The potential impacts of the digital revolution on the operational risk profiles of banks

Received (in revised form): 24th October, 2023

Michael Grimwade
Managing Director, Operational Risk, ICBC Standard Bank, UK

Michael Grimwade first worked in operational risk management almost 30 years ago. He is Head of Operational Risk for ICBC Standard Bank and has previously held senior operational risk management roles at MUFG Securities, Royal Bank of Scotland and Lloyds TSB, and he has also been a director of the Institute of Operational Risk. Michael is the author of a number of papers on quantifying emerging risks, modelling operational risk capital and climate change, and he received an award in 2014 from the Institute of Operational Risk for his contribution to the profession. His second book, ‘Ten Laws of Operational Risk’ was published in December 2021.

Operational Risk, ICBC Standard Bank Plc, 20 Gresham Street, London EC2V 7JE, UK
E-mail: michael_s_grimwade@yahoo.co.uk

Abstract  Society is undergoing a digital revolution. This is altering the business profiles of banks in terms of their systems, processes, controls and usage of third parties; their competitive landscape, ie competition from both digitising incumbents and new BigTech and FinTech entrants; and the behaviours of stakeholders, ranging from customers to cyber-criminals. This digital revolution is amplifying some of their existing risks, while also creating new risks, eg the potential for artificial intelligence (AI) tools to change behaviours over time (AI model drift). Some of these changes in operational risk profile will be transient, as they are associated with digital transformation, while others will be both ongoing and characterised by a high degree of dynamism. In this digitised endstate higher frequency/lower value human errors, may be replaced by lower frequency/higher impact systemic losses, arising from both catastrophic and silent failures. The influence of the digital revolution spans almost all of the Basel operational risk event categories, and may also lead to the enhancement of some controls (eg surveillance), while others may be undermined (eg by voice-spoofing). There is no silver bullet to mitigate these risks; instead, a portfolio of existing control frameworks need to be enhanced, including the following: change management; model risk management; third party vendor management; business continuity management, disaster recovery and operational resilience; and cybersecurity, with new controls required to address the new risks associated with AI. This will be a key factor in the operational risk losses of banks over the next decade.

Keywords: digital revolution, digitisation, artificial intelligence, AI, BigTech, FinTech, operational risk

INTRODUCTION
Society is undergoing a digital revolution which is altering the business profiles of banks, their competitive environments and the behaviours of their stakeholders. This digital revolution is enabled by a series of technological advances, such as the development of the Internet, smart phones and artificial intelligence (AI), that are underpinned by faster, cheaper and bigger processing and storage capabilities. Almost all aspects of banking are changing in response to technology-enabled competition from both incumbents and new entrants: the greater willingness of customers to use a portfolio of financial service providers and shareholder pressures to improve returns. The Organisation for Economic Co-operation and
Development (OECD) observed that ‘Banking is undergoing a transformation from being based in physical branches to using IT and big data’ and that this digital revolution has ‘greatly increased the weight of codified information and the tools that are available to process it, Artificial Intelligence (AI) [...] using big data’.1

This paper is focused on the consequences of this digital revolution on the operational risk profiles of banks, and is organised into the following four sections:

1. Why and how banking is being transformed by the digital revolution.
2. How the digital revolution may alter the operational risk profiles of banks:
   • digital transformation;
   • AI;
   • third and fourth party suppliers;
   • controls — both AI enhancements and degradations;
   • criminal innovation.
3. What actions should operational risk managers be taking to mitigate these risks.

As the risks of the digital revolution are still emerging, the examples in this paper are taken from a wide variety of industries, not just financial services, to illustrate potential consequences.

WHY AND HOW BANKING IS BEING TRANSFORMED BY THE DIGITAL REVOLUTION

Over the course of the author’s career, the business profile of retail banking has moved from branches, which sold products and services and processed cash, cheques and customer requests between the hours of 9 am and 3 pm, five days a week, to online, 24/7 banking, in which payments are increasingly digital and nearly instantaneous. In the UK, this has contributed to more than half of the bank branches that existed in the mid-1980s being closed. In 1986 there were ~21,600 bank and building society branches in the UK, but by 2022 there were just ~8,000.2 The COVID-19 pandemic may have accelerated this existing trend, for instance in the UK cash payments represented ~23 per cent of all payments in 2019, but this had declined to just ~15 per cent by 2021.3 Similarly, investment banking has moved from open outcry and phone trading to increasingly electronic and algorithm-driven trading, with the last open outcry pits on the London International Financial Futures and Options Exchange (Liffe) trading floor closing in November 2000.

The breadth of the impacts of this digital revolution on banking is illustrated in Figure 1, which represents both the internal infrastructure of banks and the key stakeholders with which they interact.

The remainder of this section illustrates the nature of the digital revolution in terms of three of the components in Figure 1, ie BigTech and FinTechs, customers and clients, and the firms themselves.

BigTechs and FinTechs

In addition to providing technology services to banks, such as infrastructure (IaaS) and software-as-a-service (SaaS), BigTechs may also act as both competitors and partners with banks by providing basic financial services to their large networks of customers, and by acting as distribution channels for third party product providers. Payment services have historically been one of the first financial services offered by BigTech. Their offerings can be split into two categories:

1. Payments in emerging economies may involve the introduction of new entities and operations outside of traditional financial and banking networks. For example, Vodafone’s and Safaricom’s M-PESA payment service in East Africa utilises the mobile phone network.
2. In more developed markets, BigTechs’ payment services are overlaid on existing banking services, such as credit and debit cards, eg Apple Pay and PayPal.

BigTechs’ inroads into payments have been greatest where the existing provision of payment services is limited and mobile phone penetration is high.4,5 In the EU, the second Payment Services
Directive (2018) has led to a growth in BigTech firms with licensed payment subsidiaries.6

BigTechs can also leverage their customer bases to offer money market funds and insurance products via their platforms. Money market funds can be linked to their payment services, i.e surplus account balances can be invested in money market funds. This primarily occurs in China, for example Alipay, but even in China the value of the BigTech funds are a fraction of bank deposits. These money market funds and insurance products may be provided by BigTech affiliates or by third party providers.

While some BigTechs have begun to lend (mainly to small and medium-sized enterprises and consumers with short maturities) the expansion into lending has been strongest in those jurisdictions with lighter financial regulation and higher banking sector concentration. Although the access of BigTech to customer data, combined with machine learning, may give them a competitive advantage in predicting default, BigTech lending is ultimately limited by their inability to take retail deposits without obtaining banking licences. Obviously this is not an issue for FinTech peer-to-peer lending platforms, as these allow individuals and companies to lend directly to borrowers, either individually or to packages of loans, without bank intermediation. Examples included LendingClub (US) and Funding Circle (UK), but in recent years both have exited from peer-to-peer lending.

BigTech incursions into financial services regarding payments — and to a lesser extent, lending — as well as the bundling of other bank services may squeeze bank profitability, driving the need for efficiency improvements through new technology. In contrast, FinTechs may similarly impact bank profitability but by unbundling...
individual services traditionally provided by banks, capitalising on an environment in which their customers are increasingly willing to embrace both new technologies and new service providers (as discussed in the next section).

Customers: Changing behaviours
Surveys of customers have found that younger generations (eg millennials and Generation Z [Gen Z]) are less satisfied with their incumbent banks than older generations (eg baby boomers). While all generations have shown a willingness to embrace new technologies and service providers such as FinTechs, the younger generations have a greater tendency to utilise multiple FinTech service providers. This is particularly important as millennials and Gen Z now represent the largest generational demographics in the US, and millennials are currently the largest driver of new net loans, although this will soon pass to Gen Z.

Banks: Improving efficiency and controls
For banks the digital revolution may variously deliver operational efficiencies through greater automation, enhancements to controls and improved data-based decision making.

Operational efficiencies may arise from the automation of low value, repetitive activities, for example, through the use of automated computer programs or bots for activities such as responding to audit letters or correcting errors in fund transfer requests, but at the potential expense of architectural complexity. One example is the introduction by JPMorgan Chase of its COiN (contract intelligence) platform, which uses machine learning to analyse legal documents in order to extract important data. Chatbots now commonly interact with customers to solve simple problems, with machine learning facilitating the identification of customer requirements and, if necessary, transferring customers to human agents. Chatbots may increasingly be used to provide internal help-desk services to staff. In addition, cloud computing may provide banks with processing and storage capabilities that are potentially lower in cost and more flexible.

Enhancements to controls include the harnessing of machine learning to identify suspicious activities (eg anti-money laundering [AML] and fraud detection) and to flag potential cases for review by bank staff. In the future, the current compliance surveillance of traders’ communications — using multi-lingual lexicons of key words — may be replaced by AI that intelligently understands the meaning of language.

Credit approvals for retail customers have long been automated and based on scorecards, while on trading floors, it is beginning to be employed to evaluate venue, timing and order size choices to optimise the filling of orders.

Conclusions
The drivers of this digital revolution seem unstoppable. Technological advances continue apace, and they are being increasingly embraced by customers. Banks that harness these technologies will reap the benefits of operational efficiencies through greater automation, enhancements to controls and improved data-based decision making. Banks that do not embrace these changes are at risk of suffering customer and revenue attrition due to competition from digitised incumbents and new BigTech and FinTech entrants, leading to poorer returns than their digitising peers. Digitisation, however, will clearly alter the operational risk profiles of firms, both during the transition and permanently in the new end-state. This is explored further in the next section.

HOW THE DIGITAL REVOLUTION MAY ALTER THE OPERATIONAL RISK PROFILES OF BANKS
The digital revolution is changing the risk profiles of firms by creating new operational risks as well as amplifying existing risks, ie by making failures potentially more systematic, creating concentrations of data, introducing new single points of failure and fuelling criminal innovation. As noted above, some of these changes in the operational risk profile will be transient, restricted to the duration of a firm’s migration to a more digitised business model, while other changes will be permanent.
The potential impacts of the digital revolution on the operational risk profiles of banks

Table 1 sets out an extract of the Basel II event taxonomy and highlights that six of the Level 1 risks may variously be influenced by five different aspects of the digital revolution, which are causal factors. These are illustrated by examples from a range of industries as well as rules-based and AI-powered algorithms. The remainder of this section considers these five causal factors in more detail:

- digital transformation;
- AI;
- third and fourth party suppliers;
- controls, both AI enhancements and degradation;
- criminal innovation.

Digital transformation

As stated by the OECD, ‘In order to achieve improved efficiency, the incumbents must restructure simultaneously with the entry of new competitors’. The execution of digital transformation may lead to five potential points of operational failure (Figure 2):

1. Stresses on business as usual (BAU) activities, caused by the redirection of resources to a firm’s overall digital transformation programme, leading to increased operational risk losses. The resources being redirected may include both experienced staff members and technology resources needed to provide fixes and enhancements to BAU processes and legacy applications.

2. Uncovering latent issues. The process of digitising and migrating data may uncover historical operational failures which have remained latent for extended periods of time, eg misbookings or deficiencies and/or omissions in legal documentation; digital transformation may accelerate the identification of these previously hidden issues. The decommissioning of legacy infrastructure may also uncover latent issues as well as functionality that has silently accreted over time. Finally, the automation of processes that results in headcount reductions may act as a trigger for staff litigation.

3. Risks of migration/transition to the next state. This includes a range of operational failures including those associated with data quality and data leaks, the functionality and performance of AI, and systems integration. The disruption suffered by TSB in 2018 is illustrative of the potential risks of migration during strategic transformation programmes.

4. Weaknesses associated with any interim-states. While phasing transformation programmes helps to mitigate the risks of migration, new potential risks may arise from running strategic digitised and legacy systems and processes in parallel. This can lead to issues from gaps or overlaps in transactional data and inconsistencies in reference data and conventions, eg the treatment of public holidays or end of day cut-off times etc. Additionally, as customers may increasingly interact with their banks regarding a single complex transaction via multiple channels, eg online, chatbots and call centres, then banks must be able to integrate these information flows between the different channels in any interim-state, as well as in the eventual end-state.

5. Weaknesses in the end-state. The strategic digitised architecture may integrate applications with bots, AI, APIs and partnerships with FinTechs, all of which may run in the cloud. This end-state may act as ‘an amplifier of existing’ risks, eg transforming the high frequency/low value human errors into much lower frequency, but much higher severity risks of systemic technology failures. It may also introduce new risks and vulnerabilities, for example new single points of failure in the form of the cloud and third parties, data privacy issues, a blizzard of AI generated false positives, silent failures over time through AI model drift (eg discriminatory lending), unintended native functionality in third party software and the potential to exclude less technologically sophisticated customers.

While none of these risks associated with change are clearly unique to digital transformation, the potential scale of change and volume of data involved makes them more significant.

The remainder of this section considers the risks associated with the digitised end-state in more detail.
Table 1: The digital revolution: analysis of how operational risks may be influenced by the five causal factors

<table>
<thead>
<tr>
<th>Baseline II, levels</th>
<th>Examples of IT and cyber operational risks, ie the risk of loss arising from . . .</th>
<th>Digital transformation</th>
<th>AI</th>
<th>Third and fourth parties</th>
<th>Controls</th>
<th>Cyber-criminals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Systems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Disruption of software, both accidental or malicious, eg TSB’s botched systems migrations (2018).</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Disruption of data and storage, both accidental or malicious.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Hardware failures, both accidental or malicious, eg Stuxnet virus disrupts Siemen centrifuges (~2007).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Disruption of own infrastructure, both accidental or malicious, eg RBS’s batch-scheduler issues (2012).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Disruption of external infrastructure, both accidental or malicious.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Disruption of vendors and suppliers of IT services, both accidental or malicious, eg SunGard UK’s administration (2022) or the ransomware attack on Ion (2023).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><strong>BDSF</strong></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Transaction capture, execution and maintenance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Various malfunctions, eg:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• System miscalculations.</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Incorrect decisions, eg a blizzard of false positives.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• System generated duplicated or erroneous transactions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>EDPM</strong></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Suitability, disclosure and fiduciary</strong></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Breach of privacy — errors.</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Breach of privacy — exceeds legal authority, eg Facebook’s EUR1.2bn fine (2023).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Breach of contract/intellectual property rights through utilising the outputs of ChatGPT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Inappropriate advice, eg robo-advice litigation in Hong Kong (2019).</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Improper business practices</strong></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Automation of improper trade/market practices, eg abuse of last-look (2015) and Dr Evil (2004).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Automation of improper trade/market practices from AI, eg discriminatory lending</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exploitation of AI through deliberate manipulation by sophisticated third parties, eg chatbot Tay (2016).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exclusion of less technologically sophisticated customers.</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The potential impacts of the digital revolution on the operational risk profiles of banks

Table 1: The digital revolution: analysis of how operational risks may be influenced by the five causal factors (continued)

<table>
<thead>
<tr>
<th>Basel II, levels</th>
<th>Level 3 Examples of IT and cyber operational risks, ie the risk of loss arising from . . .</th>
<th>Digital transformation</th>
<th>AI</th>
<th>Third and fourth parties</th>
<th>Controls</th>
<th>Cyber-criminals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Theft and fraud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber theft of firm's cash or cryptocurrency, eg Bangladesh Bank theft (2016).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber theft of clients' cash or cryptocurrency from firm, eg cryptocurrency theft from BitMart (2021).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber investment frauds, eg by establishing a clone of a bank's website — these are very numerous.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Application fraud, eg using phished data or scraped from the internet by AI.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF</td>
<td>Systems security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber theft of client data from banks, eg the JPMorgan Chase data theft (2014).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber theft of data from banks' suppliers and vendors, eg the Equifax data theft (2017).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber theft of bank customers' data from unrelated third parties, eg the eBay data theft (2014).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber theft of intellectual property, eg AI model code.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cyber extortion from banks, eg CajaGlobal (2023).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>Unauthorised activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Malicious breach of privacy, eg uploading intellectual property into ChatGPT, Samsung (2023).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Malicious destruction of assets.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Theft and fraud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Theft of intellectual property, eg computer code for AI.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Previous thefts have related to HFT code.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Theft of client data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Procurement frauds by technology staff.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPWS</td>
<td>Employee relations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Employee litigation, eg changing roles or redundancy resulting from the digital revolution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Unsuccessful candidates, eg Amazon’s reportedly discriminatory machine learning powered CV screening tool (2014).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

While the Level 1 and 2 risks reflect Basel II’s risk taxonomy, the Level 3 risks are based on an analysis of ∼900 IT and cyber operational risk events. x, increased risk; ✓/x, both enhancing controls, but also degrading controls ✓ enhancing controls.
AI
As Table 1 illustrates, there are a wide range of risks that are influenced by AI and big data, including:

- **data issues**: the malfunctioning of AI as a consequence of data issues, eg errors in training data, incomplete or unrepresentative data, historical data biases, insufficient data, poor data validation, etc.;
- **model issues**: the malfunctioning of AI as a consequence of model issues, eg inappropriate model choices, errors in model design or construction, unexpected behaviours or consequences, AI model drift, etc.;
- **inappropriate usage of AI**, eg to commit market abuse;
- **exploitation of AI**, eg again to commit market abuse;
- **data leakage and data privacy issues**; and
- **intellectual property infringements**.

**AI malfunctions: Data issues**
The quality of data used to train AI is crucial to its intended operation. This is illustrated by a widely reported issue relating to a machine learning tool, developed by Amazon in 2014, to review job applicants’ CVs. The model was trained to assess applications by studying CVs submitted to Amazon over a 10 year period, which, reportedly, led it to learn to favour male candidates for software developer jobs and other technical posts and to downgrade applicants whose CVs contained words such as ‘women’s’ or who were graduates of two all-women’s colleges. In 2015, reportedly Amazon recognised that the tool was not evaluating applicants in a gender-neutral way. Similarly, in 2019, the National Institute of Standards and Technology (NIST) found widespread evidence of racial bias in 189 facial recognition algorithms as there were higher rates of false positives for Asian and African American faces relative to images of Caucasians. A notable exception, however, was that for some algorithms developed in Asian countries there was no such bias, potentially reflecting the relationship between an algorithm’s performance and the data used to train it. In banking, comparable issues may introduce bias into credit scoring tools, which may result in discriminatory lending decisions.

**AI malfunctions: Model issues**
AI models are clearly not infallible, for example there are reported instances of them hallucinating facts such as the case of an Australian mayor who threatened to sue OpenAI if it did not correct ChatGPT’s false claim that he had served time in prison for bribery. While there seem to be no examples (so far) of AI leading to catastrophic failures and operational risk losses for financial institutions, there are examples of losses arising from malfunctioning rules-based trading algorithms, most notably Knight Capital’s US$460m loss which it suffered in just 45 minutes in 2012. There was a spate of these events from 2010, after which there have been no material losses, reflecting...
improvements in controls and increased regulation, eg Markets in Financial Instruments Directive (MiFID) II’s Regulatory Technical Standard-6. These losses may be illustrative of the potential consequences of AI model drift, as the performance of an AI tool may diverge from its original specification over time.

Malfunctioning rules-based algorithms have also been responsible for silent failures over extended periods, for example in 2006 when the Royal Bank Of Canada (RBC) repaid CAD7.2m to 23,000 residential mortgage customers after discovering two errors relating to the calculation of mortgage repayment charges, including a mismatch between RBC’s system and the stated terms between 1999 and 2004. Similar losses could arise from AI model drift. A case was also brought in Hong Kong in 2019 by an investor against a hedge fund manager for US$23m of losses that were claimed to have been caused by an AI-powered investment tool, but the outcome of this litigation, at the time of writing, is unclear.20

**AI: Inappropriate usage**

There is the potential for AI to be used to enable staff members and/or firms to act inappropriately in a much more efficient manner. While there are no current examples relating to AI, there are examples of rules-based algorithms being used in this way. For example, the New York Department of Financial Services (DFS) fined Barclays US$150m in 2015 for allegedly implementing ‘last look’21 in an asymmetrical fashion between 2009 and 2014 on its BARX trading platform, ie FX trades that moved in a customer’s favour, above a threshold, were rejected by BARX, but not if they had moved in Barclays favour.22 Similarly, in 2004, Citigroup leveraged automation to execute its Dr Evil trading strategy. This involved building up and rapidly exiting a very large long position in European government bonds, the equivalent of an average day’s trading volume, in just 18 seconds on the MTS platform. Citigroup was fined by the FSA for failing to consider the impact of their trading strategy on the market.23 In the future, misconduct may also be perpetrated by AI: without guardrails (see section below), an AI may single-mindedly pursue its goal but, in the process, do something that is harmful to markets and/or customers that was not the desired outcome. This is the financial services’ equivalent of the AI-run paperclip factory thought experiment.24

**AI: Exploitation**

The behaviours of AI can be maliciously exploited by third parties. For example, on 24th March, 2016, Microsoft launched a chatbot — Tay — that was designed to have conversations with Twitter users and to learn how to mimic humans by copying their speech patterns. Unfortunately, it quickly learned to repeat anti-Semitic and other hateful language that human Twitter users fed the program, forcing Microsoft to shut Tay down just 16 hours after its launch. Microsoft posted a statement saying:

The AI chatbot Tay is a machine learning project, designed for human engagement. As it learns, some of its responses are inappropriate and indicative of the types of interactions some people are having with it. We’re making some adjustments to Tay.25

While there are no current examples of trading AIs being exploited, there have been examples of rules-based trading algorithms being manipulated by human traders. For example, in 2015 the Financial Conduct Authority (FCA) fined a former Bank of America Merrill Lynch bond trader for market abuse. Knowing that other market participants often used algorithms to update their quotes by tracking the best bid and best offer on BrokerTec, he entered a series of quotes for low trade sizes, which became the best bids, and waited for the algorithms to raise their quotes in response. He then sold to these other banks, generating a profit, before cancelling his own quotes.26

**AI: Data leakage and data privacy issues**

There are a range of risks associated with ChatGPT regarding both data leakage and data privacy. Uploading proprietary data into AI such as ChatGPT may mean that it is used to provide future public responses. For example, it was reported in April 2023 that Samsung software engineers had sent lines of confidential code to ChatGPT on two separate occasions in order to fix bugs.27 As a
consequence, it was reported that a number of Korean companies were putting in place guidelines, training and restrictions to prevent recurrence.\textsuperscript{28} There are also potential issues associated with AI in terms of data privacy, for example on 4th April, 2023, the Italian data protection authority (Garante) issued a temporary ban to OpenAI regarding ChatGPT due to it being originally trained on personal data. The regulator had concerns over the lack of a legal basis for the use of the personal data, as well as issues over its accuracy.\textsuperscript{29}

\textbf{AI: Intellectual property infringements}
Finally, AI that is trained on large quantities of text data, including books and articles, may potentially infringe on the copyright of these works.\textsuperscript{30} For example, Getty Images filed a lawsuit in February 2023 in a Delaware court against Stability AI Inc., accusing it of misusing 12 million Getty photographs to train its Stable Diffusion AI image generation system.\textsuperscript{31}

\textbf{Third and fourth party suppliers}
The increased usage of third parties as part of digitisation, such as BigTech, can alter the operational risk profiles of firms through both disruption of software and/or infrastructure and data associated issues. The Financial Stability Board has highlighted that BigTech could potentially affect financial stability in several ways:\textsuperscript{32}

1. Risks may be magnified by their interlinkages with regulated financial entities, such as partnerships to originate and distribute financial products; and
2. They may also generate risks as they carry out systemically important activities, that are ancillary to financial services, such as, both infrastructure-as-a-service, eg cloud services\textsuperscript{33} and also software-as-a-service, as evidenced by the disruption caused by the ransomware attack on Ion in February 2023, or the potential for disruption caused by SunGard UK, a datacentre operator, entering administration in March 2022.

In addition to these systemic risks, there are also third and fourth party risks associated with the loss of data, either through accidents and/or malicious acts, as well as the breach of data privacy rules, such as restrictions on the cross-border movement of data (eg Facebook’s 2023 EUR1.2bn fine for storage of personal data outside of the EU),\textsuperscript{34} the continued storage and usage of data. Finally, connectivity with providers of SaaS may create contagion risks. Banks may be vulnerable to attack or infection via these links through the introduction of code compromised by malware, such as in the SolarWinds cyber-attack in 2020.

\textbf{Controls: Both AI enhancements and degradation}
AI and machine learning can be used in a variety of ways to improve the effectiveness of controls, but equally these new technologies may be harnessed by cyber-criminals to exploit victims more effectively and efficiently. Applications of AI to improve control include the prevention and/or detection of frauds, money laundering, market misconduct and rogue trading. Firms can use AI to identify suspicious activities and flag potential cases for review by bank staff. AI could be trained using internal data such as historical data preceding the discovery of fraud or the identification of suspicious activity (eg by using suspicious activity report [SAR] data). This may eventually lead to a rather dystopian future in which AI-powered trading algorithms may themselves be subject to monitoring by AI-powered surveillance tools. There is also the potential, as noted earlier, for existing surveillance of communications based on multi-lingual lexicons of key words to be replaced by AI that understands language.

Other uses of AI to enhance controls may include the following.

- \textit{Identification of lessons learned:} AI could interrogate operational risk data held in governance, risk and compliance (GRC) systems, such as, incidents, key risk indicators (KRIs) and risk and control self-assessments (RCSAs), in order to gain new insights.
- \textit{Automation of data collection:} Some elements of the know your customer data collection could be undertaken by AI through scraping the Internet, although this could lead to data privacy issues.
The potential impacts of the digital revolution on the operational risk profiles of banks

~20 billion digital devices will need to be upgraded or replaced with post-quantum cryptography in the next 20 years. This is reminiscent of the Y2K issue at the end of the last millennium.

Criminal innovation

Professional criminals are rationale, seeking to maximise their returns, and will therefore optimise, like foraging animals, the effort to locate victims, the effort to exploit victims, and the financial rewards obtained. Consequently, any changes in customer and/or bank behaviours (effort to locate), usage of technology by both banks and criminals (effort to exploit), or concentrations of data through digitisation (financial rewards) will lead to predictable changes in criminal behaviours. For example, text generating AI may prove to be both more efficient and effective at phishing, social engineering, encouraging people to click on links infected with malware, or to make fraudulent urgent payments in response to messages purporting to come from their relatives, friends or colleagues. This will effectively amplify existing cyber-crime risks.

As the world has digitised, there has been an increase in the occurrence of data hacks, and since 2014, there have been a series of very significant data losses involving more than 500 million records (Figure 3). The targets are often social media (eg Facebook and LinkedIn), technology (eg Yahoo and SolarWinds), retail (eg eBay), outsource service providers (eg Equifax) or governments, because 'that’s where the data is'. While only one of the firms in Figure 3 is a global systemically important bank (G-SIB), JPMorgan Chase (reflecting a

Figure 3: Distribution of data thefts of more than 50 million records in the public domain over the last 15 years
comforting level of control) the data of G-SIBs will increasingly be vulnerable through their third and fourth party relationships. For example, the customer data of multiple G-SIBs may have been compromised through the data theft from Equifax, the consumer credit agency, in 2017.

Digitisation has enabled bank robberies to be committed remotely. The first publicly disclosed Society for Worldwide Interbank Financial Telecommunication (SWIFT) cyber-payment theft seems to have taken place in 2015. The Bangladesh Bank theft in 2016 led to the transfer of five payments totalling US$101m. SWIFT and its roughly 11,000 members responded to these thefts with control enhancements. As a result, cyber-criminals evolved their techniques; after early 2018, the average fraudulent transaction amounts reduced by between $10×$ and $100×$ to between US$0.25m and US$2m.42

In the future, however, criminal innovation using AI or quantum computing may lead to a new spike in these thefts, before industry counter-measures may again lead to a decline (Figure 4).

A adoption of cryptocurrencies as payment mechanisms seem quite limited. Thefts of cryptocurrency primarily from exchanges (eg the theft of ~US$200m from BitMart in 2021), a few trading platforms and, more recently, decentralised finance companies, however, do not seem to be declining in the same way that SWIFT cyber-payment thefts have declined based on details of thefts in the public domain.

In summary, AI and digitisation are creating new opportunities for cyber-criminals through the creation of concentrations of data and single points of failure, new ways of locating and exploiting victims more efficiently and the ability to circumvent bank controls, both now and in the future.

**Conclusions**

In advanced economies, incumbents have largely fought off the twin threats of BigTech and FinTech through digitisation, but this can increase the risks of disruption both during digital transformation and from the subsequent increased dependencies on third and fourth parties.

The new digitised end-state, through the automation of human processes, may lead to the replacement of random human errors (higher frequency but lower value losses) with systemic automated errors (lower frequency but higher value losses) arising from both catastrophic and silent failures. AI may also have the potential to be harnessed by either firms and/or their staff members to automate misconduct.

BigTech, by providing both software and IaaS, is creating new interconnectivity and dependencies which may lead to systemic disruption. BigTech may also lead to significant data leakage and breaches through its movement of data across borders.

AI may allow the enhancement of preventive and detective controls, for example, for monitoring of
The potential impacts of the digital revolution on the operational risk profiles of banks

financial crime, market misconduct and rogue trading by better and more efficiently identifying unusual patterns of behaviour. Cyber-criminals may, however, also harness these new technologies to circumvent banks’ controls by using AI to produce deep-fakes such as voice-spoofing and quantum computing to break existing encryption. Cyber-criminals may also exploit customer and bank behaviour changes to optimise the proceeds from their crimes.

In aggregate, the digital revolution represents a very significant challenge to the operational risk management profession. Consequently, the next section considers the actions that they should be taking.

WHAT ACTIONS SHOULD OPERATIONAL RISK MANAGERS BE TAKING TO MITIGATE THESE RISKS

In the short-term, the digital revolution may drive a transient spike in operational risk events associated with the execution of digital transformation programmes as technology is advancing more rapidly than the control frameworks of banks through the usage of AI by staff members (eg the use of ChatGPT by several of Samsung’s software engineers) and cyber-criminals (eg the use of voice-spoofing). Thereafter, the operational risk profiles of banks will continue to be dynamic as a result of the ongoing emergence of new technologies (eg quantum computing), the potential for AI model drift and changing stakeholder behaviours (Figure 1). There are no silver bullets for mitigating these risks (as seen in Table 1) and consequently, this section provides an overview of the portfolio of existing operational risk management control frameworks that need to be enhanced in the short and medium-terms to meet these threats.

Digital transformation

The not unreasonable expectation of the EU’s Digital Operational Resilience Act (DORA) (2022) is that firms will appropriately manage ‘each major change in the network and information system infrastructure’.44 In practice this may involve:

- stresses on BAU: monitoring of a dashboard of metrics, highlighting trends in KRIs and KCIs focused on signs of stretch in a bank’s BAU activities, eg lags in the performance of regular tasks, and growing back-logs in completing remedial actions to enhance controls;
- risks of migration: standard first line mitigations of migration/transition risks include governance, documenting business and functional requirements, project plans, RAID logs (risks, actions, issues and dependencies), non-functional and functional testing, migration run books, dress rehearsals and contingency planning. All of these activities reduce rather than eliminate the likelihood of failure;
- weaknesses in interim and end-states: risk assessments need to be undertaken on the interim and end-states to identify any new vulnerabilities that have been introduced and the effectiveness of work-arounds for any non-delivered functionality. When implementing vendor technology, firms must additionally focus not just on the functionality that they specified but also native functionality which may expose them to incremental operational risks.

As the second line cannot review all change initiatives, it needs to prioritise projects with hard migration dates, tight timescales and significant complexity and scale. Any ‘Red’ rated projects, should receive additional attention due to the temptation for project managers and sponsors to descope functionality, reduce the scope of testing, run formally sequential project tasks in parallel and re-rate the significance of bugs etc. Ultimately, the second line should critically review and challenge the first line’s assumptions, assessments and activities regarding major change initiatives, calling out the inconvenient truths that are likely hiding in plain sight.

AI, including control enhancements

While AI has the potential to enhance the control frameworks and the efficiency of banks, they can clearly amplify existing risks and expose firms to new risks. Consequently, the use of AI, both informally by staff members and as part of digital
transformation, clearly needs to be formally governed. This involves firms articulating acceptable and unacceptable usage, defining the risk assessments that need to be undertaken and the approval processes, which should reflect the Prudential Regulation Authority’s (PRA) recent Supervisory Statement on Model Risk Management. Model validation is clearly a key component of any approval process of both internally developed and third party models and should variously include:

- **data quality validation** aimed at identifying incomplete, unrepresentative, biased and erroneous training data, eg to identify any biased/discriminatory lending;
- **outcome monitoring** against a benchmark or a non-AI model to ensure that the AI tool is equitable and free from bias;
- **‘black box’ testing**, which involves a developer experimenting with the model by feeding it different data inputs to better understand how the model makes its responses;
- **red-teaming**: The red-teamers attack the model, in an attempt to get it to do something inappropriate, as occurred with chatbot Tay.

While rigorous monitoring needs to be undertaken of any decision making tools, additional safeguards are required for AI due to the potential for model drift. These safeguards may include the following:

- **Alerts**: detective controls to flag unusual or unexpected actions to employees, to mitigate silent failures.
- **Human in the loop**: decisions are only executed after review and approval by a human.
- **Guardrails**: automatic termination of the AI if it produces undesired outputs. This is intended to mitigate model drift, which can occur with AI algorithms that are self-teaching.
- **Kill-switch**: to terminate rapidly if an AI algorithm is catastrophically malfunctioning.

While some of these controls are extensions of existing model/rules-based algo controls, others are new.

---

**Third and fourth party suppliers**

Both the EU’s DORA and the UK’s Discussion Paper on critical third parties envisage a role for regulatory oversight of critical technology providers. In practice, even when firms have rights to access, inspect or audit a critical third party, banks cannot monitor their controls on a real-time basis. As a consequence, there needs to be increasing focus on business continuity management, disaster recovery and operational resilience, ie on contingency planning as to how firms will maintain important/critical business services for their customers if these critical suppliers suffer disruption as well as the processes for severing and restoring connectivity, for example, once the third party has recovered from a cyber-attack. DORA envisages that these plans are tested at least annually to ensure their effectiveness. Firms should also periodically review the data that they hold with third parties to ensure that it is minimised (ie that superfluous data is not being held with third or fourth parties) in terms of both volume and age. While firms should also have exit strategies for their third and fourth parties, these must identify the timescales for exit, for example, the duration of a migration from one datacentre operator to another would be measured in months, if not years, meaning that firms must have adequate redundancy/contingency because exit is not a viable short-term contingency arrangement. The move towards more regulatory oversight of critical technology providers is reflective in that the existing concentration in service provision, ie disruption, could potentially be the source of a future banking crisis.

**Criminal innovation and control degradation**

Ongoing criminal innovation (see Figures 3 and 4) will require banks to be perpetually vigilant and to be sharing data to identify new forms of attack, eg voice-spoofing. This requires banks to be dynamically assessing themselves against industry cyber-security standards, such as NIST, and identifying and mitigating new attack vectors and future threats through implementing quantum-safe encryption. DORA encourages banks to enhance their digital operational resilience through sharing
information on cyber threats, therefore slowing the ability of a new threat to spread, supporting defence capabilities, threat detection techniques, mitigation strategies or response and recovery stages.

Anecdotally, the speed of these changes feels like it is accelerating. This requires banks to be both very vigilant of the changes happening around them and extremely agile in their decision making. All of these activities will require additional digitally-skilled resources to be operating in first, second and third lines of defence, although in the short-term demand for these digital resources will outstrip supply.

CONCLUSION
The digital revolution is unstoppable and will transform banking (Figure 1) and as a consequence the operational risk profiles of banks (Table 1). There is no silver bullet to mitigate these risks; instead, a portfolio of existing control frameworks needs to be enhanced — ie change management, model risk management, third party vendor management, business continuity management, disaster recovery and operational resilience and cyber security — to address both the amplification of existing risks and also new risks associated with AI. The digital revolution may be a key driver of the operational risk losses of banks in the next decade, and the concentration of critical technology providers could potentially be the source of a future banking crisis.

AUTHOR’S NOTE
The contents of this paper are the author’s own views rather than those of ICBC Standard Bank.

References and notes
3 Ibid.
7 Deloitte’s Center for Financial Services March 2022 digital banking survey of 3,000 US consumers, weighted inline with the banking population found that millennials and Gen Z were respectively only 62 per cent and 57 per cent satisfied or very satisfied with their banks, and 28 per cent and 20 per cent respectively were somewhat or very likely to switch whilst for baby boomers and older generations the comparable data was 81 per cent and 85 per cent and 6 per cent and 1 per cent respectively. Valenti, J. and Alderman, R. (7th September, 2021) ‘Building on the Digital Banking Momentum’, Deloitte Insights, Issue 30, Summer 2022, available at https://www2.deloitte.com/xe/en/insights/industry/financial-services/digitalization-in-banking.html (accessed 5th November, 2023).
8 McKinsey & Co found that 40 per cent of US financial decision makers have a FinTech Account, and that younger FinTech users tend to use multiple accounts, whereas older generations are more likely to use just one FinTech account. Gen Z are the most frequent users of FinTech with 29 per cent having more than one account.


11 OECD, ref 1 above.


16 Ibid.


21 Last look enables firms to ensure that high frequency traders have not detected a move in a market a few milliseconds before the bank, enabling them to arbitrage the bank’s marginally less nimble trading platforms. This is achieved by imposing a hold period between the receipt of a customer’s order and its acceptance and execution.


The potential impacts of the digital revolution on the operational risk profiles of banks


32 Financial Stability Board, ref 6 above.

33 A Bank of England survey in 2020 estimated that more than 70 per cent of the 30 banks surveyed and 80 per cent of the 27 insurers surveyed, relied on just two cloud providers.


40 The Depression-era American bank robber Willie Sutton, when asked why he robbed banks replied ‘because that’s where the money is!’


43 Adapted from Grimwade, ref 13 above.


Copyright of Journal of Risk Management in Financial Institutions is the property of Henry Stewart Publications LLP and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.