Deepfakes and financial fraud: what next for the UK financial sector?

24 April 2024

As the global AI arms race intensifies, should financial institutions in the UK be concerned about fraudsters' recent gains? Answer: yes. Beware "deepfakes".

AI is attracting a great deal of publicity from commentators keen to raise questions about how to mitigate all manner of legal risks for developers, corporate AI-users, and consumers. Intellectual property, data security and competition lawyers are poring over the data used to train AI, while public and regulatory lawyers are tracking legislative change and governmental plans to allocate responsibility for AI implementation across a range of industry sectors.

However, in this article we focus on the implications of a risk that has already arrived: the risk that bad actors capitalise on the opportunity to leverage recent advancements in generative AI to perpetrate financial fraud.

What is generative AI and what do we mean by "deepfake"?

Defining generative AI is one of the big issues facing legislators and commentators alike. With the speed of technological advancements, the definition of today may be out of date by tomorrow. While we do not seek to add our voice to that debate, we should clarify that when we say "generative AI" in this article we are using it as a collective term for technology which learns the patterns available in underlying training data (e.g. text and image content) and uses it to create new content with similar characteristics.

One of the most well-publicised recent trends in the use of generative AI is the generation of "deepfakes", which are highly realistic imitations of humans usually used to trick an audience into thinking they are watching or listening to an individual who is known to them (and therefore trusted).

In May 2023, Reuters reported that the President of Microsoft, Brad Smith, described deepfakes as his biggest concern around AI, and it is easy to see why. The risk types are vast and varied. For example, with important elections looming in both the UK and the US this year or early next, there is a focus in the mainstream media on possible implications for election fraud. But here we seek to unpick some of the risks associated with financial fraud.

What are the fraudsters up to?

Recent and rapid developments in generative AI are making impersonation fraud much easier to achieve. Forged images are not new, but generative AI is making it easier and quicker to disseminate increasingly convincing fakes on a widescale and replicable basis with less and less technical knowledge required by the user of the technology. There is now an abundance of examples of the use of deepfakes

to further illegal objectives, and we can expect to hear about them with increasing frequency over the year ahead.

Stephen Fry warned in autumn last year that convincing deepfake videos of celebrities would be coming after a computer-generated fake of his voice was used by a historical documentary maker to narrate a piece it released without his knowledge or consent.

As Brad Smith and Stephen Fry were issuing these warnings, the use of deepfake video techniques had already been deployed in the context of a high profile financial scam when an image of financial journalist and founder of MoneySavingExpert.com Martin Lewis was used, without his knowledge, in the highly realistic video promotion of a "*great investment opportunity*" connected with a fake project involving Elon Musk.

In recent weeks, deepfake video soundbites from chef Nigella Lawson, broadcaster Piers Morgan and talk show host Oprah Winfrey have been published by a US social media influencer seeking to raise publicity in the marketing of their self-help book.

Does it work?

Yes, the convincing nature of generative AI content is already yielding results for bad actors. One cautionary tale, reported in February by CNN, saw a finance employee at a multinational firm being tricked into paying out US$25m to fraudsters who digitally posed as their boss, the CFO of the company, during a video conference call. This is an example of what's commonly referred to as APP (or Authorised Push Payment) fraud, where a person authorises a payment and the bank acts on that person's instructions, only to discover subsequently that the individual authorising the payment has been duped.

It's no surprise that deepfakes are spreading beyond the world of celebrity. Granted, executives at multinational firms, like celebrities, have an online (video) presence greater than the average UK resident, but the technology is only going to become more powerful. Less and less input data will be needed to run the more sophisticated models, and it has become increasingly common, since the COVID pandemic, for individuals' images and voices to be broadcast over videoconferencing platforms at virtual seminars and talks, many of which are video-recorded and widely disseminated. Increased computing power and increased opportunities to harvest training footage will no doubt prove to be a dangerous combination for the future.

Financial scams designed to trick the recipient into taking action, using false information to imitate real life, have been around for years. It's all too common to see emails popping up with urgent warnings about needing to transfer money quickly to secure it or presenting links to websites containing malicious software.

But as society is adapting to become more sceptical of such communications, fraudsters are adopting ever more convincing AI-enhanced ways to deceive us. Five years ago, in 2019, an AI-generated voice deepfake was used to place a phone call which convinced the CEO of a UK-based energy firm to transfer around £243,000 to the bank account of a supposed Hungarian supplier, having thought the voice he was speaking with belonged to his boss, the CEO of his company's German parent company. The US$25m Hong Kong example referred to above, from February of this year, brings into focus how developments in technology make this type of fraud even more likely. In that example the duped employee was not just

the recipient of a forged email, or a participant on a two-party faked phone call, but was part of a multi-person video conference where it turned out that everyone that they saw was fake. We have now entered an era in which we do not only have to read and listen carefully, but also have to look closely. Seeing and hearing an interactive video of the person or people we're expecting to hear from is not as conclusive as it once was.

How will this be reflected in the statistics?

According to UK Finance, 40% of all fraud losses in the UK in 2022 resulted from APP fraud. That figure encompasses a number of different types of deception, including: false investments; advance fee requests; and CEO and other impersonation frauds. Deepfakes (as well as other forms of generative AI disinformation techniques) have the potential to be deployed across all these recognised forms of APP fraud.

UK Finance's most recent annual report also reflects that, in the second half of 2022, although only 18% of scams took place by telephone, compared to 78% of scams originating online, the typical value of a telephone scam was much higher (accounting for 44% of total fraud losses, compared to 36% of losses from online fraud). This demonstrates that people are more likely to place greater trust in people that are speaking to them, not just spamming their email inbox. As deepfake video content spreads, it does not take much of a leap to predict that the proportion of frauds resulting from (video) calls will have risen significantly by the end of 2024.

What does this mean for financial institutions?

There has already been an onus on banks to reimburse clients for losses they suffer from APP fraud for some time, so AI-driven increases in certain types of APP fraud are set to increase that burden. Since May 2019, a voluntary code has been in place by which many large payment service providers (PSPs) have been voluntarily reimbursing victims of APP fraud in circumstances where the victims are consumers or certain other smaller enterprises that have satisfied the conditions under the code.

According to UK Finance, of the £485.2m total value of APP frauds in the UK in 2022, an aggregate total of £285.6m (i.e. almost 60%) was reimbursed to the victims, either through a direct refund from the victim's bank or through recovery of funds from the beneficiary account.

From 7 October 2024, it will become mandatory for regulated PSPs participating in the Faster Payments scheme (FPS) to reimburse victims of APP fraud if certain basic criteria are met (and, at present, subject to a proposed cap of £415,000 per claim). (See our previous article here). This move was originally planned to go live earlier in the year, before it became clear further time was required to get the financial industry ready for the change. The Payment Systems Regulator (PSR) now ultimately considers the October date to be "*an ambitious but feasible date*" and that "*the protection of APP scams victims must be prioritised*." Fines can be imposed on regulated businesses not meeting that new obligation, up to 10% of their turnover, a serious sanction indicative of a hard line being taken by the PSR. An equivalent reimbursement requirement for APP fraud is also planned for PSPs participating in CHAPS, with the same implementation date of 7 October 2024 (see the PSR's latest Work Plan, published in April 2024).

Outside of the regulatory regimes, the position is perhaps a bit less ominous for payment services so far as APP fraud is concerned, following the decision of the UK Supreme Court in *Philipp (Respondent) v Barclays Bank UK Plc (Appellant)* [2023] UKSC 25 last year. (See our previous article here). This decision confirmed that the so-called "Quincecare duty" – a duty on a bank to refrain from executing a payment instruction from its customer when it is "put on inquiry" that the order is an attempt to defraud the customer – does not arise in the case of an individual customer instructing their bank to make a payment where that customer is the victim of APP fraud (though a more limited "retrieval duty" may continue to exist as recently considered in *CCP Graduate School Limited v National Westminster Bank PLC & Anor* [2024] EWHC 581 (KB)).

Nonetheless, there is no doubt that the confluence of the regulatory regime and external pressures to reimburse victims of fraud, with the increasing APP risk created by generative AI, creates something of a pinch point for UK payment services.

What other AI-related legal changes are occurring?

It's clear that the UK Government intends to continue to take what it calls a "pro-innovation" approach to the regulation of AI in the UK by declining the legislation-led approach we are seeing in the US and EU, instead inviting industry sector regulators to formulate their own strategies. In March 2023, the Government published a white paper outlining what it perceives to be five sector-agnostic principles to be implemented by regulators: (i) safety, security and robustness; (ii) appropriate transparency and explainability; (iii) fairness; (iv) accountability and governance; and (v) contestability and redress.

On 15 February 2024, following the publication the week before of the outcomes from the consultation on the AI white paper, the Government sent letters to various UK regulators, including the FCA and the Bank of England (BoE), requiring an update on their strategic approach to dealing with advancements in AI to be published by 30 April 2024. In these letters the Government recognises that, as the use of AI becomes more widespread across the economy, "*transparency regarding the steps regulators are taking to understand both the opportunities and the risks this creates, and the actions they are taking in response, would be valuable*".

The FCA and BoE have released their respective updates this week, showing proactive intent ahead of the end of April deadline. (See our full article on this here). The FCA has made clear that it will remain a "*technology-agnostic, principles-based and outcomes-focused regulator*", as has the PRA via the BoE's announcement, and the FCA says there will be *"close scrutiny of the systems and processes firms have in place to ensure our regulatory expectations are met*". The FCA commented that AI "*has already transformed the speed with which we can monitor and tackle scam websites, money laundering and sanctions breaches*" and has confirmed that it has recruited over 75 data scientists of its own. It is also committed to the continued provision of synthetic data for testing of AI tools through Digital Sandbox initiatives.

The financial services sector will also watch with interest the implementation of another significant legal change hoped to help cut off the supply of deepfakes and other AI-driven deception techniques at source: the Online Safety Act (OSA). Enacted in October 2023, the OSA provides OFCOM with a blueprint with which to introduce industry standards and codes of practice for tech firms, as well as empowering it to

call on them for evidence of various kinds, including most relevantly the operation of protections against fraudulent advertising.

The OSA sees the introduction of key duties for tech firms in this area, following much campaigning from the likes of Martin Lewis, as well as input from numerous financial institutions. All platforms must "*use proportionate measures relating to the design or operation of the service*" and "*operate the service using proportionate systems and processes designed*" to prevent users from accessing fraudulent content, minimise the time for which such content is present on the service, and swiftly take such content down if the provider becomes aware of it. But the OSA does not create direct liability for tech firms towards victims if a fraud does take place.

These key aspects of the OSA overlap with recent efforts made by the FCA itself to crack down on unauthorised financial promotions. Restrictions on the publishing of financial promotions in the UK have been around for some time, but the FCA has recently been proactive in seeking to collaborate with big players in the tech industry associated with the platforms used by fraudsters looking to make use of generative AI, and has recently published finalised guidance on social media financial promotions. (See our previous article here). Having been collaborating with OFCOM through the Digital Regulation Cooperation Forum (or 'DRCF'), the FCA again highlights the importance of such collaboration in this week's AI update, which is echoed by the PRA in the BoE's simultaneous update.

In a case study published by the FCA on 8 February 2024, the FCA advertised that its "*close engagement with Google, Bing (Microsoft), Meta, X/Twitter and TikTok led to them changing their policies to only permit 'paid for ads' for financial services, including investments, that have been approved by an authorised person*". The FCA noted that, since Google introduced its policy in this area, "*we have seen close to a 100% reduction in illegal paid-for ads*".

So what steps can participants in the UK financial markets be taking in the meantime in response to these recent AI-developments?

In the first place, financial institutions will need to keep an eye on the first line of defence, and develop their own systems to combat the direct threats to them. Voice recognition is used by some as an identity verification step, and evidently deepfakes make that riskier. But institutions will be investing in their own technologies to keep up with the pace of change. There lies the possible saving grace. Though the technology underlying deepfakes creates risks, wider technological developments may create opportunities. For example, sophisticated pattern-recognition capabilities that can be used to test customers' behaviour and detect fraudulent activities. Indeed, subject to restrictions in use, generative AI might be its own remedy, and organisations will need to find ways of leveraging new technologies to fight fire with fire. Reaction to the FCA and BoE Government-mandated AI updates will be particularly interesting in this context.

Beyond this, financial institutions' efforts may be best focused on collaboration and education. Collaborating with industry peers, tech companies, legislative bodies, law enforcement agencies, and regulatory bodies will provide opportunities to learn, share intelligence, develop best practices, identify emerging threats, and ultimately foster a collective defence against fraudsters. The education of customers and employees about the risks of generative AI-enabled fraud and the raising of awareness about the recognition and avoidance of AI-driven frauds will be collectively highly valuable. This is all

certainly a great challenge in a competitive market where user experience is key and customers are typically resistant to anything which might constitute an “impediment” to a smooth and quick payment process.

If you would like to discuss any of the issues raised in this article, please get in touch with one of the contacts listed.

Authored by William Robinson and Andrew Holland.