also stresses that while AI enables institutions to evaluate risks at a much more granular level, this could *disadvantage certain customers* and challenge conceptions of fairness. This is particularly applicable to the insurance sector where technological innovation and the resulting availability of data allows insurers to charge more tailored premiums reflecting the true expected costs of risk. While this leads to reduced premiums for some consumers, prices for other policyholders could increase, sometimes even beyond the point of affordability. In some cases, individuals may be able to adjust their behaviour and personal choices to reduce their risk profile and the price of their insurance cover. In some other situations, the issue may become more nuanced as it takes in factors beyond an individual's immediate control. This raises legitimate questions about fairness and requires proactive engagement from the industry to find appropriate solutions.

There is no question that technological innovation and Big Data will lead to a prolonged period of rapid evolution of the financial services industry. In this process, it is important that financial services firms show how they are contributing to building a more resilient society by using data and technology responsibly for the benefit of their customers.

Torben Thomsen
Swiss Re Chief Risk Officer for Reinsurance

Foreword

Until fairly recently, mainstream computing's problem-solving capabilities were constrained by the nature of the data available. Problems that depended on predictable data, that could be described as regular structured records, and that were stored in relational databases, could generally be solved without too much difficulty – although naturally, the solutions could be complex in certain domains. But problems that involved unstructured data, like voice, images or blocks of text, were the domain of academic research and beyond the capabilities of most mainstream computing technology.

Then, from about 2005 or 2006, an almost miraculous convergence of technologies started to change this. Vast amounts of data generated by internet-connected digital services met almost unlimited computing power in the form of public cloud computing, and these two factors collided with decades of research into machine learning algorithms, which had previously been rendered impractical because of their need for huge datasets to train on and massive computing power to process. This happy convergence of circumstances has since led to an explosion, both in theory and in practice, of machine learning, and the development of applications in the areas of predictive analytics, dataset classification, voice recognition, document interpretation, video processing and machine vision that would have been impossible only a few years before.

The potential for the application of this technology in the financial industry is huge – from the personalisation of financial products, to efficiency gains in back-office processing, to improved compliance processes, to risk and credit scoring based on novel datasets and beyond.

However, it is clear to us at Endava, as well as to many other observers and practitioners, that together with these benefits there could be real risks in applying AI-based technology to such a critical industry – one that has such a direct impact on so many people's lives.

This report from the CSFI provides a timely and thorough analysis of the opportunities that AI can offer in financial services, while also considering its limitations and the possible risks that we all need to be aware of in order to make informed decisions about how to apply it. Keyur Patel and Marshall Lincoln provide a very clear introduction to the core concepts of AI, as it relates to financial services, and explore its benefits for the industry. They then provide a clear and balanced consideration of the potential risks that it brings, particularly focusing on the ethical challenges that it may pose, the problems with the skills gap that it implies, and the changes to market dynamics that it may trigger.

The result is a thorough and balanced review of the opportunities and risks that face us as we start to apply AI technology in financial services. I believe that it will act as a useful primer for those not yet familiar with AI technology. It will also generate thought-provoking discussion for those who are currently engaged in identifying the opportunities, understanding the risks and ultimately trying to deliver the benefits of this sophisticated technology in one of the world's most fast-moving and important industries.

Eoin Woods
Chief Technology Officer
Endava

CSFI

Number 133 November 2019

It's not magic:

Weighing the risks of AI in financial services

Keyur Patel and Marshall Lincoln

Introduction

This report explores some of the risks that could arise as the use of artificial intelligence (AI) becomes increasingly prevalent in financial services – not just at the fringes, but at mainstream institutions that touch billions of people worldwide. We aim primarily to provide an accessible introduction to the subject for readers without a technical background.

It is notoriously difficult to find a universally agreed definition of AI. But in the sense of the 'narrow' AI we have access to today (as opposed to the theoretical 'general' AI that aims to replicate human intelligence), there is a consensus that it is predicated on recent advances in the field of machine learning (ML) and its subsets such as deep learning. Broadly speaking, this report is interested in the *risks* that arise as a result of collecting and processing the data that fuel ML and related statistical models.

Why do we focus on risks? No subject in financial services is sexier than AI. The World Economic Forum (WEF) underlined the scale of expected transformation by calling a horizon-scanning report on AI that it published last year: *'The new physics of financial services'*[i]. Predictions about its potential value to the industry over the next decade typically produce figures in the hundreds of billions of dollars.

Sceptics argue that such proclamations are hyperbolic, or at least premature. Some even take issue with calling it 'artificial intelligence', on the basis that this label gives a misleading impression about the capabilities of the new technology. There is disagreement about how different AI is from the types of automation that have existed in financial services for many years. Does it represent a qualitative change, or merely a quantitative one? To what extent will it be used to replace human activity, rather than better inform it? Over what timeframe? There is certainly much hype (and business opportunism) tangled up in more sober analysis of the evidence.

These debates will continue. There is, nevertheless, no denying the transformative impact of computerisation and digitisation on financial services over the past 50 years. However 'intelligent' AI is or isn't, there is a compelling argument that it represents not just an incremental improvement in computing, but a step change. For example, according to recent projections by *McKinsey*, AI will add $13 trillion of value to the global economy by 2030[ii]. That represents 1.2 per cent of additional GDP growth per year – twice the impact of IT in the 2000s.

The point is that if predictions are anywhere near the mark, it is crucial that the financial services industry gets AI right. We do not downplay the benefits the use of AI can bring to financial services. Indeed, they are explicitly the subject of Section 2. But in order to explore how the industry can better recognise these benefits, our primary focus is on risks. Increased reliance on AI will have consequences for consumers, institutions and the stability of the financial system, for better or worse.

In this report we have three main objectives:

1. To lay out the primary risks that could arise from current and near-future applications of AI in financial services.

2. To explain why these risks are important and to outline their potential consequences.

3. To provide readers with a framework for thinking about risks when presented with an unfamiliar application of AI in financial services.

We try to present our analysis in straightforward, non-technical language – which requires some simplification. Practitioners with highly technical backgrounds might think we have oversimplified in places. Our goal is to capture the essence of the key issues without misrepresenting them.

One of the greatest concerns about the upcoming AI revolution is its impact on jobs. This is an enormously important subject, but it is not directly related to the aims of this report, and therefore we do not address it.

Finally, we do not take as given that AI will fundamentally upend the whole financial services industry, much less do so all at once. For all the recent attention it has received, the foundational theory behind current applications of AI has been well-known since the 1950s, and ML has been employed in financial services to some degree since the 1990s. But, as we describe in Section 1, we do now appear to have reached a tipping point where what was once considered peripheral is becoming mainstream.

"This report was informed by dozens of conversations the authors have had over the past few months with AI specialists - including financial practitioners and risk managers, data scientists, technology providers, consultants, regulators, and academics. Interviews were conducted on the condition of anonymity, but a number of the people we spoke to subsequently agreed to be quoted by name. Our thanks to everybody who contributed."
Keyur Patel and Marshall Lincoln

Summary

This report is structured as follows:

<u>Section 1, the *Primer*</u>, looks at what AI means in practical terms for the financial services industry, at the new capabilities it enables, and at common misconceptions about the technology. It also introduces some common applications.

<u>Section 2, *Weighing the Benefits*</u>, gives an overview of the potential benefits AI could bring to financial services – including facilitating the 'democratisation' of the industry, and offering major improvements in security, compliance and risk management.

<u>Section 3, the *Risks*</u>, is the focus of this report, exploring the risks that might arise from the increasing embrace of AI in financial services. It introduces a framework of three 'risk drivers' and 12 'key risks' in three categories, as follows:

Risk drivers

- **Opacity and complexity:** A trade-off at the heart of many AI models is that, generally speaking, the more effective the algorithms, the more difficult they are to scrutinise.

- **Distancing of humans from decision making**: AI is different from previous forms of automation because it enables many actions to be taken without explicit instructions from humans.

- **Changing incentive structures:** The benefits to successful actors and the risks of getting left behind create powerful incentives for firms to collect data and implement AI solutions on a rapidly accelerated timeline.

Key risks: New ethical challenges

- **Data acquisition and aggregation:** Using AI creates strong incentives for financial institutions to collect, aggregate and centralise data, increasing concerns about data security and privacy.

- **Perverse behaviour of AI models:** AI models can lead to the propagation of biases that can be very difficult to identify and root out. They can also perform poorly in previously unencountered situations.

- **Insufficient transparency:** The difficulty of understanding and explaining decisions made or augmented by AI could damage trust in financial services.

- **Optimisation at the expense of social benefits:** AI enables institutions to evaluate risks at a much more granular level, which could disadvantage certain customers and challenge conceptions of fairness.

Key risks: Skills gap

- **Talent gap:** There is an acute shortage of specialists who can design, develop, deploy, test and maintain AI systems – particularly of those who have knowledge of financial services.

- **Knowledge gap and unrealistic expectations:** AI systems could fail spectacularly if decision-makers who don't understand the technologies do not set appropriate expectations or provide AI teams with the right resources.

- **Over-reliance on AI:** Resources could be wasted on AI if it is implemented 'for its own sake', or if the people reliant upon it are unable to interpret or work with their outputs effectively.

- **Inadequate strategic alignment and governance:** Institutions that implement AI projects without restructuring their organisational hierarchy to reflect the new technologies expose themselves to risks from poor management and leadership.

Key risks: Market dynamics

- **Market concentration:** While AI has spurred competition, it may also lead to more market concentration and the erection of barriers to entry

since its 'winners' benefit from economies of scale and powerful new network effects.

- **Increased interconnectedness:** Use of AI might create new kinds of interconnectedness in financial markets – at the data level, the IT systems level, and the decision-making level.

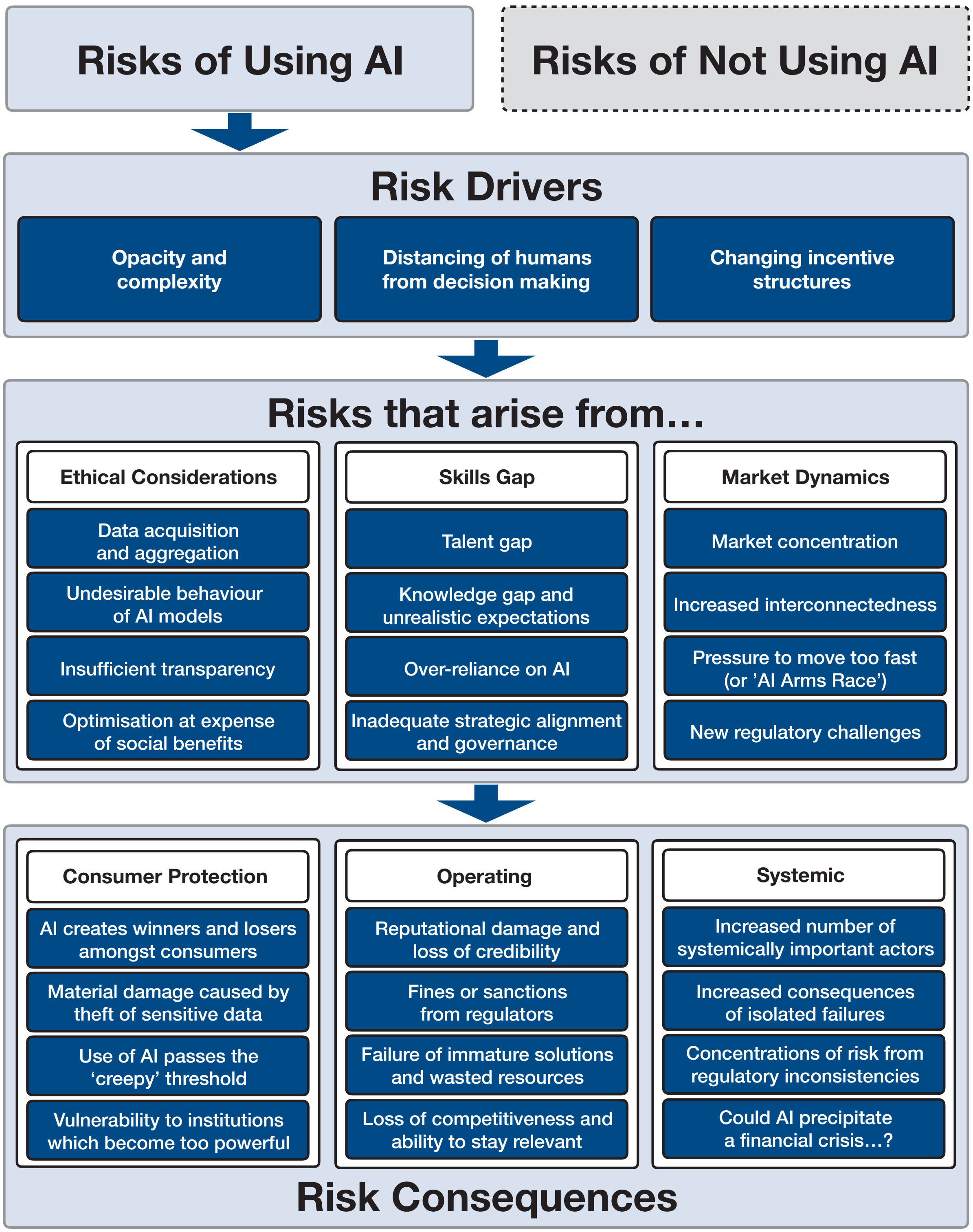
- **Pressure to move too fast:** Pressure to deploy AI solutions quickly to remain competitive may lead to risks, including insufficient testing and an overreliance on AI specialists.

- **New regulatory challenges:** AI poses new challenges for regulators and policymakers because of its technical complexity, the ethical questions it raises and its potential to fundamentally transform market structures.

Section 4, *Concluding Thoughts*, examines these risks through the lens of their potential consequences – to consumers, to institutions and to the financial system as a whole.

The risks we look at in this report are outlined in Chart 1. As we discuss in the section on benefits, risks also arise from *not* using AI. They are not part of this framework.

Chart 1: Risks to the financial services industry arising from AI



This chart does not address the impact that AI is likely to have on jobs and employment.

Chapter 1: Primer

What 'artificial intelligence' means in financial services – and what it doesn't

Pinning down what precisely the term 'artificial intelligence' means is a tricky business, fraught with often vigorous disagreement.[1] There are uncontroversial definitions, but they tend to be vague – such as "the capability of a machine to imitate intelligent human behaviour". What it might mean for a machine to 'think' is a philosophical question, and one we do not attempt to broach in this report.

Our focus here is on what AI means in practical terms for the financial services industry – as it is applied today, and as it is likely to be applied over the next few years. Many parts of the industry have been automated for decades in ways that were, at the time, considered 'intelligent'. For example, 'expert systems' – where experts in a field would codify their knowledge into 'if/then' rules – were applied for personal financial planning and market trading in the 1980s. We can now go beyond that.

The pertinent question today is whether a set of technologies currently on the brink of widespread adoption offer fundamentally different capabilities than those offered by already widely implemented forms of automation. We believe the evidence strongly suggests that they do.

Many in the financial services industry prefer to use terms like 'augmented' or 'assisted' intelligence to describe these technologies, since to do so emphasises that humans remain at the helm. However, because it has become commonplace, we use the broader term 'artificial intelligence'.

It's really machine learning...

When financial institutions say they are using AI, they are typically referring to automation based on machine learning (ML).

Most previous forms of automation have been rule-based – i.e. they involved a human programmer knowing exactly what needs to be accomplished, who specifically defined what the machine would do at each step of the process. A useful way to think about ML is that it allows us to "automate the ineffable"[2] – to enable a computer to perform certain tasks when we are unsure of how to convey explicit instructions. The idea is that, if fed enough data relevant to a given task, the computer can 'learn' from examples and infer instructions by itself. Then, in the course of performing the task, it will continue to learn from the effectiveness of its actions and improve these instructions in real time.

In order to do this, the computer needs to be provided with an *algorithm* – a sequence of actions that helps it adapt to the data to solve a given problem. There are Different types of algorithm are useful for different purposes. Many are available in public libraries: they can be picked 'off the shelf', and repurposed as necessary. An ML algorithm is 'trained' on data to build a *model*. The model can be thought of as a blueprint for the instructions the computer can infer to optimally perform a given task.

There is nothing 'magical' about this process. At its core, it is about finding patterns in data. It is based on well-known statistical techniques, which range from relatively simple (e.g. linear regression) to much more complex (e.g. multilevel neural networks). ML is generally more suited to finding correlations than establishing causality. But the correlations it does find can produce very powerful insights and accurate predictions.

1. Not least because one popular conception of AI seems to be: "whatever a computer cannot yet do".
2. A phrase attributed to Cassie Kozyrkov, Chief Decision Intelligence Engineer at Google.

Chart 2: The increasing technical complexity of automation

Type of Automation	Executed by Humans	Basic Automation	Classical Statistics	Machine Learning
Description	Humans performing all tasks (including through manual, direct control of computer programs). Highly dynamic decison-making, but naturally limited in speed and amount of data that can be used to inform decisions.	Rule-based programming ('If-Then' commands) and automation of repetitive tasks. Fully deterministic. All possible actions the machine may take have been prescripted, in response to a known range of possible inputs.	Methodologies used to predict relationships, segment objects, and otherwise draw inferences about data, in order to inform rule-based programs around which action(s) to take.	Systems centred on the ability for machines to learn, make decisions, and continuously improve the quality of decision-making, without pre-programmed, deterministic rules.
Technical Complexity				

The outputs of ML depend on the task at hand. They include:

- Prediction: identifying something as likely before the event.
- Classification: assigning predefined labels to data.
- Categorisation: splitting objects into similar groups based on their characteristics.

When we talk about artificial intelligence (AI) in this report, we are referring to computer programs that are able to automatically take some action on the basis of the outputs of ML.

We use both terms here. Given their inextricability and the lack of agreed definitions, the easiest way to read the report is to regard the two as roughly synonymous (but to know that ML is typically the core around which AI applications are built).

To use an analogy: ML[3] is the engine that powers AI. Data is the fuel for any ML model. This report looks at the risks to the financial services industry that arise as a result of collecting and processing the fuel, running the engine, and relying on its outputs.

We've surpassed critical thresholds

While there has been a surge of excitement around AI and ML in the past few years, much of the theory behind what is being applied today was formulated in the mid-twentieth century. ML has been employed for a limited range of uses in the financial services industry for more than two decades. Basic ML techniques in fraud detection, for example, can be traced at least as far back as the early 1990s.

But, for ML to be applied on a wider scale, there are two key requirements:

- The first is enough computing power. ML methodologies have intensive computer memory and storage demands, which made them impractical for most uses until quite recently. Since the 1970s, *Moore's law* has broadly held true: the principle that overall computer processing power doubles around every two years. Generally speaking, the critical thresholds in hardware availability have now been crossed. Advances in cloud computing are also enabling financial institutions to rent the hardware they need from third parties rather than having to invest in their own servers – making ML capabilities much more widely accessible.

3. (and related statistical techniques)

- The second requirement is enough data. So frequently is data proclaimed "the new oil" that the metaphor is sometimes seen as trite. But there is another reason the comparison is useful. Data is the 'raw resource' of ML, like oil to an internal combustion engine. When collected, aggregated, processed, refined, and made available in the right quantities and in the right ways, it becomes the fuel that powers ML. But in order for ML to produce insights that are useful, it requires a lot of data. To give one example of scale: last year the social media platform Facebook used 3.5 billion public photos to train an algorithm to categorise images for itself.

The collective sum of the world's data is predicted to grow from 33 trillion gigabytes in 2018 to 175 trillion by 2025, according to *IDC* [iii]. Much of this will be produced by internet-connected devices with embedded sensors (IoT devices) – more than 75 billion of them worldwide by 2025, according to some forecasts.

… leading to new capabilities

The upshot is that while AI and ML have not suddenly emerged from nowhere, they can now be economically used for a much wider range of applications. The expansion of computing power and data availability, combined with the development of new ML algorithms, has led to promising new opportunities in financial services.

One of the most significant comes from improvements in natural language processing (NLP). NLP enables computers to analyse and 'understand' human language. Most of the data the world generates is not structured, and therefore cannot normally be input into algorithms in its raw form. This includes freeform text in emails and websites, news stories and social media, research and reports, and so on.

Rule-based tools to equip computers with an understanding of human language have been around for decades. But they are useful in only a narrow range of cases, due to the complexities of natural language and the practical difficulties of explicitly defining rules for every contingency. Modern NLP techniques both use and enable ML capabilities. ML has underpinned advances in NLP, which have in turn greatly increased our ability to organise text data into inputs for other ML models. In other words, computers can now be unleashed on an ocean of previously unusable data.

ML's pattern detection capabilities are also being widely applied to image recognition, in fields such as computer vision. Uses range from extracting text from scanned documents (to which NLP can then be applied) to automatically identifying objects, locations and people in photographs and videos. AI-powered facial recognition technologies are becoming increasingly effective and can be used in financial services, for example, for instant identity verification. But they are also a magnet for controversy because of their potential for abuse.

The most exciting advances in AI in recent years have come from a subset of ML called deep learning – which makes use of very complex algorithms with multiple layers, typically inspired by biology. While there are a growing number of deep learning applications in financial services, simpler ML algorithms are sufficient for many purposes. This report uses ML as an umbrella term which includes deep learning.

A few examples in financial services

'Supervised' and 'unsupervised' ML for fraud detection. One of the earliest applications of ML in financial services was to help combat identity fraud (for example, in credit card transactions) – deployed alongside, and now increasingly replacing, rule-based statistical models.

The most widely applied ML methodologies in this field are 'supervised' – where the algorithm has been trained on many past transactions which are labelled as fraudulent or not. The characteristics of new transactions are assessed in real time to determine which group they probabilistically belong to.

A growing concern is that criminals are devising innovative techniques to evade screening. Financial institutions are, therefore, experimenting with ML algorithms based on 'unsupervised learning', which enables models to discover patterns they have not been explicitly primed to look for. The basic idea is to cluster together unlabelled data points with shared characteristics that aren't known in advance, to identify suspicious connections.

Implications: financial needs and the life cycle

"Financial crime is a really good example of where we have to become more efficient. People are transacting in different ways, more digitally, more internationally, with a much higher frequency of micro-payments. Traditional rule-based models generate a high volume of false positives. Perhaps 0.05 per cent are true matches. That's a huge cost. Each alert has to be manually evaluated by someone in the operations team to determine whether it needs escalation for further investigation. For example, under a rule-based model we may have a threshold for an account dependent on average monthly spend. The model will automatically send out an alert whenever that threshold is breached, and every month that will keep happening, until the rules are recalibrated.

"Using machine learning we can improve our true match volumes and reduce false positives. An adaptive learning model can automatically identify what looks out of place for a customer with a particular profile and get down to a much more personalised level of data analysis. It can refresh the limits of previous account thresholds based on new information. This improves our ability to detect fraud and reduces operational costs. It also stores that data, meaning if there's another alert in the future, we don't have to duplicate all our efforts."

Ajwad Hashim,
Vice President, Innovation and Emerging Technology, Barclays

NLP for sentiment analysis and chatbots. By analysing natural language, computers can parse a large number of documents to gain insights into their underlying tone. 'Sentiment analysis' can, for example, be used to classify each sentence in a financial report as expressing a positive, negative or neutral opinion. This turns unstructured text into the kind of structured data necessary to feed ML models.

Advances in NLP have also made 'chatbots' increasingly viable – automated text systems which can (to some degree) take part in dialogue with customers, make sense of the context it is delivered in, and even mimic human empathy. Chatbots have become commonplace on financial institutions' websites, social media platforms, and in their contact centres.

ML applied to larger and new datasets for credit scoring. When assessing whether an individual or company is likely to repay a loan, ML enables creditors to analyse a much larger number of data points, perform more complex pattern analysis, and make decisions more quickly – sometimes instantly. Where applicants lack traditional credit histories, ML algorithms are increasingly being deployed to analyse alternative data – from utility bills to data from social media accounts – to evaluate creditworthiness.

Smart automation for personalised marketing. ML can help institutions identify to whom they should market and what services to sell them. Rather than the shotgun approach of traditional marketing, AI can automate the management of thousands of unique campaigns, personalised to each prospect.

The Internet of Things (IoT) for increasingly granular risk assessment. The ability to run data produced by IoT sensors through an ML engine has important implications for insurance pricing. One of the key opportunities here is to reduce information asymmetry. For example, sensors in cars that monitor an individual's driving style can be used to assess risk at an individual rather than a group level, and price policies accordingly. Sensors in the home can more accurately predict loss events (for example, if an electronic application shows signs of malfunctioning) – and automatically alert customers if risks they face are not covered by existing insurance policies.

What AI isn't...

Many misconceptions about AI arise either because people overestimate its capabilities (often by comparing it with human intelligence), or because they disregard AI as not fundamentally different from existing technologies.

Don't anthropomorphise: AI has no access to common sense. In the media it's common to read that AI has 'gone rogue', e.g. an automated chatbot that makes inflammatory comments, or a self-driving car that runs red lights. This can create a misleading impression.

An AI application requires very specific instructions about the parameters of the task it is charged with. It will then follow the letter of the instructions; it has no capacity to gauge their spirit. If the best path to an objective happens to be something the system's designers wanted to avoid but forgot to specify – or did not anticipate – the algorithm will find it. But it is incapable of 'bending the rules'.

The approach that ML algorithms use to achieve an objective isn't the kind of refined strategy that we intuitively associate with the human mind, but sheer brute force. The machine makes a guess, refines this guess based on instant automated feedback, and converges on to a target. While algorithms are capable of narrowing the range of remaining options with each iterative guess, the general approach is often one of trial and error. So, while ML can perform specific tasks orders of magnitudes faster than humans, in a computational sense it is very *inefficient* – which is one of the reasons it requires so much processing power.

> "Behaviours get reported as 'rogue AI' despite them being very far from that. AI technologies behave exactly as they're told to; but without the common sense and contextual information a human would see as obvious. This can make it seem as though the AI is doing something irrational from a human point of view."
>
> Senior AI consultant in financial services

But don't miss the point: AI is not just faster computing. Traditional rule-based automation can be complex, multivariate, and result in a broad set of possible outcomes that references a broad set of inputs. However, it is deterministic. A human has to think through every step, and define in advance what action the machine should take.

ML is fundamentally different. It enables programmers to approach problems for which they do not know the optimal solution, and which they do not need to discover for themselves.

This is important for at least three reasons:

First, it's impossible for human programmers to write deterministic programs that consider thousands (or millions) of data points – resulting in billions of potential combinations – and tell a machine what to do in every one of those situations. When a machine is able to infer what to do, it can make use of much more data and make better predictions.

Second, rule-based programs are static. If fed more accurate data they will generate better outcomes, but the process remains the same until it is modified by a human. ML algorithms automatically evaluate the outcome of their decisions in real-time. The greater the quantity of relevant data the algorithm is fed, the better the process – and not just the outcome – gets.

Third, ML can detect non-linear correlations. This can be, for example, the difference between an approach that assigns a given weight to each variable to make a prediction, versus a non-linear approach that is capable of considering concepts such as 'and', 'or', and 'not'. In very large data sets that model the world in much more detail than was previously possible, many correlations cannot be discovered by linear analysis. ML allows us to uncover these relationships and use them to make predictions.

More complex algorithms, such as multi-layer neural networks, have been referred to as 'universal approximators'. This means that a well-designed neural network should (in theory) be able to detect and approximate all possible relationships within the data, and represent those in its model of a system to generate its outputs.

On the next page, we present a table of applications for which AI is currently being applied most widely in the financial services industry, and areas in which it will likely be applied over the next few years. Section 2 explores some of the key benefits to the industry that these applications can bring about.

Table 1: Selected AI Use Cases

Use Case	Potential Applications
Security and Compliance	
Regtech	Automation and assistance of regulatory reporting, compliance, and monitoring (e.g. for AML and KYC).
Fraud detection	Identification and blocking of fraudulent transactions in real time (e.g. credit card purchases) – with greater precision (fewer false positives) and recall (fewer false negatives).
Cyber security	Models that identify malicious actors and enhance network security with more effective intruder detection.
Identity verification	Use of computer vision technology to enable more accurate – and often instant – identity verification, reducing costs for institutions and reducing friction for customers.
Capital optimisation	Optimisation of risk-weighted assets to improve margins in line with regulatory requirements.
Operational Efficiencies	
Process automation	Automation of repetitive tasks to increase efficiency and speed, reduce error rates, and cut labour costs.
Trading efficiencies	Reduction in operational and transaction costs associated with trading activities.
Insurance claims validation	Claims-handling with ML-based fraud-detection, triage, and verification of damages (e.g. using image recognition).
Hiring efficiencies	Automation of processes to improve screening, and cut process time and drop-off rates. Reduction of bias by hiding personal information from human reviewers.
Marketing, Sales & Retention	
Web and mobile optimisation	Enablement of relevant and personalised site experiences to improve click-through and conversion ratios on web and mobile.
Personalised marketing	Personalisation and automated messaging drives better conversion rates, increases average revenue per customer, and reduces cost of acquisition.
Automated buying experience	Reduced friction in purchasing and reduced sales costs.
Customer support	Use of chatbots, analytics, and other tools to increase efficiencies, deflect interactions, and identify primary case drivers.

Table 1: Selected AI Use Cases (Continued)

Risk Pricing	
Credit scoring	More accurate risk assessment, and improved speed and efficiency.
Insurance underwriting	More granular and personalised assessment of risks, including through the use of IoT sensors.
Price optimisation & bundling of services	Better understanding of customer purchasing patterns, which allows institutions to better address markets by refining, bundling and effectively pricing their services.
Use of proxy and non-traditional data	Creation of larger markets of potential customers through the use of non-traditional sources of data and AI in order to more accurately price risk.
Product & Strategy	
Financial advice	Democratisation of value-added services which would otherwise be inaccessible to segments of the population.
Predictive advisory services	Provision of real-time information to customers to anticipate and prevent risks. New, value-added services to customers.
New investment strategies and services	AI-driven investment services (e.g. automated portfolio management) which can be offered to customers at a fraction of the cost.
Business Intelligence & decision support	Improved understanding of an organisation's operations, revenue streams, and customer base. Deeper insights into the driving forces behind metrics.
AI trading	Algorithmic trading based on ML can outperform rules-based approaches. Differentiated AI-driven investment strategies, including use of non-traditional data.

Chapter 2: Weighing the benefits

"There is potential for more and better data, combined with AI, to transform finance for the better – for consumers, for risk managers, for financial inclusion, and many other goals. Used alongside cell phones and new distribution systems, we can democratise finance, make it accessible, make it affordable, make it fair. It's like no technology in the financial industry that has ever existed before… it can be the most democratising force."

Jo Ann Barefoot, CEO and Founder of the Alliance for Innovative Regulation (AIR)

Though this report primarily focuses on risks, the potential benefits to financial services of using AI and ML effectively are, in our view, compelling.

AI can be used to improve the efficacy and efficiency of existing activities in the industry, often in ways which are powerful enough to transform business models. It can also make entirely new activities possible, and create new markets. Applications are being developed and used in virtually every part of the industry. Many are customer or client facing, including in retail and corporate banking, insurance, wealth management and trading. Others are designed to improve financial institutions' back office performance and risk management.

Broadly speaking, the benefits to the industry of using AI fall into one or more of six categories.

First, institutions are using AI to <u>increase the speed and reduce the cost of their operations</u>[4] – for example, by streamlining processes, automating regulatory compliance, and analysing incoming data in real time.

A second is to <u>curtail or eliminate human error</u>, while freeing human talent to pursue more creative or higher-value work. Among the most prosaic AI applications are those that deal with the automation of repetitive tasks, such as reviewing large numbers of text-based legal documents.

A third is to <u>offer 'always-on' services through a variety of digital channels</u>. For example, AI-powered chatbots can respond to customer enquiries 24-7. A smartphone with a camera can be used alongside ML to facilitate a wide range of activities – from verifying a customer's identity to offering insurance quotes on a product that has been photographed by the applicant.

A fourth is to offer <u>customised products and advice</u>, tailored to the unique characteristics of customers or clients. For example, data can be pulled together from across a user's financial life and run through an ML engine to personalise investment decisions. Data from IoT devices can be used to assess insurance risk at an individual level, and policies can be priced flexibly (e.g. pay-per-mile car insurance).

A fifth is to make <u>more accurate predictions</u>, enabling better-informed decision-making. This is one of the core capabilities of ML when it is 'let loose' on large datasets to discover new correlations. This can be used, for example, to more accurately assess creditworthiness, underwrite insurance policies, detect fraud and develop investment strategies. These decisions can increasingly be made on dynamic datasets.

A sixth is to <u>draw insights from new *types* of data</u>, which may enable financial institutions to reach new markets. This can include, for example, satellite images to assess insurance claims in rural areas where verification is otherwise costly, and psychometric tests that attempt to gauge character traits to assess the creditworthiness of thin-file loan applicants.

4. Of course, one of the main ways AI is likely to reduce costs to the industry is by substantially cutting headcounts. Whether this is a benefit is a matter of perspective…

The risks of *not* using AI

An argument we came across frequently was that the financial services industry needs to embrace AI not only because of its potential benefits, but because it is the only way to manage evolving risks. As the accessibility of these technologies grows, so does the potency of bad actors within the financial system that use them.

Lael Brainard, a member of the U.S. Federal Reserve's Board of Governors, said in a speech last year [iv]:

> *"The wide availability of AI's building blocks means that phishers and fraudsters have access to best-in-class technologies to build AI tools that are powerful and adaptable… Supervised institutions will likely need tools that are just as powerful and adaptable as the threats that they are designed to face, which likely entails some degree of opacity… Where large data sets and AI tools may be used for malevolent purposes, it may be that AI is the best tool to fight AI."*

There is a wider argument here: that powerful new technologies should not just work in favour of the most well off by, for example, developing profitable investment strategies for hedge funds that serve the super-rich. In other words…

Can AI democratise financial services?

There is much excitement around this question because AI has at least two characteristics with wide-reaching implications:

The first is that it provides new ways to evaluate risk — for instance, making it economically viable for institutions to offer financial services to individuals or small businesses without traditional credit histories. Billions of people worldwide have limited or no access to banking, savings and insurance products. Most reside in markets where financial participation is low. Many others are locked out of financial services in mature economics.

AI has many potential applications here. For example, insurers can apply ML to unstructured or semi-structured data to underwrite policies that would previously have been unviable – not because the risks were too high, but because there was no way of pricing them. AI can also be used to mitigate risks – for example, by using data from sensors and satellite images to monitor agricultural produce and flag early warning signs (such as crop damage or drought threats).

When it comes to fraud detection and anti-money-laundering (AML) activities, AI can be used not only to uncover criminal activity, but also to dramatically reduce false positives.

A second democratising capability is that AI enables institutions to provide low-cost alternatives to financial services which have traditionally incurred high fees. The automated alternative may lack the human touch of the original, but it is often good enough for its purpose (And, as technology matures, potentially even better).

Notable examples are financial advice and investment management, which are being opened up to customers with relatively low balances, or who are unwilling to pay fees. Other potential uses include, for example, triaging payments to avoid late fees, optimising saving rates, highlighting inefficient financial behaviour, refinancing and consolidating credit, and switching providers for commoditised products such as utilities.

> *"At the moment, with screening requirements like KYC, we create huge barriers that stop people getting into the financial system, and then try to police the transactions that make it through. But you could imagine an alternative with much lower barriers – because we really want to have all of the transactions in the system – and AI could be used to do the monitoring and spot people up to no good. But this is only possible if the people in charge of policing have access to AI."*
>
> Dave Birch, author, advisor and commentator on digital financial services

But the argument that AI will decrease inequality is a double-edged sword. Just like any other technology, AI is a tool – its uses are determined by those in a position to wield it. There are widespread concerns that AI will be used in ways that exacerbate inequality: between individuals, companies, and countries. The collection and aggregation of data on an unprecedented scale can provide valuable services at a low cost, but can also concentrate power in the hands of those with access to data. AI can be used both to augment human work and to replace human workers cheaply. It has enormous implications for income distribution and employment. These raise fundamental questions about how policy and law should be formulated at a national and international level.

Compliance and risk management are among the most valuable uses of AI for institutions

While customer-facing AI applications such as chatbots receive most publicity, financial institutions often see the most value arising away from the public eye, in the back office.

One opportunity is to reduce compliance costs that have mounted since the financial crisis, as institutions grapple with the growing volume and prescriptiveness of regulatory rules.

ML is increasingly being applied in the field of regulatory technology (regtech) to, for example:

- Automate compliance with Anti Money Laundering (AML) and Know Your Customer (KYC) requirements.
- Track regulatory changes from multiple authorities and automatically identify those that are relevant.
- Screen all emails by employees before they are sent, to comply with data regulations.

In a recent global survey of financial decision-makers by IBM, around two-thirds of respondents said they were using AI to help meet compliance requirements. The main benefits cited included cost and time savings, more accurate analytics, and the ability to reveal previously unavailable information.

Some regulators are experimenting with machine-readable handbooks, which would allow computers to interpret and implement rules largely without human intervention. For example, the UK's FCA recently developed a proof of concept which supported the idea that regulatory reporting requirements could be converted into code and executed by machines, and is currently in a pilot phase to assess the economic viability of rolling this out widely.

Institutions are also exploring use of AI to optimise how they allocate capital, in line with tightening regulatory requirements. This is an area that already relies heavily on mathematical approaches and modelling, and is likely to benefit from ML's superior pattern detection capabilities and ability to process larger and broader datasets. Other possibilities include:

- Use of alternative datasets to improve the coverage of risk models.
- Better pre-trade risk analysis – including modelling the effect of an institution's own trading on market prices – and real-time tracking of risk exposure.
- More accurate stress testing, which can account for a wider range of potential scenarios.

How widely is AI being used in financial services?

There's a lot of variation by location, sector and application – as well as in how institutions self-report AI usage. That said, survey evidence suggests that ML is already widely used in financial services, and that decision makers plan to implement much more of it over the next few years.

Some recent global surveys, by:

IDC[iv] **(2019):** Worldwide cross-industry spending on AI systems is forecast to reach $35.8 billion in 2019, an increase of 44 per cent over the amount spent in 2018. Banking is the second largest industry by spend (after retail), spending $5.6 billion.

Financial Times[vii] **(2018):** Of 18 major banks surveyed, 17 reported using AI in the front office. Eight reported AI in the front office, middle office, back office and data analytics. Of six which gave details of AI spending, the sums ranged from €5m to €15m, with one institution planning to increase spending from below $3m to $50m a year. Eight were involved in joint ventures, while four had made investments in AI-related companies

Narrative Science and the National Business Research Institute[viii] **(2019):** 32 per cent of financial services executives said they were using AI technologies such as predictive analytics, recommendation engines, and voice recognition.

SAS and the Global Association of Risk Professionals (2019): 81 per cent of risk professionals in the financial services industry reported having seen benefits from AI technologies, including improved process automation (52 per cent), credit scoring (45 per cent) and data preparation (43 per cent)[ix].

Deloitte (2018)[x]**:** 25 per cent of financial institutions said they already used ML in their risk functions, and a further 47 per cent were planning to use it

Intertrust (2018)[xi]**:** 77 per cent of respondents said that AI would play a bigger role in revolutionising the financial services industry over the next five years than other disruptive technologies - ahead of blockchain (56 per cent) and robotics (27 per cent).

… but the hype factor remains

It is clear that the use of AI in financial services is no passing fad. However, the *Financial Times*, reporting last year on the survey listed in the box, cautioned: "Not only is there little consensus on how AI should be used in banking, [but] many of the current efforts to apply machine learning are relatively modest. Rather than racing towards an AI-enabled future, the industry is feeling its way forward."

One comment we heard widely was that many of the companies that claim to use AI may only be using it at the periphery of their services – if at all – for marketing purposes. A department head at one large financial institution told us: "So many fintech start-ups are marketing themselves as using AI, but when you really dig into what they're doing, typically they are not AI businesses. Institutions are dressing themselves up to be something they're not – and also looking for AI solutions when actually perhaps simple rule-based systems or other types of algorithmic approaches could be applied to those problems, and be as or more successful".

The AI ecosystem

There is a mix between financial institutions building AI solutions in-house and procuring services from third parties. One fairly widespread approach is to purchase cost-saving technologies – such as chatbots and financial crime modelling – as commodities, under a licensing model. However, where institutions see AI offering a competitive advantage – for revenue generating technologies, such as credit modelling or sales lead identification – they often prefer to retain more control and develop the technology in-house.

The third parties active in this space can be divided into three broad groups: providers of AI applications, the cloud infrastructure upon which it runs, and providers of the third-party data that feeds the algorithms. Many of the fintech start-ups which are both competing with incumbents and providing them with services are small. Conversely, providers of cloud and technology infrastructure are often among the world's largest and most well-resourced tech companies. They are increasingly exploring ways to bypass financial institutions altogether, and offer services directly to customers.

How this will all play out is a much-debated question. In the CSFI's 2019 *Insurance Banana Skins* survey, which identifies the main risks facing insurers over the next 2-3 years, we received many comments along the lines of the following, from a life insurance respondent in Japan: "Digitisation of distribution access is allowing powerful new entrants in… we should fear the technology giants."

AI is a global business. The countries leading the way in AI investment and deployment, by some distance, are the US (in first place) and China. Then there is a gap, with the UK probably in third place among the chasing pack.

According to the IDC survey listed above, AI spending growth in the Asia-Pacific region is expected to outpace the rest of the world over the next three years – two thirds of which will be driven by China. At least four of the world's largest fintechs – Ant Financial, Qudian, Lufax, and ZhongAn – are Chinese businesses.

Many concerns have been raised about differing attitudes towards data privacy and surveillance in China, especially with the impending introduction of the government's Social Credit System in 2020. It's still unclear what this system might entail, or the extent of its reach. In the West, fears are hardly less pronounced about the rise of 'surveillance capitalism'. These will be pervasive themes in future discussions about AI.

A word on blockchain and quantum computing. AI is not the only technology expected to revolutionise financial services over the next few years. Much has been said, for example, about the industry's adoption of blockchain and distributed ledger technologies.

In a nutshell, these involve the use of cryptographic methods to securely store records that anyone can verify – not just a central authority. This could have profound implications, for example, for the payments industry and identity management. There are a number of ways in which blockchain and AI might complement one another. For example, blockchain might be used to facilitate micro-transactions (such as the sale of data from or across IoT devices), with ML applied to analyse the resulting datasets.

Though it is still largely at a conceptual stage, quantum computing could also have massive ramifications for how we use technology.

The ML algorithms at the heart of AI applications rely on the ability to make inconceivable numbers of computations every second. Computer hardware has caught up with these requirements – but any mature quantum computer would be orders of magnitude more powerful still. Much about this field is speculative, but there are predictions that ML techniques which take days to run with traditional computers could be performed virtually instantaneously using quantum computing hardware. This could bring about a step change in what AI applications can accomplish – in terms of scale, power and complexity.

End-to-end AI business models in the insurance industry

Applications of AI in financial services are often not clear-cut. Institutions regularly use ML alongside rule-based systems, or are in the process of transitioning from one to the other. AI currently tends to be used for relatively narrow tasks. This is expected to change. Using technologies that are already viable, it is plausible to envision large financial services providers which have AI applications at the centre of their business models.

Consider, for example, this wide range of applications for AI in the insurance industry:

~ Claims prediction ~ Risk assessment ~ Lifetime value prediction ~ Behavioural policy pricing ~ Price optimisation ~ Customer segmentation ~ Automating life event marketing ~ Recommendation engines ~ Personalised marketing ~ Automated buying experience ~ Automated claims handling ~ Fraud detection ~ Claims settlement

On the surface, these may appear to be unconnected. Taken together, however, they represent a fundamental shift in the end-to-end business.

Product design and pricing. Insurance providers could offer customers wearable IoT devices, or the option to install IoT sensors in their cars and homes, in exchange for discounted rates. ML could be applied to the ocean of data these sensors generate to predict future claims and determine risk as a function of user behaviour. This could be used to create behavioural pricing models that optimise prices charged to customers. Using ML, these prices could be weighed against the risk profile of each customer, to determine a 'lifetime expected value' of each customer or lead.

Personalised marketing and customer acquisition. Insurers already segment prospective customers by risk and reward profiles. They could use ML models to form groups of similar individuals, and then run targeted advertising campaigns against them. By pulling third-party data into their own databases of leads, they could gain visibility into the timing of major life events – such as marriage, the birth of a child, the purchase of a new house or car, or moving to a new city. This data could then be fed as inputs into recommendation engines which deliver personalised messaging, products, and pricing.

Automated sales and sign-up. The lead – now an interested prospect – would not need to interact with a human agent to purchase a policy. AI-powered chatbots are already being used by some insurers to automate the entire buying experience.

Customer support and claims handling. Once the lead becomes a customer, chatbots can be used to handle basic support functions and help customers navigate online databases. To handle phone calls, AI embedded in the Interactive Voice Response (IVR) at the heart of the company's call centre platform can be used to anticipate questions and provide answers on the fly. If the caller needs to speak with a human, the IVR can use available data on the customer, their policy, and their case history to route the right call to the right agent and provide that agent with the relevant case history.

Customers can file claims entirely online, through a browser or app. ML fraud detection models could then flag fraudulent claims without a human having to provide the first line of defence.

In short, insurance companies can use AI to design personalised services, optimise pricing around those products, and personalise both the content and timing of marketing material to generate new leads and customers. It can also be used to automate, optimise, and personalise nearly every interaction with leads, customers, and claimants.

On the relationship between humans and AI. None of this is to say that humans will not remain at the centre of the business – designing, overseeing, monitoring and working alongside AI. Roles that entail routine, highly repetitive actions may be replaced by AI. Other employees, who might have little functional understanding of the systems, will nonetheless be enabled to do their jobs better by working with AI applications.

Elsewhere, new jobs will be created. Organisations will increasingly require data scientists who design and build AI applications. They will also need specialists across the business with a functional understanding of how AI works, who can use it to design and build financial products, to acquire customers, and to inform business strategy.

Chapter 3: The risks

This section looks at the risks that arise from collecting and processing the data that fuels ML, from implementing the algorithms and from relying on its outputs. They are not specific to any sector of the financial services industry.

Our aim is to show why AI might present different risks to those associated with the technologies that have transformed financial services over many years. While many of the themes we discuss are not unique to a world where use of AI is widespread, they are heightened and manifest themselves in new ways. For example, the risk of private data being stolen or misused has been a serious concern in financial services for some time. But the ability to harness ML creates a strong incentive for institutions to collect new kinds of data, and to centralise and aggregate it – creating even greater 'prizes' for a successful cyber attack.

There are at least three relevant characteristics of AI technologies that arise repeatedly in our discussion. These help to explain why AI is different from other forms of automation, and why it may bring about new kinds of risks. We call them 'risk drivers'.

Risk drivers

Opacity and complexity

One of the trade-offs at the heart of ML is that (generally speaking) the more powerful the methodology, the harder it is to scrutinise. This is because:

- Many algorithms are set up to find complex, non-linear correlations between an immense number of disparate data points – correlations which may be invisible to human eyes.

- There tends to be an extremely low signal-to-noise ratio in data. In other words, while ML is capable of mining very large volumes of data to find obscure connections and predictive insights, only a small amount of that data may actually be useful for a prediction. The relationship between the data and the output may only be discoverable with a great deal of effort, if at all.

- Some of the most powerful algorithms – such as *multi-layer neural networks* – are notorious for being black boxes. Roughly inspired by the structure of the human brain and nervous system, they employ 'hidden layers' and other intermediary steps as part of the learning process. Their designers understand conceptually how they work, and can measure how accurate the outputs are, but may have little conception of what kind of relationships are being formed, tested or acted upon. It is, therefore, often difficult to understand the relative importance or specific role of a variable in determining the final output of the model.

Opacity and complexity can be a concern even when the ML technique being used is fairly straightforward. Some algorithms, such as *decision trees* (which have a flowchart-like structure), are conceptually quite transparent, insofar as the algorithm's reasoning can be traced along its branches. But applied to huge datasets, a decision tree might formulate millions of independent rules – rendering it practically (if not theoretically) opaque.

Distancing of humans from decision making

By definition, any kind of automation – AI or not – involves machines performing tasks previously carried out by humans. But there are important differences.

Rule-based automation is fundamentally predictable. For repetitive processes (e.g. most manufacturing automation), the task is always performed in the same way, with predetermined inputs and steps. True, more complicated forms of rule-based automation models can be fed a dynamic range of inputs, and given evaluation criteria. But the machine uses these criteria to determine which of a predefined selection of responses to execute. The outputs then fall into a predictable range.

With AI automation, a human gives the machine an objective (optimisation criteria) and an algorithm to guide its approach. The machine is then instructed to learn for itself how best to accomplish the objective. In other words, humans provide a goal and an engine, but are largely removed from how the machine gets to its destination.

The key distinction is between when a decision is *made* and when it is *implemented*. In rule-based automation, all expected scenarios have been thought through and all possible actions mapped out. The decision was made ahead of time by a human, and the relevant action simply taken in real time by a machine. With AI, humans provide the machine with an objective ahead of time – but decisions are evaluated and made in real time by the machine in an attempt to achieve that objective.

An important point, emphasised by many in the financial services industry, is that AI is generally used to support or augment decisions made by humans, not to replace them. It is critical, for example, that people relying on AI employ a ‘common sense check’. But risks can arise from humans not knowing how and when to rely on machines.

Moreover, even if the final or overarching decision remains in the hands of a human, there may be thousands of automated evaluations which inform the options that a human decision-maker chooses from. Computers can perform these evaluations much faster than humans. It is impossible (and self-defeating) to try to manually check them all.

It is also likely that as AI technology matures and its use becomes widespread, financial institutions will become more comfortable entrusting machines with broader business decision-making responsibilities.

Changing incentive structures

The benefits to successful actors and the risks of getting left behind create powerful incentives for firms to collect data – and to implement AI solutions that use that data – on an accelerated timeline.

Scarcely a day goes by without some new proclamation about how AI will fundamentally change the way we live and work. National governments are racing to lead innovation. ‘AI-first’ increasingly seems to be the default strategy in large swathes of the technology industry: last year, Sundar Pichai, chief executive of Google, even suggested that AI could have “more profound” implications for humanity than electricity or fire.

The financial services industry has not been at the forefront of AI deployment, largely due to regulatory factors. But there is a growing feeling that it cannot afford to miss out. One recent study estimated that the aggregate potential cost savings for banks from AI applications could be around $450 billion by 2023[xii]. We heard accounts of ML-driven fintechs being able to provide the same products 90 per cent more cheaply than traditional institutions, and still be more profitable (though we heard plenty of accounts, too, of wildly inflated claims).

The CSFI’s latest *Banana Skins* surveys of the greatest risks facing the insurance and financial inclusion industries both had ‘technology risk’ at No. 1. The dominant theme was that institutions that fail to adapt to the pace of technological change face losing relevance or worse. Many respondents explicitly referred to AI.

On the other hand, the rewards of getting AI right are huge, including first-mover advantage in vast new markets (and maintaining relevance with younger generations that are more likely to prioritise speed and cost over high-touch services). The emergence of large, dominant players in AI-driven technology markets suggests the degree of change that is possible – due not just to economies of scale, but to new kinds of network effects.

This all means that, in spite of regulatory constraints, there are mounting incentives for financial institutions to adopt AI quickly. The risk is that they feel compelled to move faster than they otherwise would – at the expense of proper controls and safeguards – which could lead to wasted resources, technological failures, and regulatory breaches.

We look at three categories of risks that AI and ML can pose to financial services.

- The first is the ethical challenges that can arise – including questions around data security and privacy; potential biases in ML models and other undesired behaviour; the extent to which a lack of transparency can be tolerated; and whether efficiency improvements might come at the expense of socially optimal outcomes.
- The second is risks around the implementation of these technologies – including a shortage of human talent and insufficient knowledge about how to use AI effectively.
- The third looks at how widespread adoption of AI might affect market dynamics – its impact on market concentration; whether it might increase interconnectedness between financial institutions, etc.

New ethical challenges

Challenges around data privacy and security

There aren't hard rules on how much data ML models require to be effective – but generally speaking, it's a whole lot more than was used before.

The upshot is that where once data was mainly viewed as a by-product of doing business, it's now increasingly seen as an invaluable *strategic tool*. This means that financial institutions are investing huge amounts of money to collect data they previously weren't collecting, to coalesce it into massive, centralised data stores, and to put it through increasingly sophisticated processes to prepare it for ML methodologies.

> "The problem is that as we bring all this data together, we need hyper-security. Banks are centralising a lot of data about their customers, which creates an enormous target. It's not just a bit of your identity, but everything we know about you that can disappear into the criminal underworld – and when it's gone, it's gone. I don't think the industry has cracked that problem and produced a really safe environment in which we can store identity" – Former executive, international bank

As well as acquiring more data internally, firms are increasingly purchasing it through third party vendors. One recent study found that the global financial services industry spent $9bn on Big Data vendors last year [xiii].

The risks include cyber attacks, data theft, and identity fraud. Cyber crimes has emerged as one of the most serious threats to the financial services industry in recent years. Institutions of all sizes face a barrage of daily attacks – from lone opportunists, criminal gangs and state actors. Compounding the problem, many institutions are encumbered with legacy IT systems that have been patched together over decades. Widespread use of AI could further amplify the risk, because:

- Previously, hackers might have had to break into multiple databases, often at different institutions, to access the quantity of data now increasingly stored in a single place.
- Aggregating and processing data doesn't always 'move' the data, but replicates it. So, too, does sharing it with third parties. This creates more targets – hackers only have to find the weakest link.
- The prize for successful data theft is more valuable when the attacker knows that data which may have been less useful in its unstructured form has been cleaned and mapped together.

Privacy issues arising from data sharing. Another concern is that customers who agreed to share their data with one entity, or for a specific use, may find it shared with other entities they were not aware they had consented to. We heard accounts of legal teams at financial institutions scrambling to update their Terms of Service to reflect expanded data-sharing activities. This is a fast-moving area, and mistakes are often unwitting. A risk is that financial institutions without clear and well-managed data policies could find themselves inadvertently violating their own customer contracts.

The pace of change is creating grey areas. In many cases, data collection technologies are moving faster than regulation can keep pace. Especially in jurisdictions with weak legal and regulatory frameworks, it seems inevitable that there will be grey areas around what firms can legally collect, how they can gather (or purchase) it, how they are allowed to use it, and how much of this activity they must disclose.

Where regulations are strong, fines for breaches of data rules can be hefty. For example, under the EU's General Data Protection Regulation (GDPR), institutions can be fined up to 20 million euros, or four per cent of annual global turnover, for severe violations. A widespread concern is that some GDPR requirements aren't just costly to comply with but, given technical limitations, actually impossible. There is much uncertainty around how much leeway regulators will give in cases where firms fail to comply despite their best efforts, or how consistently the rules will be applied.

The 'creepy' factor. The collection of much more – and more invasive – data raises fundamental questions about societal norms. The issue is not just that individuals may be reluctant to share their data; they also may be unaware of how much information about them it can reveal. Multiple studies have shown, for example, that ML analysis of public social media posts can, with a high degree of accuracy, predict a host of characteristics that haven't been shared – from income levels to political views to sexuality.

> "Lots of people, when they say they want AI, mean they want magic. And they often don't like it when you explain how it works. They say: 'that's not magic!'. The risk of people thinking AI is different is if they think there's something magical about it. That's often why people run data through algorithms and seem surprised when they get bad answers."
>
> Richard Quick,
> Principal Data Scientist, Featurespace

Why might ML models misbehave?

Many of the people we spoke to stressed that while ML models can be much more powerful than rule-based alternatives, they are just as fallible in at least four ways.

First, data quality is paramount. ML can only derive insights and make predictions based upon the data it has access to. Algorithmic biases might arise:

- <u>When training the model</u>: The sample data might not be representative of the population it is supposed to reflect (especially if the sample size is not large enough). Furthermore, historical data is not necessarily a good predictor of the future. It might, for instance, omit important contextual factors. ML's capacity to detect patterns in the data does not necessarily include the ability to determine causation – which can lead to biased outputs if the inputs to a model were shaped by historical biases.

- For example, in one widely publicised case[xiv,] an AI recruitment algorithm by *Amazon* was scrapped after it was found to discriminate against women, including penalising graduates of all-women's colleges. The machine had taught itself that male hires were preferable, due to the historical under-representation of women in the tech industry.

- <u>Once the model is live</u>: New biases can be introduced, or existing biases reinforced by feedback from earlier decisions.

In addition, different algorithms are able to explore data in different ways, and discover different kinds of correlations. Biases can be introduced into the model if the right algorithm is not chosen for a particular application.

The UK's Centre for Data Ethics and Innovation (CDEI), a government advisory body, identified financial services as one of four key sectors (alongside crime and justice, recruitment, and local government) in which biases in algorithmic decision-making have the potential to cause serious harm. A recent report [xv] noted that:

> *"The history of finance includes profoundly discriminatory practices, for example, redlining majority ethnic minority postcodes and requiring a woman's husband's signature on a mortgage. While the most egregious of these no longer happen, these organisations still operate in a socio-economic environment where financial resources are not spread evenly between different groups. As a result these embedded historical biases may be reflected in the data held."*

Second, human biases invariably seep into AI models*…* even if the data is generally of a high quality. A purported benefit of AI is that it dispassionately draws conclusions from the data it has access to without the prejudices that humans might exhibit. The reality is much more complicated.

The teams that design ML models do not simply throw in all the data at their disposal. They first perform a process known as 'feature engineering', to identify what's likely to be relevant in the raw data for the problem at hand, and transform it into 'features' – the inputs for ML algorithms. For example, to predict creditworthiness, 'income' might be one of many thousands of possible features. 'Income-to-debt ratio' is another feature which can be engineered, by applying calculations to multiple data points.

Determining which features will be used in the model can be as much of an art as a science, because they must be chosen with some hypothesis in mind. Data scientists must make at least some subjective calls when defining the scope for a project, choosing data elements for a model, and selecting hypotheses to explore. This means that ML models are incontrovertibly affected by the beliefs, values and biases – whether conscious or unconscious – of the humans who build them.

Third, undesirable outcomes can be difficult to root out. Programmers have tools to restrain ML processes to prevent 'undesirable' outcomes. One is to set a range of boundaries on the ways data can be used. For example, the output can be required to fall within a certain numerical range. Another is to prohibit the algorithm from using certain variables in its computation.

The challenge is that this ability depends on the programmer being able to anticipate ahead of time what actions should and shouldn't be allowed, and on giving the algorithm very specific instructions around the parameters within which to accomplish its objectives. Developers must anticipate that AI may find and use any correlation in the data it is fed – unless it is explicitly instructed not to. It excels at finding precisely those kinds of complex and non-linear 'hidden' correlations it is difficult or impossible for a human to anticipate or discover. This means that the machine might pick up on some correlated variable which wasn't explicitly excluded – effectively getting to the same answer in a different way (as depicted below). The relationship with this variable could be very subtle, or even counterintuitive.

X → Y	This rule is forbidden
Z → X → Y	The AI discovers this correlation
Z → Y	And therefore incorporates this one into its knowledge of the system – which is not forbidden

So, in effect, X *plays a role in the value determined for* Y*, though that role is not explicit.*

An important point is that this is not the machine trying to 'get around' its constraints by finding loopholes. On the contrary, it is very literally following the directions it is given; it cannot interpret the spirit of these instructions.

Fourth, we don't know how ML models will perform in unencountered situations. In financial markets, the most impactful events have often been fundamentally unpredictable. ML-powered risk management systems might be more effective than traditional models during market conditions which lie within the boundaries of previously observed events. But there is a great deal of uncertainty about how they might perform in rare, extreme conditions, where past data can provide little or no guidance. Prior experiences have shown that ML can perform very poorly under these circumstances.

A data scientist at a large bank told us: "When we come up with a model, what the algorithm knows is derived from that exact set of data it's trained on. The way we measured its efficacy was also based on that sample of data. It knows nothing about any other events that may take place within the system. There is no way to measure performance based on speculative events we haven't seen before."

The risks from a lack of AI transparency

Concerns about algorithmic biases are intensified if the internal mechanism of the ML model used to make decisions cannot be explained. As AI is used for more consumer-facing applications, there are risks that customers will find it more difficult to challenge decisions that are not in their favour, and to understand the actions they need to take to be approved in the future. This is especially important when the financial product is potentially life changing, e.g., a mortgage or life insurance policy.

Existing laws provide some cover. For example, regulations in many jurisdictions stipulate at least some right to an explanation for rejected credit decisions, such as the Equal Credit Opportunity Act in the US, and the GDPR in the EU. But the latter, which was implemented in 2018, appears to deliberately leave wiggle room for the kinds of complexity arising as a result of ML decisions. Regulators are grappling with a tricky balancing act: to require sufficient transparency to protect customers without too severely muzzling the benefits of AI.

Fairness in consumer outcomes is only one issue among many. For example, interpretability is essential for regulators and senior managers to ascertain that AI trading applications aren't colluding in ways that would be illegal for human traders.

What are reasonable transparency requirements? This is a much-debated question. The interpretability of opaque models can be increased by, for example, creating second simpler models that run alongside the first, and offer some verification of its actions. This is an area where considerable effort is being expended and progress made by ML researchers.

Research last year by BaFin, Germany's Federal Financial Supervisory Authority, concluded that:

Why are deep learning models especially difficult to explain?

"With rule-based approaches, you are generally able to recognise why a given recommendation has been made. The logic is built into the system.

"However, deep learning is all about identifying characteristics within the data that we as humans would sometimes have a hard time realising or even explaining. Even where we have an intuition for why a model is using a given set of features, it would be really hard for us to describe the model in a rule-based approach. When you ask why a model arrived at its decision, it will be very hard to provide an answer. You can provide an answer with respect to what features the trained model has identified as the most important – as having the most impact on a predictive value – but may not be able to provide a clear answer as to why a decision has been made."

Yoram Zahavi,
Sr. Vice President Artificial Intelligence, Citi

> *"It is the responsibility of supervised institutions to ensure the explainability and traceability of BDAI [Big Data AI]-based decisions. At least some insight can be gained into how models work and the reasons behind decisions, even in the case of highly complex models, and there is no need to categorise models as black boxes."*

Most people we spoke to expressed a view broadly in line with this. However, a few questioned the wisdom of expending a lot of resources on providing what they said might be very rough approximations of the workings of complex models simply to tick a box.

The question of explainability – i.e. explaining why a decision was made in ordinary language – raises practical questions. Christopher Woolard, Executive Director of Strategy and Competition at the FCA, said in a recent speech[xvi]: "What level does that explainability need to be? Explainable to an informed expert, to the CEO of the firm or to the consumer themselves? When does a simple explanation reach a point of abstraction that becomes almost meaningless?"

It should also be noted that many consumer-facing, rule-based technologies that have been in use for decades are far from transparent. For example, existing credit scoring methods used by agencies are based on closely guarded proprietary data collection and weighting techniques. We frequently heard the argument made that while transparency is important, current methodologies should be the benchmark against which the opacity of AI applications are measured.

When can optimisation be a bad thing?

At its core, ML is about more effective discrimination – in the neutral rather than normative sense of the word.

Better discrimination in data (e.g. isolating the signal from the noise, and identifying variables and inputs with predictive power from those without) enables *better discrimination in action* (e.g. identifying who is creditworthy, or personalisation of insurance pricing).

This can be a powerful tool to expand the provision of financial services, particularly if new types of data can be analysed to reveal risk characteristics of individuals that do not have a traditional financial history to assess.

There are concerns, however, that this can sometimes come at the expense of social benefits.

For example, in the insurance sector, more granular risk differentiation could lead to high risk individuals (who are often likely to need insurance the most) being priced out of the market – as premiums are increasingly based on individual characteristics rather than groupings. While this also occurs under traditional forms of risk assessment, the argument is that AI widens the range of possibilities for exclusion because it can assess many more factors.

A number of people we spoke to raised concerns that AI might undermine financial services which rely on risk pooling – particularly in fields such as health and life insurance – as advances, for example in genetic research, will enable much more accurate predictions.

A counter is that risk pooling will remain relevant because there will always be 'unknown unknowns'. (AI won't any time soon be able to tell us where lightning will strike). A report[xvii] published last year by the Geneva Association, an insurance research group, said:

> *"As long as individual risks retain some level of uncertainty and are not predictable with certainty, risk pooling has a role to play, even when big data allows a much better assessment of the risks. It is true, though, that the better individual risks can be predicted, the lower the value of insurance for policyholders and hence the lower an individual's willingness to pay."*

Equally, ML might be used to group together characteristics of different individuals that previously couldn't be linked, in ways that might challenge our concept of fairness. For example, an institution might be able to lower an individual's credit limit based on data on where he or she has made purchases, after an AI model has determined that others who have also shopped there have a poor repayment history.

Another potential risk is that the more personalised financial products enabled by AI could come at the expense of price transparency. Institutions could use ML to assess not only customers' individual risk profiles, but also their willingness to pay – which theoretically opens the door to customer surplus being extracted through higher prices for those personalised products.

The wider debate here is about what data sources institutions should be permitted to use to price and customise products. For example:

- Will customers who refuse to share their data with financial institutions, or to use sensors which enable firms to better assess risks, have to pay more for financial products?

- In what ways might individuals feel compelled to change behaviour because more areas of their lives are being monitored? While in some cases the links are clear (e.g. between giving up smoking and life insurance costs), in others they might be much more nebulous.

"A key question from a consumer perspective is how far you should go with Machine Learning and related technologies in assessing and pricing risk? How granular should you be with risk pools, with individual pricing of every policy being the extreme case?

"You can argue that this isn't a problem when the pricing of the policy is based on behavioural factors. If you are a reckless driver, why should others cross subsidise you? But what about when risk factors aren't behavioural? If you are born with certain medical conditions, you'll incur high costs regardless of your behaviour. This is largely a question for government policy, not just insurers."

Ermir Qeli,
Head Stargate Services,
Director, Swiss Re

New ethical challenges – key risks

Data acquisition and aggregation

The usefulness of ML is largely rooted in its ability to make use of a much greater volume and variety of data. This includes data which institutions already collect but that would previously have been unusable (particularly of the unstructured type, such as freeform text).

Firms' assessments of which data to use are often speculative, since it is uncertain which inputs will ultimately determine the outputs of a model. Therefore, there are mounting commercial pressures on financial institutions to expend resources to acquire and store data that could *potentially* be valuable. Once collected, this data must be aggregated, centralised, and processed before it can be used as the fuel for ML. It is also increasingly being shared between financial institutions and with third parties – through private arrangements and because of regulations such as open banking.

An increasingly important debate is around the commercial collection of personal data. There is mounting unease about cyber crime and fraud, erosion of privacy, and 'surveillance capitalism'. Collecting, preparing, and sharing data could all create new risks.

Undesirable behaviour of AI models

ML has the ability to perform much more impressive feats of prediction than previous computing capabilities – *if* the humans designing the application are able to unambiguously convey the objectives it should achieve and feed it the right raw materials. But if these conditions aren't met, AI models can produce undesirable outcomes. For example:

- <u>Algorithmic biases can arise</u>. The engines that determine credit eligibility, insurance premiums, or candidates for jobs might disproportionately penalise people with shared characteristics, such as race, gender, or background.
- <u>AI systems can perform poorly in previously unencountered situations</u>. For instance, a model designed for risk management could become ineffective during abnormal financial conditions – potentially amplifying the seriousness of 'black swan' events.

A key point is that ML can only derive insights from the data it has access to; it cannot establish context. If that data is incomplete or is skewed by historical factors that are no longer relevant, the result can be *garbage in, garbage out.*

Insufficient transparency

One of the more disconcerting qualities of many AI applications is that even the experts who designed them may find it is difficult (or impossible) to understand why the machine has taken an action – much less convey this to customers at the wrong end of a decision, or managers demanding answers when something goes wrong.

This strikes at fundamental questions about trust, both in technologies, and in the institutions that use them. An entire field has formed around 'explainable AI.' Even so, there is a fundamental trade off. Many of the characteristics which underpin the effectiveness of the most powerful ML algorithms are inextricably tied up in their opacity.

There is a spectrum of transparency. A lower threshold is *interpretability*. This is the extent to which the machine's inputs can be connected to its outputs, even if its internal mechanism remains foggy. Interpretability might, for example, be determined by tweaking a weighting in the algorithm and drawing inferences from changes in what it does. *Explainability* goes further, and generally refers to being able to explain cause and effect in ordinary language.

Optimisation at the expense of social benefits

One of the main benefits of ML is that it can parse many disparate datasets to make recommendations at a granular level – enabling financial institutions to offer personalised products to many more customers at little marginal cost.

However, there are concerns that in certain circumstances, a focus on optimisation might come at the expense of social benefits. For example, it could lead to insurance becoming prohibitively expensive for individuals with high risk factors (e.g. when genetic factors are used to determine health insurance costs), and it could undermine traditional risk pooling business models.

There are wider questions, too, about what data sources financial institutions should be permitted to use in order to price and customise products. For example, should customers be required to share their data to qualify for lower prices? And how might the prospect of being monitored and assessed in previously private areas of our lives challenge social norms around privacy?

Skills gap

AI talent is in short supply

"AI technologies combine computation skills, business knowledge, and statistical and mathematical frameworks. The number of people who understand all of those worlds and can work across them is quite limited. Traditionally our education establishments have been training for one of those buckets rather than for all of them. This is one of the key drivers of the skills gap. I think the way to address this is by bringing these skills together in collaborative groups – where each individual understands one of those dimensions."

Alaister Moull,
Financial Services Lead, PwC

Widespread demand for AI solutions is a relatively recent phenomenon, and has emerged simultaneously in many different industries. Financial institutions' struggle to find and retain AI talent was a pervasive theme in the interviews we conducted.

Hiring expert programmers who lack financial services knowledge is a pronounced risk. This was widely emphasised. It is more difficult to address the talent gap in financial services by bringing in experienced AI specialists from other industries than for the typical IT role, because the required skillsets are less transferable. (A mobile app developer at a tech company, for example, can generally transition into a similar role at a financial institution with minimal retraining).

Programming and statistical skills are necessary to build and test algorithms and interpret their outputs – in other words, to solve a problem in the right way. But a thorough understanding of the domain is essential to answer a more fundamental question: where should AI be incorporated into the business, and what problems should it attempt to solve?

We heard many concerns about technically clever AI solutions being designed without a comprehensive grasp of the company's business model, operating challenges and market conditions. This could lead to resources being wasted – or the assumption of new risks, such as unwitting regulatory violations.

A common sense fix is that the teams that implement AI applications should be composed of people with different skill sets, technology and business-related. But this is often easier said than done. In particular, communication is seen as a challenge.

One data scientist who has worked at multiple financial institutions told us: "One of the main problems I see is a lack of effective communication between technical and non-technical people. As data scientists, we need to understand where our stakeholders are coming from and to simplify what we've created to make ourselves understood. This is an essential skill set which is oftentimes neglected."

"There is a huge new wave of optimism around AI but this needs to be balanced with realism about expected outcomes. A data science degree doesn't mean that you can solve all the issues in the world. It's easy to naively take an off-the-shelf algorithm and put it up against a problem. But the person doing the analysis may not have thought about the data that's going into the analysis, how it's been collected, where there could be biases, or where their approach may not be robust enough.

"You need to have people with domain experience who understand the broader problem, as well as people who can do the number crunching. The danger is a lack of process maturity associated with the implementation of new AI solutions. The risk isn't the systems themselves, but the people doing the work not considering the complete context."

Tony Wicks,
Head of Financial Crime Compliance, SWIFT

On the other side of the coin, the financial services industry's proclivity for confusing terminology and jargon was noted as a concern for people coming into the industry for the first time.

The anatomy of an AI project

'Data scientist' was dubbed "the sexiest job of the 21st century" by the *Harvard Business Review* in 2012, and topped *LinkedIn*'s[xviii] list of the 'most promising jobs' in 2019. The designation is often thrown around as a catch-all to describe an umbrella of skill sets needed to design and implement AI systems. But projects typically require a broad range of professionals with proficiency in different areas. While the following list is not exhaustive, they can generally be grouped into a few categories:

- The AI product manager works with both the business stakeholder and the technical teams to gauge business requirements, and define an appropriate AI solution.
- The technical teams may consist of machine learning researchers, who search for patterns in the data to inform the application's algorithms, expectations, and potential approaches. Machine learning engineers then write production code to implement the algorithm. They work with data engineers and data platform specialists who build the 'plumbing' and maintain the systems which aggregate, clean, and transform the data into inputs for the algorithms – as well as feeding the algorithm's outputs to downstream systems.
- Throughout the life cycle of a project, a project manager will oversee day-to-day management and alignment of different teams.

'Data scientist' is not just a catch-all term, but also a specific role, in high demand. Data scientists usually own responsibility for producing results. They bring to the table an understanding of machine learning, statistics, and relevant domain expertise to determine whether a given approach addresses the right question in the right way – and guide the business in its implementation.

...but does the business side get it right?

"People on the business side need to learn, at a high level, about data science concepts: the methodologies and the vocabulary we are using. They need to be familiar with our mode of work, how things are being researched, and what it takes to deliver a project. This is critical, because it sets expectations about what it takes to generate results, and the timeframe over which it's possible.

"I think that the people who make decisions at banks often don't understand them well enough. They are often not familiar enough with different use cases or approaches – to recognise, for example, the difference between data analytics versus descriptive capabilities versus predictive capabilities. They need to realise that in each case there are often very different types of challenges, depending on what the goal is.

"Or, they don't understand the importance of data and data quality. They may underestimate the complexity of an event and give us a small data set, but over time realise it didn't capture the variety of events we need to deal with."

- Senior data scientist, global bank

Risks arising from the knowledge gap and unrealistic expectations

The extent to which decision-makers at financial institutions are 'AI-literate' varies widely. Nevertheless, concerns are widespread that decision-makers with an inadequate grasp of AI and its workings could:

- Have unrealistic expectations about what can be achieved. For example, they may be unable to define reasonable expectations for AI projects, accurately assess timelines, or take a realistic view of necessary resource allocation. A cross-industry survey by *KPMG* this year found that 47 per cent of executives expected "a significant return on investment" from AI within three years – dropping from 62 per cent in 2018 – which suggests that more decision makers may be finding that their initial expectations were too ambitious [xix].

- Underestimate the resources AI teams need to achieve their targets. These include clearly defined business objectives, access to the right technology infrastructure, and a sufficient quality, quantity, and breadth of data.

- Fail to ensure organisational readiness to integrate new AI applications into the business, such as by providing training on how to use the technologies effectively. In a recent survey of bank executives by *Accenture*, respondents said that on average, only a quarter of their employees are ready to work with AI.[xx]

- Fail to effectively evaluate the risks versus the benefits of AI, and make informed decisions around how to fold it into their overall business strategy.

Misinterpreting the outputs of AI

A key consideration when working with AI is knowing how and when to rely upon it. Due in part to hype and inflated expectations, there are invariably cases in which AI is applied when it shouldn't be. But even when the application is an appropriate one, there are risks to institutions from over-reliance.

One that was widely emphasised to us is the risk that the people who use AI as part of their jobs are unable to effectively interpret the machine's outputs.

On the one hand, blind trust is a problem. There are studies that reveal a tendency for people working with AI to unquestioningly accept machine decisions without checking them for common sense – even if they might have directly questioned the same decision from a colleague or manager. One reason is the accountability factor – if something goes wrong, the perception is that the fault lies with the computer.

On the other hand, the benefits of AI are largely obviated if users are too sceptical of its findings. The rationale for using AI is often that it is able to find hidden insights in data that humans may not be able to grasp. A common consequence of being overly critical is a tendency towards confirmation bias – where the people working with AI only trust its conclusions if it supports their inclinations.

The following is an amalgamation of 'best practices' that were conveyed to us for interpreting AI outputs:

- *Effective interpretation of an AI system's outputs requires a framework for understanding the process by which a model evaluates available options, and arrives at a decision or recommendation.*

- *The people who work with AI need to have a basic understanding of how the machine handles data and arrives at decisions. They need to be satisfied that the process has been thoroughly vetted, the data coming in is accurate and unbiased, and the assumptions the model is built upon are sound.*

- *Once they are satisfied that the AI has come to a decision in the right way and for the right reasons, they should feel comfortable trusting the outcome regardless of whether it supports or rejects their suspicions – provided it first passes a commonsense filter.*

Strategic realignment is essential

In a global, multi-industry survey [xxi] conducted last year by McKinsey, the most frequently cited barrier to AI adoption was the lack of a clear strategy for AI (cited by 43% of more than 1,600 respondents). The third most cited (by 30%) was the existence of functional silos that constrain end-to-end AI solutions.

A number of the people we spoke to said that taking a haphazard approach is one of the main reasons financial institutions are failing to implement AI effectively. However, organisational realignment can be a daunting task. One senior risk manager at a large financial institution told us: "It's hard for large companies to organise AI because the knowledge transfer is so expensive. You can't get away with splitting it like a production line; the whole is interconnected, and it interconnects in different ways. That makes it an extremely tough structural challenge."

One key requirement is to create clearly defined hierarchies and lines of accountability. This is likely to include appointing a C-level executive who has clear ownership of AI initiatives and a mandate to implement long-term solutions across many functional groups – for example, a 'chief data officer' or 'chief analytics officer'. It might also involve an explicit separation of duties between this executive and the heads of functional groups that incorporate AI into their processes.

Another is to bring together previously siloed data, which may have been managed by different groups, and to centralise the infrastructure required to house that data and make it available. This might involve establishing new functional groups which serve the entire business, including:

- <u>Data groups</u> – which gather, aggregate, clean, and maintain corporate and user data, and provide it to different business functions, in accordance with a single set of standards.

- <u>Business intelligence</u> – which establishes standardised Key Performance Indicators and definitions across the organisation, and produces reports and analysis.

Getting strategic alignment wrong can result in many risks. These include:

- <u>A lack of ownership and accountability</u>. In particular, we heard concerns around accountability for decisions made by AI, the quality and availability of data being fed to algorithms, compliant usage of the data (both regulation and customer contracts), and employee training.

- <u>Creating point solutions to incorporate AI into the business that do not work well together</u>. Incremental introduction of AI initiatives may be necessary if firms don't have the resources to implement a top-level strategy all at once. However, significant risks arise when initiatives are deployed independently of one another.

- <u>Fragmented data aggregation and clean-up initiatives, with overlapping goals.</u> This can result in expensive and time-consuming duplication of efforts, as well as increased vulnerability of systems to cyber crime.

Where should boardrooms be paying attention?

In a recent speech[xxii], James Proudman, the Bank of England's executive director for supervision of UK deposit-takers, laid out three principles for governance of AI/ML at financial institutions. He said that boards should focus on:

1. The governance of data – what data should be used, how it should be modelled and tested, and whether the outcomes derived from the data are correct.

2. The oversight of human incentives and accountabilities within AI/ML-centric systems.

3. The range of skill sets and controls that are required to mitigate these risks both at senior level and throughout the organisation.

There are, however, widespread concerns that boards often lack the ability to critically evaluate technologies. For example, corporate governance was a significant and rising concern in the CSFI's latest *'Finance for all'* survey. Many respondents to the survey took the view that more short-term, profit-oriented business styles were eroding governance standards and tilting the balance of power towards management. One said that: "Many boards seem unprepared to lead the digital strategies needed at their institutions to stay relevant."

Boardrooms lacking in IT literacy

"The biggest problem is that many banking boards have very little understanding of basic IT, let alone advanced technologies like AI. That makes it very difficult to ask insightful questions. There is far too much acceptance of what they're being told, because they don't know any differently. Consequently, there's a real danger of boards signing off on things they don't understand – without questioning policies, procedures, strategies, and potential alterations to the business model.

"For board members to effectively challenge executives, they need to have a good idea of what constitutes a reasonable answer to their challenge. Without that, they are likely to be too easily brushed aside."

Brandon Davies, trustee of the Responsible Finance Institute Foundation and non-executive director in financial services.

Skills gap – key risks

Talent gap

There is a global shortage of people who can design, develop, deploy, test, and maintain AI systems. Financial institutions – especially those in less lucrative sectors – are struggling to attract and retain talent with the right skill sets. Experienced data scientists and AI specialists are sought after by many institutions, in many sectors – meaning they can typically expect high pay and a credible promise of fulfilling work.

A consequence is that many of those who are available have little experience working in the financial services industry. New entrants into the field typically have advanced degrees in subjects such as computer science, without domain expertise. There are widespread concerns about financial institutions building AI projects which, while technologically imaginative, may fail to address business priorities and risks.

In the last few years, universities have begun to offer specialised data science degrees. There are also some programs that are starting to combine data science with broader financial services understanding. The talent shortage will likely be addressed, but it could take many years to catch up with demand, especially at senior levels.

Knowledge gap & unrealistic expectations

On the other side of the talent gap is a perceived 'knowledge gap': a lack of understanding among many of the non-technical people at financial institutions around the workings – and limitations – of AI technologies. These people include the executives and other decision makers who set goals and sign off on projects, line of business managers who oversee them, and employees expected to work alongside new systems. This can lead to risks, including:

- Inflated expectations from decision makers about what projects can achieve, and over what timeframe.
- The relevant people at financial institutions failing to effectively work and communicate with data scientists and/or third-party vendors.
- A failure to provide AI teams with the resources they need to do their job well (e.g. sufficient high-quality data or IT infrastructure).
- Institutions lacking a 'culture' of how to work with AI, which becomes especially important as more employees are expected to incorporate these technologies into their jobs.

Overreliance on AI

While failing to adopt AI can pose a competitive risk, on the other side of the coin are concerns that institutions are trying to implement AI 'for its own sake' – when a simpler and cheaper alternative could have been as effective. (One expert in financial regulation told us that: "Lots of businesses will say 'we need to have an ML expert'… so they'll hire a ML expert and instruct them to 'go and find problems'".)

Even when the application is appropriate and well-implemented, there are risks to institutions from over-reliance. One is an inability to critically evaluate the outputs of AI. This could be due to the people working with systems trusting them blindly without a 'commonsense check', or overriding machine decisions that don't support their preconceptions.

Another is removing humans from business processes prematurely. A poorly managed transition to AI could lead to gaps in services – especially when technology fails. This could arise, for example, in customer support models, where human agents are replaced too hastily by chatbots and automated phone assistants. Moving to low touch models without alternative options could also exclude certain market segments (for example, older generations.)

Inadequate strategic alignment & governance

Due to the scope and scale of data that is generally needed to implement AI effectively, initiatives at financial institutions are often large and require a high degree of cross-functional cooperation. If firms fail to align their organisational hierarchy and structure accordingly, they expose themselves to risks from poor management and leadership.

Inadequate strategic alignment might lead, for example, to a lack of ownership and accountability, implementation of point solutions that don't work well together, and/or fragmented data aggregation and clean-up, leading to costly duplication of efforts. The result is that it may take institutions longer to accomplish less with more money, and expose them to security and compliance risks. On the other hand, it was emphasised to us that effective strategic alignment is one of the best ways to mitigate risks around AI implementation.

Corporate governance should play a central role in defining and setting controls around a top-level AI strategy. One concern is that many financial services boardrooms lack IT literacy. If boards aren't able to evaluate the costs and benefits of AI applications, they may sign off on decisions without understanding their implications.

Market dynamics

What impact could AI have on market concentration?

A wave of AI-powered fintechs is already challenging incumbents in many parts of the industry. Among these are retail lenders, insuretechs, robo-advisors and AI traders – generally nimble outfits that are often staffed by no more than a few dozen people. Many can operate at a fraction of the cost of incumbents, and embrace new distribution channels that appeal to tech-savvy customers.

In addition, technology sector giants that already have access to huge quantities of data and the resources to exploit it – such as (in the US) Google and Amazon, and (in China) Alibaba and Baidu's financial arm, Du Xiaoman – are increasingly exploring ways to deliver financial services.

For the moment, at least, AI appears to be spurring rather than suppressing competition in the financial services industry.

There are, however, questions about how this will play out in the longer term. Much research has been conducted into the 'winner-takes-all' tendency of technology markets, due, for example, to the high fixed costs of implementing technologies, low marginal costs of serving new customers, and network effects that companies benefit from as they scale.

These tendencies could become even more pronounced. One of the key findings of the *World Economic Forum's* recent horizon-scanning study of AI in financial services [XXIII] is that it could lead to the "bifurcation of market structure". The argument is that scale-based institutions have a natural cost advantage which will enable them to capture customers from mid-sized firms. However, smaller, nimbler players – including from outside the traditional financial services industry – might carve out niches by optimising their algorithms to the needs of under-served customers. AI also lends itself to a new kind of network effect that could make markets more vulnerable to the formation of monopolies.

These are *data network effects* – which can arise because ML models automatically improve as they gain access to more data. The idea is that the more widely a product is used, the more data it will have access to and be able to train on. This, in turn, will make it more valuable to customers, which will increase its usage, in a virtuous feedback loop. In theory, with every cycle, its predictions will become more accurate and error rates lower.

Two perspectives on AI's potential impact on the concentration of financial markets:

"With the rise of blockchain networks, the boundaries between financial markets and other industries are being brought down. That's why I think we're not going to have a single player in financial markets, because everything is going to be mixed. Financial markets will get data, for example, from supply chains, food networks, etc., so that they can provide better financial services to their clients. They will engage outside their industry much more than they're doing now. That opens up space for innovation – and for smaller fintechs to collaborate with some of the bigger players. I think there's room for many players: one player in financial markets isn't going to rule it all."

Søren F. Mortensen, Director of Global Financial Markets at IBM

"At the moment, lots of companies trying to enter the space are challenging big financial institutions who are competing on standardised products and price. New financial institutions are coming up with much more bespoke services. But, in five years' time, we might see network effects coming into financial services. We might then see more of a dark side of the technology. Every other sector has built tech giants – one in each sector. In five years, we're going to get a feel for who are going to be the winning companies. Perhaps it will be winner takes all, or oligopolies. At that time, we'll begin to see how AI will reshape the financial services sector."

Professor of Economics

Data network effects could lead to substantial barriers to entry for new entrants. The effect is not just a theoretical one. It has already been widely observed in technology markets which make heavy use of AI, for example, translation services and media recommendation engines.

Where this could potentially arise in the financial industry – and whether it might be traditional players or newcomers that take advantage – is an open question. But, generally speaking, it's likely that the financial services that could benefit most from continual improvement – and therefore might be most vulnerable to this kind of monopoly – are those which deal with forecasting and prediction rather than, for example, the automation of repetitive tasks.

There are clear benefits to the users of financial products that continuously improve. The risk is that data network effects may create barriers to entry which are difficult to overcome, making effective competition in certain markets increasingly implausible. A difficult question for regulators here is how to police monopolies that may have arisen as a natural consequence of the technologies being used, rather than from anti-competitive behaviours such as predatory pricing.

AI could lead to new forms of interconnectedness

In particular, risks can arise as ML-driven algorithmic trading (including high frequency trading) becomes increasingly prevalent.

These risks are not new. A number of 'flash crashes' over the past decade have been attributed to rule-based algorithms that exacerbated herd behaviour by human traders and triggered vicious cycles of selling.

But there are additional concerns around AI trading models, because of a heightened risk that arises at the confluence of automation, speed, and 'learning'. The fact that models optimise themselves in real time (rather than following pre-programmed rules) introduces an extra degree of opacity and uncertainty, which is magnified by the lightning speed at which algorithms make decisions. Several of the people we spoke to said that this could increase the probability of very sharp flash crashes.

Widespread AI use could lead to co-variance between previously uncorrelated systems. According to a 2017 report[XXIV] by the *Financial Stability Board* (FSB):

> *"Institutions' ability to make use of big data from new sources may lead to greater dependencies on previously unrelated macroeconomic variables and financial market prices… As institutions find algorithms that generate uncorrelated profits or returns, there is a risk these will be exploited on a sufficiently wide scale that correlations actually increase. These potentially unforeseen interconnections will only become clear as technologies are actually adopted."*

The point being made here is that AI and ML applications could create new and unexpected interconnectedness in financial markets. If a critical segment of the market relies on the same data sources and algorithmic strategies, a shock to these sources: "could affect that segment as if it were a single node… even if, on the surface, [it] is made up of tens, hundreds, or even thousands of legally independent financial institutions", the FSB warns.

Furthermore, competing algorithms implemented by many different firms may become interconnected – the output of one being used as the input into another many times over.

One question this raises is whether data streams could be vulnerable to manipulation. For example, if multiple organisations are using NLP to perform sentiment analysis on social media feeds, and their AI applications act in a similar way at the same time, a bad actor might be able to put out a cluster of 'pseudo news' stories to influence an indicator of data and manipulate market prices. While human traders are also vulnerable to this kind of manipulation, the speed at which algorithms act may make it harder to rectify an error before it gets much worse.

A growing reliance on third parties. Both market concentration and interconnectedness could increase as financial institutions become dependent on a small number of third-party providers to provide essential services. For example, the cloud technologies that provide computing infrastructure for AI are currently dominated by a few large players, such as Amazon Web Services and Google Cloud Platform.

As AI usage becomes more widespread, third parties are likely to arise that can provide certain AI applications more effectively and cheaply than financial institutions can produce in-house. These are most likely to be commoditised services that firms rent to minimise costs. For example, third parties might specialise in offering ML-powered back office processes or chatbots 'as a service' to many institutions. Those which are widely adopted could, due to their scale and acquisition of data, build 'best in class' services which would be virtually unchallengeable.

A number of the people we spoke to expressed concerns that as such interdependencies grow, an isolated failure at one third party institution might become magnified and quickly spread through the system.

What are the risks of moving too quickly – or too slowly?

The potential size of the AI prize, and the perception of significant first-mover advantage, has created an environment in which decision-makers at financial institutions face pressure to act quickly to capitalise. On the other side of the coin, a senior banker told us: "Most banks are most concerned with falling behind, rather than necessarily the rewards of being first. The danger is that, if we don't do this, our competitors will." A failure to act might mean that firms:

- Incur higher operating costs than their peers.
- Fail to retain customers.
- Fail to get in early enough to capture significant share in emerging markets which can be made accessible through AI.
- Are relegated to peripheral or less profitable parts of the industries they serve.
- Lose human talent to more innovative institutions.

The constraints of regulatory rules around the financial services industry should not be underplayed. Regulation – and the threat of hefty fines for mishaps – was emphasised by many of the people we spoke to as a significant speedbump, particularly at larger institutions.

Whatever, it is clear that the potential costs of inaction are weighing increasingly heavily on the sector. It's likely that as AI is widely adopted over the next few years, many firms will perceive the risks of their business model being undermined as so great that they will consider many of the other risks they face second order.

The problem is that rushing into AI creates risks, both at the institutional level and at the market level. These include:

- ***Implementing AI solutions without proper testing.*** This was flagged as the chief risk that AI poses to the financial services industry by a number of data scientists that we spoke to. One told us: "A lot of the problems that arise ultimately come down to a lack of care and due diligence – for example, failing to pay attention to monitoring, logging, audits and testing of models. It's the boring stuff which is critical".

> "There's a lot of pressure to succeed. It's hard when you're coming up with a project at a large bank, and you're thinking: what's the business problem I'm trying to solve, and how are we going to apply machine learning to it? Also, is there enough data to support machine learning approaches, and is that data good quality? You often need to demonstrate value in a relatively short period of time. Juggling these questions is not easy at all; these are all recipes for problems down the road.
>
> "Are people going through an appropriate process to ensure the model they are running is real as opposed to random chance? And are they monitoring performance appropriately as the model is run in real time and traded, to check to see whether performance is being maintained or not? This can be even harder with AI techniques, because of the complexity of the model involved. There's an additional element of risk which should not be underestimated."
>
> Aric Whitewood, Founding Partner at XAI Asset Management, and former head of data science at a global bank.

- ***The accumulation of 'technical debt',*** i.e. the deferred costs of implementing a technological solution that can be put into operation quickly – but which increases future difficulty in development and maintenance – over a more rigorous approach that that would take longer. Several academic studies[XXV] have observed that ML-based technologies have a particular propensity to generate technical debt. This can mean that:

- Systems become increasingly buggy over time and future integrations may be more difficult to perform.

- It is difficult or even impossible to comply with certain regulatory requirements (the GDPR's 'right to be forgotten' was noted as a particular concern).

- Institutions become over-reliant on certain individuals who are familiar with the idiosyncratic workarounds that have been implemented into systems.

- It is difficult to replace systems or shift direction in the future.

- There is a lack of interoperability of systems acquired through M&A.

Like debt of the monetary kind, technical debt accrues more 'interest' as time passes – meaning that shortcuts implemented in the short term can create more problems the longer they go unaddressed.

- ***Over-leveraged talent.*** Several of the people we spoke to expressed concerns that financial institutions might become over-reliant on AI specialists with highly technical skill sets that decision-makers do not sufficiently understand. There are parallels here with the industry's uncritical trust in quantitative analysts in the lead up to the global financial crisis. A senior risk manager at one financial institution told us: "In many cases it isn't just that there are similarities… it's literally the same people repackaging their skills for data science roles, since there's a lot of crossover".

Stephen Blyth, Professor of Statistics at Harvard University, said recently in the *Harvard Business Review:* [XXVI]

> *"There are many unsettling parallels between the recent advances in machine learning and algorithmic trading and the explosive growth of financial engineering prior to the crisis… Like today, financial engineering around the millennium brought great commercial success: the mathematical tools developed by derivative desks built businesses, boosted profits and delivered superior investment returns. We were taming financial uncertainty, or so we believed.*
>
> *"…The financial crisis exposed that mindset as a 'quant delusion,' an episode which we may now be at risk of repeating."*

Striking a regulatory balance

The debate about how AI should be regulated in financial services is extremely broad, and we do not go into it in detail. However, Germany's Federal Financial Supervisory Authority last year laid out ten implications of the proliferation of big data and AI which are key issues for regulators and supervisors. These provide a very good overview. They are:

1. Emergence of new business models and companies.

2. Connecting markets and market participants (e.g. risks from increased interconnectedness).

3. Using technology to limit undesirable developments (e.g. via technological safeguards).

4. Redefining and addressing systemic importance.

5. Governance.

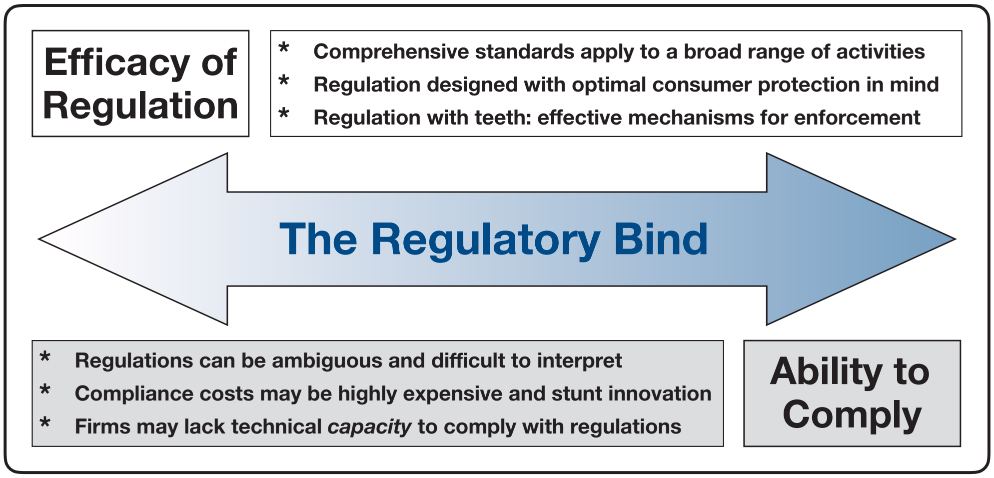
6. Fighting financial crime and preventing conduct violations.

7. Subjecting internal models to supervisory approval.

8. Handling information security risks.

9. Risk of discrimination.

10. Ensuring trust in the financial market.



In our conversations, the issue that arose most frequently was around the need to strike a balance between ensuring that regulations are rigorously and evenly enforced, and not stifling good innovation.

The biggest concerns we heard about the potential for undercooked regulations related to customer-facing applications of AI – in particular, data security and privacy, the potential for biased outcomes, and transparency of AI models. At many financial institutions, it isn't overregulation that is causing the most anxiety, but the potential for an uneven playing field and a lack of clarity in the rules.

On the other hand, there's a feeling that financial regulators often aren't supportive enough of experimentation that has the potential to offer benefits to the industry. For example, Jane Jee, Barrister and CEO of Kompli-Global, an AI-driven regtech provider, said: "A bank thinks: why experiment when the regulator will give us no credit for it and in the process we may risk being fined or sanctioned? Indeed, why take the risk even if we can see that the new technology would reduce financial crime? Regulators should issue praise where a bank adopts an effective new technology and issue examples of good practice – provide some carrots for combatting financial crime."

Implementing rigorous regulation without stifling innovation

"It's extremely difficult to get the right tech people at regulators – and it's also difficult to get these tech people positioned where they can really have impact. We've got traditional IT departments at both regulators and financial institutions walled off in their own world; what we need is tech people at regulators who are sitting knee to knee with product designers and marketing people and innovators to really transform how we do things.

"I think it is very important that as we look at these new systems we should not be comparing them to perfection, but rather to how we do things today. There are many challenges with using AI in many areas, but there are a lot of problems with the current systems too. Take driverless cars – if we expect them to never have an accident, we'd never allow them onto the road, but they are probably safer than human drivers.

"I would emphasise that regulators and firms together need to develop standards for best practice on the design of safe and fair AI systems. We need to create the ability to audit an AI system: is it using enough data to be statistically valid, is the training data biased, is the data accurate? We need to be able to audit outcomes in areas like discrimination. You can run those tests in parallel, test the AI against traditional underwriting systems, and analyse it to see whether it was more fair and inclusive."

Jo Ann Barefoot, CEO and Founder of the Alliance for Innovative Regulation (AIR)

Market Dynamics – key risks

Market concentration

There has been a flurry of AI-powered disruptors challenging traditional financial services models. These include agile fintech start-ups and some of the world's largest technology companies. New competition is a serious risk for incumbents that cannot adapt, but it offers many benefits to customers and the wider industry.

There is, however, uncertainty around how the market will evolve. Many analysts have noted the tendency of technology markets to be dominated by a few big players, due to economies of scale and network effects. Technologies based on ML can also realise *data network effects* – a virtuous cycle of automated product improvements that may make it very difficult for others to effectively compete in the market.

Concentration might also arise among the third parties that offer services to financial institutions. In many cases institutions will rent these services rather than attempt to develop their own.

The dominance of a small number of vital players in the financial services ecosystem could lead to risks – for example, dependent institutions being locked into relationships on bad terms, regulatory capture by dominant players, and increased barriers to entry. It could also create new systemically important institutions, with associated concerns around 'too big to fail' and moral hazard.

Increased interconnectedness

One of the most alarming revelations of the financial crisis was the degree to which global financial markets were interconnected.

How – and to what degree – AI might exacerbate the risk of contagion is speculation. But there is a plausible case that these technologies will create new kinds of interconnectedness – at the data level, the IT systems level, and the decision-making level.

This might include:

- Widespread use of the same publicly available datasets as inputs into decision making, which could create covariance between previously uncorrelated systems.
- Algorithms implemented by many different firms becoming increasingly interconnected over time – the output of one being used as the input into another, many times over, in a mesh of interdependencies.
- Many financial institutions becoming over-dependent on a few third-party tech providers, leading to vulnerability to single points of failure

Pressure to move too fast (AI 'arms race')

Financial institutions face pressure not only to deploy AI, but to act *quickly*. While regulatory constraints serve as a speedbump, there's a sense that many institutions fear their business models being undermined, or even made obsolete, by new technologies. The rewards for being first to market, meanwhile, are potentially vast.

Moving too fast, or failing to put in place safeguards, exacerbates most of the risks outlined in this report. One particular concern was insufficient testing, auditing, and monitoring of ML models. Another is the accumulation of 'technical debt' – the cost of implementing a technological solution quickly, at the expense of future difficulty in development and maintenance.

There is also concern that decision-makers might become over-reliant on AI specialists with abstruse skillsets and knowledge that they don't understand – with parallels to the uncritical trust in quantitative analysts in the lead up to the financial crisis.

New regulatory challenges

Regulators and policymakers are increasingly aware that AI poses challenges that are fundamentally different from previous technologies – both in terms of the ethical questions it raises and its potential to fundamentally transform market structures. Questions they must confront include:

- How to foster a regulatory environment which protects consumers without suppressing innovation?
- Which institutions should fall under the scope of financial services regulation, as more non-traditional firms challenge incumbents and lines between sectors become increasingly blurred?
- How to protect competition in financial markets, while also acknowledging that AI needs scale to be effective?
- How to ensure regulatory consistency between sectors and jurisdictions, and to avoid institutions taking advantage of regulatory arbitrage, for example, where data laws are more permissive?

Chapter 4: Conclusion

There is nothing magical about AI – and anyone who expects magic will be disappointed. At its core, AI is about finding patterns in data. It has no agency of its own, and no access to common sense.

But it does not 'just' mean faster computing. A brand new suite of technologies, centred around ML, will change financial services in fundamental ways – both in terms of efficiency and new capabilities. The scale of this transformation could swiftly match or surpass the impact that IT has had on the industry over the past few decades.

The potential benefits are compelling. AI opens the door to lower costs, lower error rates, and more personalised services that are accessible all the time. It can help democratise financial services and change for the better how institutions manage and mitigate risks.

At the same time, the proliferation of AI exposes financial institutions to important new or magnified risks. Some are inextricable from the characteristics of new technologies and the methodologies used to implement them – collecting and processing data, implementing algorithms, and relying on their outputs. Others stem from, or are exacerbated by, a lack of human understanding and preparedness.

In this report, we have identified three characteristics of AI technologies that may bring about new kinds of risks: their complexity and opacity, their ability to distance humans from decision making, and the changes in incentive structures they are likely to bring about. We then introduced a framework of three broad groups of risks: 'new ethical challenges', 'skills gap', and 'market dynamics'.

Another way to frame these risks is by looking at where the consequences might be felt.

Consequences for consumers

Opportunities…

In mature economies with high levels of financial participation, AI could benefit customers by enabling the provision of personalised and higher value products, while lowering fees, and offering new tools to equip them with the ability to better manage their finances. In markets with widespread financial exclusion, the benefits may be even more significant. AI can help bring an end to the exclusion of billions of people from savings, credit and insurance products.

…and risks.

AI could create winners and losers among consumers. While it enables institutions to use insights from past data to make more accurate predictions, these predictions might entrench biases or create new ones, at the expense of marginalised groups with shared characteristics. A lack of transparency around why decisions have been made could mean that customers are locked out of the financial system. More granular assessment of risks – right down to the individual level – could benefit some to the detriment of others, where customers deemed higher risk might find that the financial services they need are prohibitively expensive.

A second major consumer protection question is around how personal data might be used. Consumers could suffer from material damage caused by the theft of sensitive data – especially as institutions collect and centralise more of it, and process it in ways which make it even more valuable to cyber criminals and fraudsters.

Related to this is the question of where AI might cross the 'creepy' threshold. This is largely subjective and will vary widely between individuals and societies. Among other factors, it depends on the types of data that consumers are comfortable with institutions holding about them, and how it might be used to make predictions about their characteristics and behaviour.

For example, we are accustomed to car insurance providers using risk factors at a group level, such as age and experience, to price policies. What about sensors placed in the vehicle that determine how an individual is driving? Or, going further still, AI-powered facial detection technology that scans a driver in real time for signs of distraction? Should lower prices be predicated on the willingness of consumers to relinquish their data? Will the perception that previously private behaviour can be monitored and fed into an algorithm lead to changed behaviour? What correlations should an ML engine used, for example, for scoring be permitted to evaluate – and should we require that these correlations clearly demonstrate causality?

Big data and AI put more knowledge about consumers in the hands of institutions and governments, for better or worse. To what extent might consumers be vulnerable to institutions which become too powerful – and might the winner-takes-all tendency of AI technologies reduce consumer choice? There are concerns that AI might increase the risk of predatory behaviour by dominant institutions, for example, by enabling them to extract too much consumer surplus.

Consequences for institutions

Opportunities…

Institutions see AI as a way to slash their operating costs and bolster revenues. For example, considerable savings might arise from the automation of repetitive work, reduction in resources and time spent on regulatory compliance, improved detection of crime and fraud, and better capital allocation. On the revenue side, AI enables firms to sell a greater range of products, identify likely customers and improve marketing, and gain access to new markets. Firms that use AI early and effectively may enjoy network effects that give them substantial first-mover advantages over their competitors.

…and risks.

On the other hand, ineffective use of AI – exacerbated by competitive pressures to move faster than would otherwise be optimal – can hurt institutions' bottom lines. AI solutions which are implemented prematurely might fail – due to bandwagon-jumping, short-term thinking, and models that are put into production without a full understanding of their consequences.

For example, robotic process automation can dramatically increase the speed and precision of performance of repetitive tasks, but an error in the model could result in the same mistake being made many times over before it is spotted. In more advanced AI applications – such as trading and capital allocation – entrusting decisions to algorithms might cause institutions to unwittingly take on new or magnified risks. Models may become ineffective in untested market conditions. Failure could also arise at institutions that do not invest in the right kind of IT infrastructure needed to support AI, leading to outages and vulnerability to attacks.

Pressure to implement AI could also lead to wasted resources, if the wrong application is applied, or if the use case is poorly suited to AI. This might be due to inflated expectations about what AI can achieve, or the development of multiple applications which work badly together and result in duplication of efforts.

Institutions which are perceived to push use of AI beyond socially acceptable boundaries could suffer serious reputational damage and loss of credibility.

For example, in the CSFI's 2019 *Insurance Banana Skins* survey, the chief executive of a UK insurer said a main risk to the industry is: "A breakdown in trust between customers and their insurers – driven primarily by pricing practices, which could be… because of a rejection of the amount and sources of data being used to price products in a way that is (at best) far from transparent, and (at worst) seen as being unfair". There's a growing view in the financial services industry that the loss of trust from a high-profile data breach is likely to be more damaging to an institution than the material loss, and might even pose an existential threat.

Another major concern is that using untested technologies might expose firms to large fines or sanctions from regulators. For example, several of the people we spoke to worried that the opacity of AI systems and the data that feed them leads firms to unwittingly violate regulations and incur serious penalties. How might poorly formulated rules put firms in a situation where they lack even the technical capacity to comply?

Beyond the short term, the biggest questions for many financial institutions are around whether they might lose competitiveness and fail to remain relevant in a world where use of AI is widespread. To what extent might AI undermine existing business models? Firms that fail to evolve quickly and effectively enough may lose market share, face higher operating costs than their peers, and be unable to attract and retain critical talent. Adapt or die?

Consequences for the financial system

Opportunities…

AI could help create a global financial system that is more efficient, more inclusive, and more effective at detecting criminal behaviours such as fraud and money laundering. A greater quantity and wider range of data, from new macroeconomic indicators to measures of sentiment, can be used to anticipate threats to financial stability before they arise. Both supervisors and institutions see great promise in using AI to better enforce, monitor, and comply with regulations.

…and risks.

AI-driven business models could give rise to more actors that are systemically important in financial markets. These might be providers of financial services that take advantage of the economics of scale and network effects that AI applications benefit from. They might also be third parties that large numbers of institutions rely upon – for example, providers of the IT infrastructure upon which AI is run, or data aggregators.

Greater market concentration and new kinds of interconnectedness between institutions, the algorithms they deploy, and the data they rely upon to feed these algorithms, raise potential concerns around actors that are 'too big to fail' and 'too connected to fail'. This might increase the magnitude of isolated failures, where a single programming error can result in a cascade of downstream effects. Data source streams could plausibly be vulnerable to manipulation by a bad actor. Failures by external technology providers could also have consequences that reverberate throughout the financial system.

A related point is that regulatory inconsistencies – between sectors or regions – could increase the concentration of risk in financial services. For example, will the tech companies that become an increasingly important part of the financial services ecosystem be held to the same data standards as traditional financial institutions? Might there be scope for institutions to exploit regulatory arbitrage in jurisdictions where data laws are more permissive?

Finally, could use of AI contribute to a future financial crisis? This is a highly speculative question, but it is starting to be raised and is worth exploring.

One trigger might be a particularly sharp "flash crash", where many interconnected AI trading programs react in the same way to some market event. One AI trader we spoke to said: "While algorithms are supposed to be adaptive and we want to build robustness and the ability to deal with different market regimes into them, people do end up using models which actually do very similar things. If there's crowdedness or some sort of similarity in behaviour, does that increase the risks of losses for a large proportion of people?"

A second trigger might be an event which undermines public faith in the financial system – for example, a coordinated cyber-attack against many institutions, possibly conducted by a state actor, which cripples critical IT infrastructure. This might lead to massive amounts of data falling into the wrong hands, or widespread disruption of vital services. This could be more plausible as AI increases data collection and aggregation, and perhaps the number of systemically important institutions.

A recent publication[xxvii] by the Bank of England argued that while the connection is not self-evident, there is a credible case to link cyber risk to systemic risk in the financial sector. The authors said: "We are seeing a further growing gap between the technology environment we operate in and our ability to understand and secure it. As we build automated processes and artificial intelligence into its services, this will, by definition, compound the problem; making the mitigation of attacks significantly more challenging."

A third relates to financial institutions using AI for risk management. For example, how will ML-powered risk management models trained on data when market volatility was relatively low – which may work very effectively the vast majority of the time – react to extremely rare 'black swan' events?

A paper[xxviii] published a couple of years ago by the Systemic Risk Centre at the *London School of Economics* said that AI will likely result in considerable benefits as far as managing institution-level risks are concerned, for both risk managers and supervisors. But it also argued that when it comes to financial stability, AI is likely to miss out on the most dangerous threats, because systemic events are rare and unique. The authors concluded:

> *"Ultimately, the increased use of artificial intelligence in financial policy may result in us becoming very good at managing day-to-day risk at the expense of tail risk. Lower volatility and fatter tails."*

Outcomes ultimately depend upon humans, not machines…

<u>It is critical that the financial services industry gets AI right.</u> And this means not just understanding the opportunities, but anticipating what could go wrong.

This is a big subject. We selected the risks laid out in this report because we believe they are illustrative of the wider issues. We do not argue that the consequences are inevitable, though they are certainly plausible. Many of the people we spoke with made the point that many financial institutions are erring on the side of caution, waiting for regulators to tell them what is and isn't permissible and limiting their use of AI to applications where it cannot cause much damage if something goes wrong. Much will depend on the quality of regulation and governance. However, as we've described, AI creates strong first-mover incentives – and in a winner-take-all environment, risk assessment is quickly skewed.

The idea for this report came from conversations with non-technical finance professionals who were aware of the rapidly growing importance of AI in their sectors, but expressed concern that they lacked even a basic grasp of its workings. In coming years, it will be increasingly common for practitioners to work alongside these technologies. It is therefore crucial that they – and particularly decision-makers – are able to critically evaluate them.

A decade after the global financial crisis, the world is still grappling with the ramifications of the industry's over-eager embrace of complex financial instruments. Any comparisons to be made with the impact of AI are speculative, but the parallels should not be dismissed out of hand.

References

i **The New Physics of Financial Services – How artificial intelligence is transforming the financial ecosystem . (2019). World Economic Forum.** https://www.weforum.org/reports/the-new-physics-of-financial-services-how-artificial-intelligence-is-transforming-the-financial-ecosystem

ii **Notes from the AI frontier: Modeling the impact of AI on the world economy. (2019). McKinsey & Company.** https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy

iii **The Digitization of the World: From Edge to Core. (2018). An IDC White Paper, sponsored by Seagate.** https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

iv **Speech by Governor Lael Brainard on what we are learning about artificial intelligence in financial services. (November 2018). Board of Governors of the Federal Reserve System.** https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm

v **AI in RegTech: a quiet upheaval (2019). IBM and Chartis.** https://www.ibm.com/downloads/cas/NAJXEKE6

vi **Worldwide Spending on Artificial Intelligence Systems Will Grow to Nearly $35.8 Billion in 2019. (2019). IDC.** https://www.idc.com/getdoc.jsp?containerId=prUS44911419

vii **AI in banking: the reality behind the hype. (2018) Financial Times.** https://www.ft.com/content/b497a134-2d21-11e8-a34a-7e7563b0b0f4

viii **The Rise of AI in Financial Services (2019). Narrativescience.com.** https://narrativescience.com/wp-content/uploads/2018/11/Research-Report_The-Rise-of-AI-in-Financial-Services_2018.pdf

ix **SAS, GARP survey: 81 percent of risk professionals already seeing value of AI. (2019). Sas.com.** https://www.sas.com/en_is/news/press-releases/2019/february/artificial-intelligence-risk-garp-survey.html

x **Global risk management survey, 11th edition. (2019) Deloitte.** https://www2.deloitte.com/content/dam/insights/us/articles/4222_Global-risk-management-survey/DI_global-risk-management-survey.pdf

xi **Global financial services industry bullish towards disruptive technology but widescale adoption yet to come . (2019). Intertrustgroup.com.** https://www.intertrustgroup.com/news-and-insights/insight-news/2018/global-financial-services-industry-bullish-towards-disruptive-tech

xii **The $450B opportunity for the applications of artificial intelligence in the banking. (2019). Business Insider.** https://www.businessinsider.com/the-ai-in-banking-report-2019-6?r=US&IR=T

xiii **Big Data in the Financial Services Industry: 2018 – 2030 – Opportunities, Challenges, Strategies & Forecasts - SNS Telecom. (2019).** http://www.snstelecom.com/bigdatafinance

xiv **Amazon scraps secret AI recruiting tool that showed bias against women. (2018). U.S.** https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

xv **Interim report: Review into bias in algorithmic decision-making. (2019). GOV.UK**
https://www.gov.uk/government/publications/interim-reports-from-the-centre-for-data-ethics-and-innovation/interim-report-review-into-bias-in-algorithmic-decision-making

xvi **The future of regulation: AI for consumer good. (2019). FCA.**
https://www.fca.org.uk/news/speeches/future-regulation-ai-consumer-good

xvii **Big Data and Insurance: Implications for Innovation, Competition and Privacy. (2018). Geneva Association.**
https://www.genevaassociation.org/sites/default/files/research-topics-document-type/pdf_public/big_data_and_insurance_-_implications_for_innovation_competition_and_privacy.pdf

xviii **LinkedIn's Most Promising Jobs of 2019. (2019). LinkedIn**
https://blog.linkedin.com/2019/january/10/linkedins-most-promising-jobs-of-2019

xix **Here's When CEOs Expect to See Financial Returns on Artificial Intelligence. (2019). Fortune.**
https://fortune.com/2019/07/02/artificial-intelligence-financial-returns/

xx **Realizing the full value of AI in Insurance | Accenture. (2019). Accenture.com.**
https://www.accenture.com/gb-en/insights/insurance/future-workforce-insurance-survey

xxi **AI adoption advances, but foundational barriers remain. (2019). McKinsey & Company.**
https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain

xxii **Managing Machines: the governance of artificial intelligence. Bank of England. (2019)**
https://www.bankofengland.co.uk/-/media/boe/files/speech/2019/managing-machines-the-governance-of-artificial-intelligence-speech-by-james-proudman

xxiii **See (i)**

xxiv **Artificial intelligence and machine learning in financial services, Financial Stability Board.**
https://www.fsb.org/wp-content/uploads/P011117.pdf

xxv **For example: Machine Learning: The High-Interest Credit Card of Technical Debt. D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Google, Inc.**
https://storage.googleapis.com/pub-tools-public-publication-data/pdf/43146.pdf

xxvi **Big Data and Machine Learning Won't Save Us from Another Financial Crisis. (2018). Harvard Business Review.**
https://hbr.org/2018/09/big-data-and-machine-learning-wont-save-us-from-another-financial-crisis

xxvii **Could a cyber attack cause a systemic impact in the financial sector? (2018). Bank of England**
https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2018/could%20a%20cyber%20attack%20cause%20a%20systemic%20impact%20final%20web

xxviii **Artificial intelligence, financial risk management and systemic risk, London School of Economics. Systemicrisk.ac.uk.**
http://www.systemicrisk.ac.uk/sites/default/files/downloads/publications/SP13.pdf

About the authors

Keyur Patel is an independent journalist and consultant, specialising in economics, technology and financial services. He is a research associate at the CSFI and co-author (and since 2018, principal author) of the Centre's 'Banana Skins' surveys, which assess the greatest risks facing the international banking, insurance, and financial inclusion sectors. He was formerly Marjorie Deane Fellow at the Financial Times, and is a graduate in economics from University College London.

Marshall Lincoln is a data scientist and former AI specialist with a number of firms in Silicon Valley. He specialises in machine learning and applied statistics, with a background in economics, price modelling, and experiment design. He was formerly Director of Business Intelligence at 8x8, a NYSE-listed cloud communications company.

Keyur and Marshall are co-founders of the Lucid Analytics Project, which conducts cross-industry research into the responsible and effective use of AI technologies.

Supporters

The CSFI is an educational charity. It has no endowment income. It receives financial and other support from a wide range of public and private bodies, as well as from individuals. Among the institutions that have provided the Centre with financial support are:

Accenture
Arbuthnot
Citi
City of London
Deloitte
EY
HSBC
JPMorgan
Finance & Leasing Association
London Stock Exchange Group
Moody's
PwC
Royal Bank of Scotland
Ruffer
Swiss Re
Tradeweb

ACCA
Association of British Insurers
Aviva
Bank of England
Bank of Italy
Brunswick Group
Building Societies Association
Euroclear
Eversheds
Financial Conduct Authority
Financial Reporting Council
Fujitsu
Gate One
ICMA
IHS Markit
Investment Association
Japan Post Bank
Link Scheme Holdings Ltd
Jersey Finance
KPMG
Legal & General
LendInvest
Lloyds Banking Group
Meiji Yasuda
Morgan Stanley
Nomura Institute
Orrick
Oxera Consulting
PIMFA
Schroders
Triple Point
UK Finance
Wipro
World Federation of Exchanges

3 Lines of Defence Consulting
Absolute Strategy
AFME
Association of Corporate Treasurers
Bank of Japan
Brigade Electronics
Chartered Banker Institute
C. Hoare & Co.
CISI
CMS
Cognito Media
EBRD
Embassy of Switzerland
Endava
ETF Securities
Euro IRP
EVIA
Fairbanking Foundation
Farrer & Co
Greentarget
Hardman & Co.
HM Treasury
Intrinsic Value Investors
Ipsos MORI
Kreab Gavin Anderson
MacDougall Auctions
Meritus Consultants
Money Advice Service
NM Rothschild
Nutmeg
OMFIF
Raines & Co
Sapience Communications
Skadden, Arps, Slate Meagher & Flom
Taiwan Financial Supervisory Commission
TheCityUK
Zopa
Z/Yen

The CSFI has also received support in kind from, *inter alia*

Allen & Overy
Burges Salmon
Charles Russell Speechlys
Clifford Chance
Dentons
DLA Piper
Financial Times
The London Institute of Banking & Finance
Kemp Little
Linklaters
Norton Rose Fulbright

CSFI
Registered Charity Number 1017352
Registered Office: North House, 198 High Street, Tonbridge, Kent TN9 1BE
Registered in England and Wales limited by guarantee. Number 2788116

UK £30
US $40
EUR €40