COMMENTS OF SUSAN VON STRUENSEE, JD, MPH
to the
Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence,
Including Machine Learning
86 FR 16837-38 (March 31, 2021)
Agency/Docket Numbers:
Docket ID OCC-2020-0049
Docket No. OP-1743
Docket No. CFPB-2021-0004
Docket No. NCUA-2021-0023

Thank you for the opportunity to help shape an updated framework designed to address the constantly evolving space of Artificial Intelligence (AI). I urge the financial regulators that AI Ethics is a meaningful activity that provides real oversight and produces useful information and impact for the public. I recommend a triage system to identify the AI risks that pose the greatest risks to individuals and prioritize reviewing those risks and solutions first, so as to discourage lax procedures, minimize harm and prioritize readiness over after-the-fact review.

The use of AI by financial institutions is not new, from using AI to analyze contracts and financial statements, to oversight of traders, to chatbots to communicate with customers, AI and machine learning provide financial institutions efficient opportunities to analyze large volumes of data and improve customer service, among other benefits. Financial technologies are becoming an integral part of all types of financial services: lending, payments and remittances, savings, investment, insurance, etc. They transform business models and enhance their customer focus. Both large financial organisations, such as banks, and specialist fintech companies offering a narrow range of services have introduced various fintech solutions. Such technological transformation of the financial market requires that the regulator should revise its approaches. As AI use cases evolve, financial institutions will face new opportunities and challenges to meet regulatory expectations with respect to safety and soundness and consumer protection. In the introduction to the RFI, the regulators highlight three areas that could pose risk management challenges: (1) explainability—how the AI uses inputs to produce outputs; (2) data usage—including potential bias or limitations in data; and (3) dynamic updating—the ability to validate AI use cases in a constantly changing environment.

In response, this comment particularly emphasizes the issues of AI ethics and deep fakes (especially for third-party oversight: the challenges institutions face in using AI developed or provided by third parties) to the attention of the Comptroller of the Currency, the Federal Reserve System, the Federal Deposit Insurance Corporation, the Consumer Financial Protection Bureau, and the National Credit Union Administration, in response to the Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning.

Biometrics – working in concert with a combination of artificial intelligence (AI) and machine learning (ML) to scan, analyze and then to create what could be a varied biometric identity database capable of verifying and storing fingerprints, facial features, even voice and device data – could allow for not only tougher, more meticulous identity security, but also a deeper understanding of a financial institute’s customer profile – giving banks and other fintech a truer way to “know your customer” preventing identity theft (and money laundering). Thus deep fake technology to prevent the hijacking of biometric data is needed.

Regulatory and policy developments reflect a global tipping point toward serious regulation of artificial intelligence ("AI") in the U.S. and European Union ("EU"), with far-reaching consequences for technology companies and government agencies. In late April 2021, the EU released its long-anticipated draft regulation for the use of AI, banning some "unacceptable" uses altogether and mandating strict guardrails such as documentary "proof" of safety and human oversight to ensure AI technology is "trustworthy."

The U.S. federal government's national AI strategy continues to take shape, bridging the old and new administrations. Pursuant to the National AI Initiative Act of 2020, which was passed on January 1 as part of the National Defense Authorization Act of 2021 ("NDAA"), the White House Office of Science and Technology Policy ("OSTP") formally established the National AI Initiative Office (the "Office") on January 12, 2021. The Office—one of several new federal offices mandated by the NDAA—will be responsible for overseeing and implementing a national AI strategy and acting as a central hub for coordination and collaboration by federal agencies and outside stakeholders across government, industry and academia in AI research and policymaking.

The National Defense Authorization Act of 2019 created a 15-member National Security Commission on Artificial Intelligence ("NSCAI"), and directed that the NSCAI "review and advise on the competitiveness of the United States in artificial intelligence, machine learning, and other associated technologies, including matters related to national security, defense, public-private partnerships, and investments."

On March 1, 2021, the NSCAI submitted its Final Report to Congress and to the President. At the outset, the report makes an urgent call to action, warning that the U.S. government is presently not sufficiently organized or resourced to compete successfully with other nations with respect to emerging technologies, nor prepared to defend against AI-enabled threats or to rapidly adopt AI applications for national security purposes. Against that backdrop, the report outlines a strategy to get the United States "AI-ready" by 2025. The Commission explains:

The United States should invest what it takes to maintain its innovation leadership, to responsibly use AI to defend free people and free societies, and to advance the frontiers of science for the benefit of all humanity. AI is going to reorganize the world. The National Security Commission on Artificial Intelligence, The Final Report (March 1, 2021), available at **https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.**

Artificial Intelligence

Another example of new technology that does not fit neatly into the existing data privacy framework is artificial intelligence and machine learning. This includes robo-advisors and robo-solutions where systems are automated and the owner of the platform (i.e., the technology firm or financial institution) does not have interaction-by-interaction control over the system.

Technology firms are offering up these types of platforms to financial institutions to use in test environments and in some cases with dummy customer data. It is unclear in these cases who would be the data controller and who would be the data processor. While robo-advisors or robo-solutions typically require user input of personal information in order to generate results, should a user decide not to proceed, data may be deleted, and therefore, may not technically be used or controlled by the entity that offers the platform. This makes it problematic for the parties involved with the platform to determine what their obligations are to users under data privacy rules. AI solutions are also challenging one of the founding notions of most transparency requirements under data privacy regimes: telling users what their data will be used for. The core benefit of AI is that it may have the ability to perform tasks and offer products and services to a customer that are perhaps not contemplated by a human being tasked with the same role. Does this mean that the AI is using personal data in a way that is not technically articulated as the purpose for which data was collected in the relevant privacy notice? Create a fiduciary duty for data brokers. There is a profound yet relatively easy to implement step to address this manipulation. The G20 and other governments could make their AI Principles practical by extending the regulatory requirements they have for doctors, teachers, lawyers, government agencies, and others who collect and act on individuals' intimate data to also apply to data aggregators and their related AI implementations. Any actor who collects intimate data about an individual should be required to act on, share, or sell this data only if it is consistent with that person's interests. This would force alignment of the interests of the target/consumer/user and the firm in the position to manipulate. Without any market pressures, data brokers who hold intimate knowledge of individuals need to be held to a fiduciary-like standard of care for how their data may be used. This would make data brokers responsible for how their products and services were used to possibly undermine individual interests.

Promoting Innovation

The US National Security Commission on Artificial Intelligence made clear-promote AI innovation.

The Center for Long-Term Cybersecurity (CLTC) has published a new report, A New Era for Credit Scoring: Financial Inclusion, Data Security, and Privacy Protection in the Age of Digital Lending, that examines the trade-offs associated with digital lending platforms in India. By providing small loans to consumers through their mobile phones, lending apps have broadened access to credit for low-income borrowers. But they have also introduced new threats to fairness, privacy, and digital security, as lenders use an array of personal data — including age, location, and even personal contacts — to gauge an individual's willingness and ability to pay.

“Digital lenders in India can (and do) use data points as far-ranging as individuals' GPS location history and phone contacts as proxies for financial responsibility,” wrote the report's author, Tarunima Prabhakar, who undertook her study while serving as a research fellow with CLTC. “These personal details are also in some cases leveraged in debt recovery, as some lenders contact borrowers' friends and families to pressure them to repay debts. Alternative lending enables people to access credit, but with far fewer safeguards than traditional banks would provide.”

To develop the paper, Prabhakar analyzed industry reports, academic papers, and government publications, and she conducted interviews with stakeholders in the emerging “fintech” sector, including data scientists and venture capitalists. She also used text-mining techniques to analyze a corpus of 70,000 comments from lending app users, revealing that many of those who sign up to receive loans are unaware how their mobile phone data will be used.

Prabhakar's paper compares the emerging digital lending industry in India with the example of the United States, where financial lending is regulated through laws such as the Fair Credit Reporting Act and Equal Credit Opportunity Act, which prevent lenders from making loan decisions based on factors like race and gender. Discrimination is a major concern with lending apps that rely on "alternative data," particularly as decisions may be made by algorithms with limited transparency or accountability.

"The rise of digital lending — and more specifically, alternative credit scoring in India —provides a useful framework for considering the social and ethical consequences of algorithmic decision-making more broadly, and highlights trade-offs that governments and institutions must consider in weighing factors such as privacy and fairness against access to credit and other social goods," Prabhakar wrote.

The report aims to help inform policymakers and financial industry leaders around the world who may be confronted with new risks and opportunities as the fintech sector evolves."Lenders (and their regulators) should leverage the power of technology to expand access to credit, but should be wary of enabling lenders to violate borrowers' privacy or allowing potentially discriminatory practices," Prabhakar wrote. "The example of India highlights how, in an emerging economy with relatively weak institutions and low financial literacy, credit scoring through alternate data creates the possibility for rapid progress in financial inclusion — but under weaker consumer protection standards. The constant threat of exposure of consumer information adds to the challenge, and there is as yet no silver bullet that can enhance financial inclusion without a significant decline in consumer privacy and transparency in lending decisions." See https://cltc.berkeley.edu/wp-content/uploads/2020/06/A_New_Era_for_Credit_Scoring.pdf

The OECD Principles on Artificial Intelligence promote artificial intelligence (AI) that is innovative and trustworthy and that respects human rights and democratic values. They were adopted in May 2019 by OECD member countries when they approved the OECD Council Recommendation on Artificial Intelligence. The OECD AI Principles are the first such principles signed up to by governments.

The OECD AI Principles set standards for AI that are practical and flexible enough to stand the test of time in a rapidly evolving field. They complement existing OECD standards in areas such as privacy, digital security risk management and responsible business conduct.

In June 2019, the G20 adopted human-centred AI Principles that draw from the OECD AI Principles. https://www.mofa.go.jp/files/000486596.pdf

The OSTP/OMB Guidance on Regulation of Artificial Intelligence Applications was signed on November 17, 2020. https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

1. AI ETHICS

With voice recognition and other AI technologies being used for financial transactions, please think multiple steps ahead to ensure an expertise in AI ethics and deep fakes. Users should be given an opportunity as much as possible to "opt out" of AI functions.

Sample steps in order to ensure that AI ethics are taken seriously:

- Hire ethicists who work with corporate decisionmakers and software developers
- Develop a code of AI ethics that lays out how various issues will be handled
- Have an AI review board that regularly addresses corporate ethical questions
- Develop AI audit trails that show how various coding decisions have been made
- Implement AI training programs so staff operationalizes ethical considerations in their daily work, and

- Provide a means for remediation when AI solutions inflict harm or damages on people or organizations.

Several companies have joined together to form the Partnership for Artificial Intelligence to Benefit People and Society. They include Google, Microsoft, Amazon, Facebook, Apple, and IBM. It seeks to develop industry best practices to guide AI development with the goal of promoting "ethics, fairness and inclusivity; transparency, privacy, and interoperability; collaboration between people and AI systems; and the trustworthiness, reliability and robustness of the technology."

It is not easy to resolve any of the ethical issues surrounding the topics discussed above. Each of them raises important ethical, legal, and political concerns, and therefore are not amenable to easy resolution. Leaders dealing with these challenges will have to take considerable time and energy to work through the substantive issues.

But there are organizational and procedural mechanisms that help with some of these ethical dilemmas. Having clear processes and avenues for deliberation would help deal with particular problems. There are a number of steps that would help firms ensure fair, safe, and transparent AI applications.

As William Galston of Brookings suggests, if these reforms prove inadequate, there may need to be government legislation to mandate appropriate safeguards. Improving protections in the areas of racial bias and discrimination are especially important. In addition, resolving how the United States wants to handle technology such as for deep fakes is crucial.

1. Hiring company ethicists
It is important for companies to have respected ethicists on their staffs to help them think through the ethics of AI development and deployment. Giving these individuals a seat at the table will help to ensure that ethics are taken seriously and appropriate deliberations take place when ethical dilemmas arise, which is likely to happen on a regular basis. In addition, they can assist corporate leadership in creating an AI ethics culture and supporting corporate social responsibility within their organizations. These ethicists should make annual reports to their corporate boards outlining the issues they have addressed during the preceding year and how they resolved ethical aspects of those decisions.

2. Having an AI code of ethics
Companies should have a formal code of ethics that lays out their principles, processes, and ways of handling ethical aspects of AI development. Those codes should be made public on the firm's websites so that stakeholders and external parties can see how the company thinks about ethical issues and the choices its leaders have made in dealing with emerging technologies.

3. Instituting AI review boards
Businesses should set up internal AI review boards that evaluate product lines and are integrated into company decisionmaking. These boards should include a representative cross-section of firm stakeholders and be consulted on AI-related decisions. Their portfolio should include development of particular product lines, the procurement of government contracts, and procedures used in developing AI products.

4. Requiring AI audit trails
Companies should have AI audit trails that explain how particular algorithms were put together or what kinds of choices were made during the development process. This can provide some degree of "after-the-fact" transparency and explainability to outside parties. Such tools would be especially relevant in

cases that end up under litigation and need to be elucidated to judges or juries in case of consumer harm. Since product liability law is likely to be the governing force in adjudicating AI harm, and it is necessary to have audit trails that provide both external transparency and explainability.

5. Implementing AI training programs
Firms should have AI training programs that not only address the technical aspects of development, but the ethical, legal, or societal ramifications. That would help software developers understand that they are not merely acting on their own individual values, but are part of a broader society with a stake in AI development. AI goes beyond the development of traditional product lines with narrow social implications. With its potential to distort basic human values, it is crucial to train people in how to think about AI.

6. Providing a means of remediation for AI damages or harm
There should be a means of remediation in case AI deployment results in consumer damages or harm. This could be through legal cases, arbitration, or some other negotiated process. This would allow those hurt by AI to address the problems and rectify the situation. Having clear procedures in place will help when disasters strike or there are unanticipated consequences of emerging technologies.

1.a EXPLAINABILITY

"Another issue is explainability. Here, explainability is a term used to discuss the problem that with neural networks: we don't always know which feature or which dataset influenced the AI decision or prediction, one way or the other. This can make it very hard to explain an AI's decision, to understand why it might be reaching a wrong decision. This can matter a great deal when predictions and decisions have consequential implications that may affect lives for example when AI is used in criminal justice situations or lending applications....
Recently, we've seen new techniques to get at the explainability challenge emerge. One promising technique is the use of Local-Interpretable-Model Agnostic Explanations, or LIME. LIME tries to identify which particular data sets a trained model relies on most to make a prediction. Another promising technique is the use of Generalized Additive Models, or GAMs. These use single feature models additively and therefore limit interactions between features, and so changes in predictions cane be determined as features are added.

Yet another area we should think about more is the "detection problem," which is where we might find it very hard to even detect when there's malicious use of an AI system—which could be anything from a terrorist to a criminal situation. With other weapons systems, like nuclear weapons, we have fairly robust detection systems. It's hard to set off a nuclear explosion in the world without anybody knowing because you have seismic tests, radioactivity monitoring, and other things. With AI systems, not so much, which leads to an important question: How do we even know when an AI system is being deployed?

There are several critical questions like this that still need a fair amount of technical work, where we must make progress, instead of everybody just running away and focusing on the upsides of applications for business and economic benefits.

The silver lining of all this is that groups and entities are emerging and starting to work on many of these challenges. A great example is the Partnership on AI. If you look at the agenda for the Partnership, you'll see a lot of these questions are being examined, about bias, about safety, and about *these kinds of existential threat questions. Another great example is the work that Sam Altman, Jack Clarke and others at OpenAI are doing, which aims to make sure all of society benefits from AI. "* Quote of JAMES MANYIKA, CHAIRMAN AND DIRECTOR OF MCKINSEY GLOBAL INSTITUTE, [1] in Martin Ford, Architects of Intelligence: The Truth about AI from the People Building it, (2018) (consists of conversations with the most prominent research scientists and entrepreneurs working in the field of artificial intelligence, including Demis Hassabis, Geoffrey Hinton, Ray Kurzweil, Yann LeCun, Yoshua Bengio, Nick Bostrom, Fei-Fei Li, Rodney Brooks, Andrew Ng, Stuart J. Russell and many others. The conversations recorded in the book delve into the future of artificial intelligence, the path to human-level AI (or artificial general intelligence), and the risks associated with progress in AI.)

1. Diversify XAI objectives. Explainability techniques are currently developed and incorporated by machine learning engineers, and not surprisingly, their needs (and companies' desire to avoid legal trouble) are being prioritized.Realizing a broader set of XAI objectives will require both greater awareness of their existence and a shift in incentives for accomplishing them. XAI standards and policy guidelines should explicitly include the needs of users, stakeholders, and impacted communities to incentivize this shift. Explainability case studies are one pedagogical tool that can help practitioners and educators understand and develop more holistic explainability strategies. Diverse organizational guidance documents, recommendations, and high-level frameworks can also help guide an organizations' executives and/or developers through key questions to support explainability that is useful and relevant to different stakeholders. Consider, for example, the different needs of developers and users in making an AI system explainable. A developer might use Google's What-If Tool to review complex dashboards that provide visualizations of a model's performance in different hypothetical situations, analyze the importance of different data features, and test different conceptions of fairness. Users, on the other hand, may prefer something more targeted. In a credit scoring system, it might be as simple as informing a user which factors, such as a late payment, led to a deduction of points. Different users and scenarios will call for different outputs.

2. Establish XAI metrics. While there has been some work done to evaluate AI explanations, most attempts are either computationally expensive or only focus on a small subset of what constitutes a "good explanation" and fail to capture other dimensions. Measuring effectiveness more holistically likely requires combining a comprehensive overview of XAI approaches, a review of the different forms of opacity, and the development of standardized metrics. In particular, the evaluation of explanations will need to take into account the specific contexts, needs, and norms in a given case, and use both quantitative and qualitative measures. Further work in this space will help hold organizations accountable and promote successful AI deployment.
Mitigate risks. Explainability entails risks. Explanations may be misleading, deceptive, or be exploited by nefarious actors. Explanations can pose privacy risks, as they can be used to infer information about the model or training data. Explainability may also make it easier for proprietary models to be replicated, opening up research to competitors. Methods for both documenting and mitigating these risks are needed and emerging standards and policy guidelines should include practical measures to do so. For some high-stakes decisions, it may be better to forgo deep learning models and the need for explainability techniques.

3. Prioritize user needs. So far,explainability has primarily served the interests of AI developers and companies by helping to debug and improve AI systems, rather than opening them to oversight or making the systems understandable to users. Prioritizing user needs has received some research

[1] James is a senior partner at McKinsey and Chairman of the McKinsey Global Institute, researching global economic and technology trends.

attention, but user needs remain neglected. Key considerations in providing explanations to users include understanding the context of an explanation, communicating uncertainty associated with model predictions, and enabling user interaction with the explanation. Other user concerns include design practices for user experiences and accessibility. The field might incorporate decades of experience from the theory of risk communication. For example, this roadmap for risk communication with users developed by the Center for Long-Term Cybersecurity provides insights into the needs for two-way communication, accessible choice architecture, and protection for whistleblowers, among other mechanisms that help promote user interests.
https://cltc.berkeley.edu/wp-content/uploads/2020/12/Designing_Risk_Communications.pdf

“ The charade of consent has made it obvious that notice-and-choice has become meaningless. For many AI applications … it will become utterly impossible.” Cam Kerry
https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/

4. Explainability isn’t enough. Although explainability may be necessary to achieve trust in AI models, it is unlikely to be sufficient. Simply having a better understanding of how a biased AI model arrived at a result will do little to achieve trust. When students in England recently learned that they had been assigned standardized test scores based on a simple algorithm that had ascribed weight to schools’ historic performance and, thus, advantaged rich schools, they were outraged, sparking protests in cities around the country. Explainability will only result in trust alongside testing, evaluation, and accountability measures that go the extra step to not only uncover, but also mitigate exposed problems.

5. Explainability is seen as a central pillar of trustworthy AI because, in an ideal world, it provides understanding about how a model behaves and where its use is appropriate. The prevalence of bias and vulnerabilities in AI models means that trust is unwarranted without sufficient understanding of how a system works. Currently, there is a significant discrepancy between the vision of explainability as a principle that reaches across domains and works for diverse stakeholders, and how it is being incorporated in practice. Bridging that gap requires greater transparency about the goals being optimized, and further work to ensure those goals align with the needs of users and the benefit of society at large.

1.b DEEP FAKES

FinTech (Financial Technologies) is the provision of financial services using innovative technologies such as big data, artificial intelligence and machine learning, robotisation, blockchain, cloud technologies, biometrics, etc.

The Federal Reserve's "Guidance on Model Risk Management" (SR Letter 11-7) highlights the importance to safety and soundness of embedding critical analysis throughout the development, implementation, and use of models, which include complex algorithms like AI. The Financial Stability Board highlighted four areas where AI could impact banking. First, customer-facing uses could combine expanded consumer data sets with new algorithms to assess credit quality or price insurance policies. And chatbots could provide help and even financial advice to consumers, saving them the waiting time to speak with a live operator. Second, there is the potential for strengthening back-office operations, such as advanced models for capital optimization, model risk management, stress testing, and market impact analysis. Third, AI approaches could be applied to trading and investment strategies, from identifying new signals on price movements to using past trading behavior to anticipate a client's next order. Finally, there are likely to be AI advancements in compliance and risk mitigation by banks. AI solutions are already being used by some firms in areas like fraud detection, capital optimization, and portfolio management. Governor Lael Brainard, What Are We Learning about Artificial Intelligence in Financial Services?, November 13, 2018, https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm#f27

Deep fakes have become a buzzword discussed widely among legal and technology experts. The term ‘deep fakes’ refers to face-swapping technologies that enable a quick creation of fake images or videos which appear incredibly realistic. The technologies behind the creation of deep fakes include four categories of deep fakes (deep fake porn, deep fakes in political campaigns, deep fakes for commercial uses and creative deep fakes). We need to address ethical and regulatory aspects each of those four categories of deep fakes. Deep fakes are likely to be more widely adopted in the future, and there are various social and legal challenges which the regulators will have to face. in addition to public and private legal measures, market-driven solutions would be most desirable. Companies running content dissemination platforms have the necessary technical expertise and sufficient resources to develop deep fake detecting technologies. Some deep fake technology companies already indicated that they are aware of ethical duties associated with their businesses. From a technology point of view, it may be argued that deep-fake detecting technologies could lead to the “race-to-the-bottom” situation where the technology used to generate deep fakes is as sophisticated as the technology used to detect them. Therefore, one could suggest that deep fake technology should locked-in within certain government agencies and content dissemination platforms. On the other hand, it may be argued that deep fake detecting technologies should develop as an open source: this could ascertain that there is a common shared standard and that such commonly shared standard is more capable in dealing deep fakes.   See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3497144

Harmful lies are nothing new. But the ability to distort reality has taken an exponential leap forward with “deep fake” technology. This capability makes it possible to create audio and video of real people saying and doing things they never said or did. Machine learning techniques are escalating the technology’s sophistication, making deep fakes ever more realistic and increasingly resistant to detection. Deep-fake technology has characteristics that enable rapid and widespread diffusion, putting it into the hands of both sophisticated and unsophisticated actors. While deep-fake technology will bring with it certain benefits, it also will introduce many harms. The marketplace of ideas already suffers from truth decay as our networked information environment interacts in toxic ways with our cognitive biases. Deep fakes will exacerbate this problem significantly. Individuals and businesses will face novel forms of exploitation, intimidation, and personal sabotage. The risks to our democracy and to national security are profound as well. In-depth assessments of the causes and consequences of this disruptive technological change, and to explore the existing and potential tools for responding to it. A broad array of responses need study, including: the role of technological solutions; criminal penalties, civil liability, and regulatory action; military and covert-action responses; economic sanctions; and market developments. Solutions need to be assessed in law and policy along with the pitfalls embedded in various solutions. Chesney, Robert and Citron, Danielle Keats, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security (July 14, 2018). 107 California Law Review 1753 (2019), U of Texas Law, Public Law Research Paper No. 692, U of Maryland Legal Studies Research Paper No. 2018-21, Available at SSRN: https://ssrn.com/abstract=3213954 or http://dx.doi.org/10.2139/ssrn.3213954

Consider the points in this article at https://www.mofo.com/resources/insights/210210-ai-machine-learning-manipulation.html:

The Collision of AI’s Machine Learning and Manipulation: Deepfake Litigation Risks to Companies from a Product Liability, Privacy, and Cyber Standpoint

Benjamin S. Kagel, Erin M. Bosman, Christine E. Lyon

AI and machine-learning advances have made it possible to produce fake videos and photos that seem real, commonly known as “deepfakes.” Deepfake content is exploding in popularity.[i] After *Star Wars: The Rise of Skywalker* used visual effects with historical footage to create a visage of Carrie Fischer on the screen, fans generated competing deepfake videos through artificial intelligence models. Using thousands of hours of interviews with Salvador Dali, the Dali Museum of Florida created an interactive exhibit featuring the artist.[ii] For *Game of Thrones* fans miffed over plot holes in the season finale, Jon Snow can be seen profusely apologizing in a deepfake video that looks all too real.[iii]

Deepfake technology—how does it work? From a technical perspective, deepfakes (also referred to as synthetic media) are made from artificial intelligence and machine-learning models trained on data sets of real photos or videos. These trained algorithms then produce altered media that looks and sounds just like the real deal. Behind the scenes, generative adversarial networks (GANs) power deepfake creation.[iv] With GANs, two AI algorithms are pitted against one another: one creates the forgery while the other tries to detect it, teaching itself along the way. The more data is fed into GANs, the more believable the deepfake will be. Researchers at academic institutions such as MIT, Carnegie Mellon, and Stanford University, as well as large Fortune 500 corporations, are experimenting with deepfake technology.[v] Yet deepfakes are not solely the province of technical universities or AI product development groups. Anybody with an internet connection can download publicly available deepfake software and crank out content.[vi]

Deepfake risks and abuse. Deepfakes are not always fun and games. Deepfake videos can phish employees to reveal credentials or confidential information, e-commerce platforms may face deepfake circumvention of authentication technologies for purposes of fraud, and intellectual property owners may find their properties featured in videos without authorization. For consumer-facing online platforms, certain actors may attempt to leverage deepfakes to spread misinformation. Another well-documented and unfortunate abuse of deepfake technology is for purposes of revenge pornography.[vii]

In response, online platforms and consumer-facing companies have begun enforcing limitations on the use of deepfake media. Twitter, for example, announced a new policy within the last year to prohibit users from sharing “synthetic or manipulated media that are likely to cause harm.” Per its policy, Twitter reserves the right to apply a label or warning to Tweets containing such media.[viii] Reddit also updated its policies to ban content that “impersonates individuals or entities in a misleading or deceptive manner” (while still permitting satire and parody).[ix] Others have followed. Yet social media and online platforms are not the only industries concerned with deepfakes. Companies across industry sectors, including financial and healthcare, face growing rates of identity theft and imposter scams in government services, online shopping, and credit bureaus as deepfake media proliferates.[x]

Deepfake legal claims and litigation risks. We are seeing legal claims and litigation relating to deepfakes across multiple vectors:

1. Claims brought by those who object to their appearance in deepfakes. Victims of deepfake media sometimes pursue tort law claims for false light, invasion of privacy, defamation, and intentional infliction of emotional distress. At a high level, these overlapping tort claims typically require the person harmed by the deepfake to prove that the deepfake creator published something that gives a false or misleading impression of the subject person in a manner that (a) damages the subject’s reputation, (b) would be highly offensive to a reasonable person, or (c) causes mental anguish or suffering. As more companies begin to implement countermeasures, the lack of sufficient safeguards against misleading deepfakes may give rise to a negligence claim. Companies could face negligence claims for failure to detect deepfakes, either alongside the deepfake creator or alone if the creator is unknown or unreachable.

2. Product liability issues related to deepfakes on platforms. Section 230 of the Communications Decency Act shields online companies from claims arising from user content published on the company's platform or website. The law typically bars defamation and similar tort claims. But e-commerce companies can also use Section 230 to dismiss product liability and breach of warranty claims where the underlying allegations focus on a third-party seller's representation (such as a product description or express warranty). Businesses sued for product liability or other tort claims should look to assert Section 230 immunity as a defense where the alleged harm stems from a deepfake video posted by a user. Note, however, the immunity may be lost where the host platform performs editorial functions with respect to the published content at issue. As a result, it is important for businesses to implement clear policies addressing harmful deepfake videos that broadly apply to all users and avoid wading into influencing a specific user's content.

3. Claims from consumers who suffer account compromise due to deepfakes. Multiple claims may arise where cyber criminals leverage deepfakes to compromise consumer credentials for various financial, online service, or other accounts. The California Consumer Privacy Act (CCPA), for instance, provides consumers with a private right of action to bring claims against businesses that violate the "duty to implement and maintain reasonable security procedures and practices."[xi] Plaintiffs may also bring claims for negligence, invasion of privacy claims under common law or certain state constitutions, and state unfair competition or false advertising statutes (e.g., California's Unfair Competition Law and Consumers Legal Remedies Act).

4. Claims available to platforms enforcing Terms of Use prohibitions of certain kinds of deepfakes. Online content platforms may be able to enforce prohibitions on abusive or malicious deepfakes through claims involving breach of contract and potential violations of the Computer Fraud and Abuse Act (CFAA), among others. These claims may turn on nuanced issues around what conduct constitutes exceeding authorized access under the CFAA, or Terms of Use assent and enforceability of particular provisions.

5. Claims related to state statutes limiting deepfakes. As malicious deepfakes proliferate, several states such as California, Texas, and Virginia have enacted statutes prohibiting their use to interfere with elections or criminalizing pornographic deepfake revenge video distribution.[xii] More such statutes are pending.

Practical tips for companies managing deepfake risks. While every company and situation is unique, companies dealing with deepfakes on their platforms, or as a potential threat vector for information security attacks, can consider several practical avenues to manage risks:

•**Terms of Use development:** Companies can consider a variety of approaches to incorporating acceptable usage boundaries on their platforms, including the following:

•Craft Terms of Use with specific guidelines that define the scope of deepfakes, such as prohibiting synthetic or manipulated media that violates any applicable law or could cause harm to an individual.

•Update Terms of Use to capture deepfakes as a violation and provide enforcement mechanisms against users (e.g., removal procedures and account suspension or bans).

•Monitor how regulatory bodies and peer companies define and enforce violations to determine what constitutes a harmful deepfake and ensure Terms of Use are consistently applied.

•Consider whether to rely on Terms of Use to remove only reported violations or whether to implement a policy to proactively monitor and remove violations. Carefully weigh the potential risks for each option; failing to follow through on a policy creates exposure as well.

•**Technical approaches**: Using a fire fighting fire approach, companies can leverage AI-powered detection algorithms, such as facial recognition technology (subject to applicable legal requirements), to detect manipulated media. Implementing multifactor authentication and deploying behavior analytics technologies (which may also be AI-driven) can guard against account takeover. Keeping up with technical advancements as well as security alerts from law enforcement and government agencies can reduce exposure to class action litigation or regulatory enforcement associated with deepfakes.

•**Law enforcement cooperation:** Companies can serve their users and their own interests through proactive outreach and cooperation with law enforcement on deepfake issues. Law enforcement officials at various agencies have increasingly engaged the private sector to combat malicious threat actors that leverage deepfakes for criminal ends.

•**Civil enforcement.** Companies can develop an enforcement program involving a spectrum of action, including account restrictions, pre-litigation outreach, and appropriate escalation to civil litigation as needed. Such enforcement programs can help to reduce and deter online platform abuse.

While the future of deepfakes is uncertain, it is apparent that the underlying AI and machine-learning technology is very real and here to stay—presenting both risks and opportunity for organizations across industries.

---

[i] https://www.cnet.com/pictures/26-deepfakes-that-will-freak-you-out/.

[ii] https://www.youtube.com/watch?v=mPtcU9VmIIE&feature=emb_title.

[iii] https://www.youtube.com/watch?v=4GdWD0yxvqw&feature=emb_logo.

[iv] https://wiki.pathmind.com/generative-adversarial-network-gan.

[v] https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/.

[vi] https://arxiv.org/abs/2005.05535.

[vii] https://en.wikipedia.org/wiki/Deepfake_pornography.

[viii] https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html.

[ix] https://www.reddithelp.com/hc/en-us/articles/360043075032.

[x] https://www.ft.com/content/8a5fa5b2-6aac-41cf-aa52-5d0b90c41840; https://www.biometricupdate.com/202002/advancing-facial-technology-to-fight-identity-fraud-through-liveness-detection.

[xi] Cal. Civ. Code § 1798.150.

[xii] https://jolt.law.harvard.edu/digest/manipulated-media-examining-californias-deepfake-bill; https://www.jdsupra.com/legalnews/texas-law-could-signal-more-state-37742/; https://www.bbc.com/news/technology-48839758.

2. Fair Lending, Bias in Data, Overfitting

Fair lending: The ability to evaluate compliance of AI-based credit decisions with fair lending laws; risk of bias or discriminatory impact of AI; application of model risk management principles; challenges of performing fair lending risk assessments on AI; and ability to identify reasons for adverse actions from AI decisions.

Bias in data (including raw and alternative data): How do institutions manage risks related to data quality and data processing, and are there specific AI use cases where alternative data are effective.

Overfitting (where the algorithm learns from data that is under representative): How do institutions manage the risks of overfitting.

2.1 Focus on fair lending:

The largest set of questions in the RFI relate to fair lending considerations associated with the use of AI, including challenges that institutions face in evaluating bias and discriminatory impact on protected groups, and the potential limitations of model risk management principles in making those determinations. The RFI also seeks information on how institutions comply with requirements under the Equal Credit Opportunity Act and its implementing regulation, Regulation B, to notify consumers of the reason(s) for taking an adverse action on a credit application where the reason for a decision made by an AI-powered decision engine may not be transparent. The latter has been of particular interest to the CFPB. In June 2020, the CFPB reminded institutions of the “regulatory uncertainty” in this space, and encouraged institutions to use the Bureau’s trial disclosure program, no action letter program, and compliance assistance sandbox to potentially address the regulatory uncertainty associated with AI and adverse action notice requirements. Patrice Alexander Ficklin, Tom Pahl, and Paul Watkins, CFPB Blog, Innovation spotlight: Providing adverse action notices when using AI/ML models (July 7, 2020), available at https://www.consumerfinance.gov/about-us/blog/innovation-spotlight-providing-adverse-action-notices-when-using-ai-ml-models/.

Although regulators continue to express concerns about fair lending risks associated with AI, they have yet to articulate expectations for financial institutions’ use of these technologies or how they will be evaluated in the future. See also Federal Deposit Insurance Corporation, FIL-82-2019, Interagency Statement on the Use of Alternative Data in Credit Underwriting (Dec. 13, 2019), available at https://www.fdic.gov/news/financial-institution-letters/2019/fil19082.pdf.

If an AI or machine learning model has a disparate impact on a prohibited basis, what documentation will regulators accept to demonstrate that the model is supported by a legally sufficient business justification? https://www.mayerbrown.com/en/perspectives-events/publications/2021/04/rfi-on-financial-institutions-use-of-ai-provides-opportunity-to-shape-future-regulatory-framework#_edn2

Sian Townson, PhD., a Director in Oliver Wyman's Data and Analytics practice, writes that many financial institutions are turning to AI reverse past discrimination in lending, and to foster a more inclusive economy. But many lenders find that artificial-intelligence-based engines exhibit many of the same biases as humans. How can they address the issue to ensure that biases of the past are not baked into algorithms and credit decisions going forward? The key lies in building AI-driven systems designed to encourage less historic accuracy, but greater equity. That means training and testing AI systems not merely on loans or mortgages issued in the past, but instead on how the money should have been lent in a more equitable world. Armed with a deeper awareness of bias lurking in the data and with objectives that reflect both financial and social goals, we can develop AI models that do well and that do good.

As banks increasingly deploy artificial intelligence tools to make credit decisions, they are having to revisit an unwelcome fact about the practice of lending: Historically, it has been riddled with biases against protected characteristics, such as race, gender, and sexual orientation. Such biases are evident in institutions' choices in terms of who gets credit and on what terms. In this context, relying on algorithms to make credit decisions instead of deferring to human judgment seems like an obvious fix. What machines lack in warmth, they surely make up for in objectivity, right?

Sadly, what's true in theory has not been borne out in practice. Lenders often find that artificial-intelligence-based engines exhibit many of the same biases as humans. They've often been fed on a diet of biased credit decision data, drawn from decades of inequities in housing and lending markets. Left unchecked, they threaten to perpetuate prejudice in financial decisions and extend the world's wealth gaps.

AI and Equality

Designing systems that are fair for all**.**

The problem of bias is an endemic one, affecting financial services start-ups and incumbents alike. A landmark 2018 study conducted at UC Berkeley found that even though fintech algorithms charge minority borrowers 40% less on average than face-to-face lenders, they still assign extra mortgage interest to borrowers who are members of protected classes. http://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf

Recently, Singapore, the United Kingdom, and some European countries issued guidelines requiring firms to promote fairness in their use of AI, including in lending. Many aspects of fairness in lending are legally regulated in the United States, but banks still have to make some choices in terms of which metrics for fairness should be prioritized or de-prioritized and how they should approach it.

So how can financial institutions turning to AI reverse past discrimination and, instead, foster a more inclusive economy? In our work with financial services companies, we find the key lies in building AI-driven systems designed to encourage less historic accuracy but greater equity. That means training and testing them not merely on the loans or mortgages issued in the past, but instead on how the money should have been lent in a more equitable world.

The trouble is that humans often cannot detect the unfairness that exists in the massive data sets that machine-learning systems analyze. So lenders increasingly rely on AI to identify, predict, and remove the biases against protected classes that are inadvertently baked into algorithms.

Here’s how, according to Towson:

Remove bias from data before a model is built.

An intuitive way to remove bias from a credit decision is to strip discrimination from the data before the model is created. But this requires more adjustment than simply removing data variables that clearly suggest gender or ethnicity, as previous bias has effects that ripple throughout. For example, samples of loan data for women are usually smaller because, proportionally, financial institutions have approved fewer and smaller loans to women in decades past than to men with equivalent credit scores and income. This leads to more frequent errors and false inferences for the under-represented and differentially treated female applicants. Manual interventions to attempt to correct the bias in data can also end up in self-fulfilling prophecies, as mistakes or assumptions made may be repeated and amplified.

To avoid this, banks can now use AI to spot and correct patterns of historic discrimination against women in raw data, compensating for changes over time by deliberately altering this data to give an artificial, more equitable probability of approval. For example, by using AI, one lender discovered that, historically, women would need to earn 30% more than men on average for equivalent-sized loans to be approved. It used AI to retroactively balance the data that went into developing and testing its AI-driven credit decision model by shifting the female distribution, moving the proportion of loans previously made to women to be closer to the same amount as for men with an equivalent risk profile, while retaining the relative ranking. As a result of the fairer representation of how loan decisions should have been made, the algorithm developed was able to approve loans more in line with how the bank wished to extend credit more equitably in the future.

Pick better goals for models that discriminate.

Yet even after data is adjusted, banks can often need an extra layer of defense to prevent bias, or remaining traces of its effects, from creeping in. To achieve this, they “regularize” an algorithm so that it aims not just to fit historical data, but also to score well on some measure of fairness. They do this by including an extra parameter that penalizes the model if it treats protected classes differently.
For example, one bank discovered by applying AI that very young and very old applicants were not getting equal access to credit. To encourage fairer credit decisions, the bank designed a model that required its algorithm to minimize an unfairness score. The score was based on the gap between outcomes for people in different age brackets with the same risk profile, including intersections between subgroups, such as older women. By taking this approach, the final AI-driven model could close the mathematical gap between how similar people from different groups are treated by 20%.

Introduce an AI-driven adversary.

Even after correcting the data and regularizing the model, it is still possible to have an apparently neutral model which continues to have a disparate impact on protected and non-protected classes. So many financial institutions go one more step and build an additional, so-called “adversarial” AI-driven model to see if it can predict protected-class bias in decisions made by the first model. If the adversarial challenger successfully detects any protected characteristic such as race, ethnicity, religion, gender, sexuality, disability, marital status or age, from the way the first credit model treats an applicant, then the original model is corrected.

For example, adversarial AI-driven models can often detect ethnic minority zip codes from the outputs of a proposed credit model. This can often be due to a confounding interaction with lower salaries being associated with overlapping zip codes. Indeed, we have seen adversarial models show that an original model is likely to offer lower limits to applications from zip codes associated with an ethnic minority, even if the original model or data available did not have race or ethnicity as an input to check against.

In the past, these issues would have been dealt with by attempting to manually change the original model's parameters. But now we can use AI as an automated approach to re-tune the model to increase the influence of variables which contribute to equity and reduce those that contribute to bias, partially by aggregating segments, until the challenger model is no longer able to predict ethnicity by using zip codes as a proxy. In one instance, this resulted in a model that still differentiated between zip codes but reduced the mortgage approval rate gap for some ethnicities by as much as 70%.

To be sure, financial institutions should lend wisely, based on whether people are willing and able to pay debt. But lenders must not treat people differently if they have similar risk profiles, whether that decision is made by artificial neural networks or by human brains. Reducing bias is not just a socially responsible pursuit — it also makes for more profitable business. The early movers in reducing bias through AI will have a real competitive advantage on top of doing their moral duty.

Algorithms can't tell us which definitions of fairness to use or which groups to protect. Left to their own devices, machine-learning systems may cement the very biases we want them to eliminate.

But AI need not go unchecked. Armed with a deeper awareness of bias lurking in the data and with objectives that reflect both financial and social goals, we can develop models that do well and that do good.

There is measurable evidence that lending decisions based on machine-learning systems vetted and adjusted by the steps outlined above are fairer than those made previously by people. One decision at a time, these systems are forging a more financially equitable world. AI Can Make Bank Loans More Fair https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair

2.2 Explainability: How do institutions manage AI explainability risks, the types of post-hoc methods institutions use to evaluate conceptual soundness, and the types of use cases that present particular explainability challenges.

Precursor to additional oversight: The tone of the RFI reflects regulators' concerns about the explainability and auditability of AI systems, including how regulators will test these systems going forward. International financial regulators have been seeking information about financial institutions' use of AI for years, and some regulators have already begun issuing guidance and scrutinizing AI use cases. For example, in 2019 the Hong Kong Monetary Authority published two sets of risk management guidelines for the use of big data and AI. Hong Kong Monetary Authority, Consumer Protection in respect of Use of Big Data Analytics and Artificial Intelligence by Authorized Institutions (Nov. 5, 2019), available at https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2019/20191105e1.pdf; Hong Kong Monetary Authority, High-level Principles on Artificial Intelligence (Nov. 1, 2019), available at https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2019/20191101e1.pdf.

Among other things, these principles encourage financial institutions to maintain audit logs associated with the design of AI, provide avenues for consumers to request information about decisions made by AI applications, and ensure that AI models produce "objective, consistent, ethical and fair outcomes to customers." The Monetary Authority of Singapore ("MAS") published a set of principles governing the use of AI and data analytics in Singapore's financial sector. Monetary Authority of Singapore, Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector (Nov. 12, 2018), available at https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20

MAS has been partnering with financial institutions to test its "FEAT" (fairness, ethics, accountability, and transparency) principles against actual AI use cases through Project Veritas. This public-private partnership resulted in the development of open source metrics to help financial institutions test fairness in the use of AI for credit risk scoring and consumer marketing. Monetary Authority of Singapore, Veritas Initiative Addresses Implementation Challenges in the Responsible Use of Artificial Intelligence and Data Analytics (Jan. 6, 2021), available at https://www.mas.gov.sg/news/media-releases/2021/veritas-initiative-addresses-implementation-challenges.

The UK Information Commissioner's Office has proposed an AI auditing framework that its investigations teams will use when evaluating the compliance of organizations using AI, and has encouraged entities to use this framework to audit their own AI systems. Information Commissioner's Office, Guidance on the AI auditing framework: Draft guidance for consultation (Feb. 14, 2020), available at https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

Although it is not clear whether US regulators will follow these international trends, the RFI suggests that the US financial regulators are, at a minimum, beginning to think strategically about issues of fairness, governance, and risk management in anticipation of potential future guidance or regulation. See Tori K. Shinohara Matthew Bisanz Alex C. Lakatos, Mayer Brown, *RFI on Financial Institutions' Use of AI Provides Opportunity to Shape Future Regulatory Framework*, April 2021, https://www.mayerbrown.com/en/perspectives-events/publications/2021/04/rfi-on-financial-institutions-use-of-ai-provides-opportunity-to-shape-future-regulatory-framework#_edn4

Reminder of existing regulations and guidance: The RFI contains a "non-exhaustive" list of laws, regulations, supervisory guidance, and agency statements that are relevant to AI, such as the agencies' longstanding model risk and third-party risk management guidance. While some of these statements contain broad-based principles, the piecemeal nature of this laundry list of guidance highlights the challenges financial institutions face in constantly retrofitting old regulations and guidance to new products and services. For example, the interagency guidance on model risk management was published almost a decade ago and articulates supervisory expectations for how institutions should evaluate conceptual soundness. However, models and associated model risk management has evolved over the last 10 years, and as highlighted in the RFI, current industry practices for evaluating conceptual soundness now incorporate post-hoc methods. Similarly, existing agency statements on managing third-party risk address model-related issues, such as maintaining intellectual property rights and negotiating access and audit rights, in a cursory manner that does not contemplate the unique features of AI use cases, including dynamic updating and alternative data.

Explainability: While transparency provides advance notice of algorithmic decision-making, explainability involves retroactive information about the use of algorithms in specific decisions. This is the main approach taken in the European Union's General Data Protection Regulation (GDPR). The GDPR requires that, for any automated decision with "legal effects or similarly significant effects" such as employment, credit, or insurance coverage, the person affected has recourse to a human who can review the decision and explain its logic. This incorporates a "human-in-the-loop" component and an element of due process that provide a check on anomalous or unfair outcomes.

A sense of fairness suggests such a safety valve should be available for algorithmic decisions that have a material impact on individuals' lives. Explainability requires (1) identifying algorithmic decisions, (2) deconstructing specific decisions, and (3) establishing a channel by which an individual can seek an explanation. Reverse-engineering algorithms based on machine learning can be difficult, and even impossible, a difficulty that increases as machine learning becomes more sophisticated. Explainability therefore entails a significant regulatory burden and constraint on use of algorithmic decision-making and, in this light, should be concentrated in its application, as the EU has done (at least in principle) with its "legal effects or similarly significant effects" threshold. As understanding increases about the comparative strengths of human and machine capabilities, having a "human in the loop" for decisions that affect people's lives offers a way to combine the power of machines with human judgment and empathy.

Risk assessment: In the 1974 Privacy Act, risk assessments were originally developed as "privacy impact assessments" within the federal government. They have since evolved as widely used privacy-management tools to evaluate and mitigate privacy risks in advance, and are required by the GDPR for novel technology or high-risk uses of data. Proposals for privacy legislation from Sen. Ron Wyden (D-Ore.) and Intel Corporation would require that any automated decision-making be preceded by an assessment of its risks to individuals. Wyden has also filed a separate, stand-alone bill on algorithmic decision-making, the Algorithmic Accountability Act. Risk assessments for algorithmic decision-making provide an opportunity to anticipate potential biases in design and data as well as the potential impact on individuals. For the regulatory burden to be proportionate, the level of risk assessment should be appropriate to the significance of the decision-making in question, which depends on the consequences of the decisions, the number of people and volume of data potentially affected, and the novelty and complexity of algorithmic processing.

Audits: Audits evaluate privacy practices retrospectively. Cam Kerry of Brookings notes that most legislative proposals contain some general accountability requirements to ensure companies comply with their privacy programs, and some include self-audits or third-party audits. Paired with proactive risk assessments, auditing outcomes of algorithmic decision-making can help match foresight with hindsight; although, like explainability, auditing machine-learning routines is difficult and still developing. https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/

One of the clear lessons from the AI debate, as summarized in a review of best practices by Brookings scholar Nicol Turner Lee with Paul Resnick and Genie Barton, is that "it's important for algorithm operators and developers to always be asking themselves: Will we leave some groups of people worse off as a result of the algorithm's design or its unintended consequences?" (emphasis in original). https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Because of the difficulties of foreseeing machine learning outcomes as well as reverse-engineering algorithmic decisions, no single measure can be completely effective in avoiding perverse effects. Thus, where algorithmic decisions are consequential, it makes sense to combine measures to work together. Advance measures such as transparency and risk assessment, combined with the retrospective checks of audits and human review of decisions, could help identify and address unfair results. A combination of these measures can complement each other and add up to more than the sum of the parts. Risk assessments, transparency, explainability, and audits also would strengthen existing remedies for actionable discrimination by providing documentary evidence that could be used in litigation. Not all algorithmic decision-making is consequential, however, so these requirements should vary according to the objective risk.

Below reprint from Jessica Newman, *Explainability Won't Save AI*, May 19, 2021, https://www.brookings.edu/techstream/explainability-wont-save-ai/

Much of artificial intelligence, and particularly deep learning, is plagued by the "black box problem." While we may know the inputs and outputs of a model, in many cases we do not know what happens in between. AI developers make choices about how to design the model and the learning environment, but they typically do not determine the value of specific parameters and how an answer is reached. The lack of understanding about how an AI system works, in some cases even by the people who have developed it, is one of the reasons AI poses novel safety, ethical, and legal considerations, and why oversight and governance are especially important. Black box deep learning models are vulnerable to adversarial attacks and prone to racial, gender, and other demographic biases. Opacity is especially problematic in high-stakes settings such as health care, lending, and criminal justice, where significant harms have already been reported. See Jessica Newman, Explainability Won't Save AI, May 19, 2021, https://www.brookings.edu/techstream/explainability-wont-save-ai/

Explainable AI (XAI) is often offered as the answer to the black box problem and is broadly defined as "machine learning techniques that make it possible for human users to understand, appropriately trust, and effectively manage AI." Around the world, explainability has been referenced as a guiding principle for AI development, including in Europe's General Data Protection Regulation. Explainable AI has also been a major research focus of the Defence Advanced Research Projects Agency (DARPA) since 2016. https://www.darpa.mil/program/explainable-artificial-intelligence

However, after years of research and application, the XAI field has generally struggled to realize the goals of understandable, trustworthy, and controllable AI in practice.

This gap stems largely from divergent conceptions of what explainability is expected to achieve and unequal prioritization of various stakeholder objectives. Studies of XAI in practice reveal that engineering priorities are generally placed ahead of other considerations, with explainability largely failing to meet the needs of users, external stakeholders, and impacted communities. By improving clarity about the diversity of XAI objectives, AI organizations and standards bodies can make explicit choices about what they are optimizing and why. AI developers can be held accountable for providing meaningful explanations and mitigating risks—to the organization, to users, and to society at large.

The explainability ideal

The end goal of explainability depends on the stakeholder and the domain. Explainability enables interactions between people and AI systems by providing information about how decisions and events come about, but developers, domain experts, users, and regulators all have different needs from the *explanations of AI models. These differences are not only related to degrees of technical expertise and understanding, but also include domain-specific norms and decision-making mechanisms. Achieving explainability goals in one domain will often not satisfy the goals of another.*

Consider, for example, the different needs of developers and users in making an AI system explainable. A developer might use Google's What-If Tool to review complex dashboards that provide visualizations of a model's performance in different hypothetical situations, analyze the importance of different data features, and test different conceptions of fairness. Users, on the other hand, may prefer something more targeted. In a credit scoring system, it might be as simple as informing a user which factors, such as a late payment, led to a deduction of points. Different users and scenarios will call for different outputs.

For now, users and other external stakeholders are typically afforded little if any insight into the behind-the-scenes workings of the AI systems that impact their lives and opportunities. This asymmetry of knowledge about how an AI system works, and the power to do anything about it, is one of the key dilemmas at the heart of explainability. Accessible and meaningful explanations can help reduce this asymmetry, but explanations are often incomplete and can be used (intentionally or not) to increase the power differentials between those creating AI systems and those impacted by them.

Domain differences

To understand the ways practitioners in different domains have different expectations for what they hope to achieve by building explainable AI systems, it is helpful to explicitly compare their goals. Below, see how three different domains—engineering, deployment, and governance—articulate the goals of explainable AI.

Engineering. In 2018, the Institute of Electrical and Electronics Engineers (IEEE) published a survey on explainable AI that illustrates how the technical and engineering domain conceptualizes the goals of XAI:

- *To justify an AI systems' results, for example to ensure that an outcome was not made erroneously.*
- *To provide better control over the system, for example by providing visibility into vulnerabilities and flaws.*
- *To continuously improve the system, for example by identifying and fixing gaps in the training data or environment to make it smarter and improve its utility.*
- *To discover new information and knowledge about the world, for example by identifying and relaying new patterns and strategies.*

Deployment. As AI applications are rolled out, the technology will increasingly interact with human beings, and the deployment domain seeks to understand how explainability impacts the human relationship with an AI system, including in military and other high-stakes contexts.

An overview of DARPA's XAI program provides an example of the deployment domain's goals for XAI:

- *To explain an AI system's rationale, describing not just what happened, but why.*
- *To characterize its strengths and weaknesses, by letting a user know under what conditions the system will successfully accomplish its goals.*

- *To convey an understanding of how the system will behave in the future, enabling a user to know when its use may be warranted and reliable.*
- *To promote human-machine interaction and enable partnership and coordination.*

Governance. A policy briefing on explainable AI by the Royal Society provides an example of the goals the policy and governance domain imagines XAI will achieve:

- *To give users confidence that an AI system is an effective tool for the purpose.*
- *To safeguard against algorithmic bias, for example through the identification of biased correlations due to skewed datasets or model design choices.*
- *To adhere to regulatory standards or policy requirements.*
- *To meet society's expectations about how individuals are afforded agency in a decision-making process.*

All three domains agree about the importance of explainability providing assurance about the effectiveness and appropriateness of a system at achieving its intended task, but the domains also differ in key ways. The engineering domain highlights the importance of control, which is either assumed in the other domains or not prioritized. And while the governance domain stresses the value of human agency, this is not a necessary outcome of goals in other domains. The engineering domain treats AI systems as constantly in flux and capable of regular improvement, while the other domains apparently expect greater consistency to enable informed expectations and adherance with policies. All three domains imagine different feedback loops. In the engineering domain, it is engineers' input that is incorporated; in the deployment domain, it is users' input that is incorporated; only in the governance domain is the impact on broader communities and the technology's relation to the broader world taken into consideration.

Explainability in practice
The reality of organizations' use of explainability methods diverge sharply from the aspirations outlined above, according to a 2020 study of explainable AI deployments.
https://dl.acm.org/doi/pdf/10.1145/3351095.3375624
In this study of 20 organizations using explainable AI, the majority of deployments were used internally to support engineering efforts, rather than reinforcing transparency or trust with users or other external stakeholders. The study included interviews with roughly 30 people from both for-profit and non-profit groups employing elements of XAI in their operations. Study participants were asked about the types of explanations they have used, how they decided when and where to use them, and the audience and context of their explanations.

The results revealed that local explainability techniques that aim to understand a model's behavior for one specific input, such as feature importance, were the most commonly used. The primary use of the explanations were to serve as "sanity checks" for the organization's engineers and research scientists and to identify spurious correlations. Participants looking for a more holistic understanding were interested in deploying global explainability techniques, which aim to understand the high-level concepts and reasoning used by a model, but these were described as much harder to implement. Study participants said it was difficult to provide explanations to end-users because of privacy risks and the challenges of providing real-time information of sufficiently high quality. But most importantly, organizations struggled to implement explainability because they lacked clarity about its objectives.

This study highlights the current primacy of engineering goals for explainability and how the needs of users and other stakehodlers are more difficult to meet. It shows that engineers often use explainability techniques to identify where their models are going wrong and that they may not have sufficient incentives to share this information, which is percieved as sensitive and complex, more broadly. While users and regulators want to see the vulnerabilities of AI systems, they may also want to see plans to fix uncovered problems or mitigate any negative impacts. The findings of this study are consistent with other examples of XAI in practice. For example, one machine learning engineer's account of explainability case studies documents her experiences of how they were used (successfully) for internal debugging and sanity checks, but not for user engagement.

Another 2020 study documents insights derived from interviews with 20 UX and design practitioners at IBM working on explainability for AI models and further explains the challenges practitioners face in meeting users' needs. https://dl.acm.org/doi/abs/10.1145/3313831.3376590

The study identifies a range of motivations for explainability that emerged from the participants' focus on user needs, including to gain further insights or evidence about the AI system, to appropriately evaluate its capability, to adapt usage or interaction behaviors to better utilize the system, to improve performance, and to satisfy ethical responsibilities. The study participants said that realizing these motivations was difficult due to the inadequacy of current XAI techniques, which largely failed to live up to user expectations. Participants also described the challenge of needing to balance multiple organizational goals that can be at odds with explainability, including protecting proprietary data and providing users with seamless integration.

These studies highlight that while there are numerous different explainability methods currently in operation, they primarily map onto a small subset of the objectives outlined above. Two of the engineering objectives—ensuring efficacy and improving performance—appear to be the best represented. Other objectives, including supporting user understanding and insight about broader societal impacts, are currently neglected.

Bridging the gaps

The five recommendations below are intended primarily for organizations developing XAI standards and practices. They offer an initial roadmap, highlighting relevant research and priorities that can help address the limitations and risks of explainability.

Diversify XAI objectives. Explainability techniques are currently developed and incorporated by machine learning engineers, and not surprisingly, their needs (and companies' desire to avoid legal trouble) are being prioritized. Realizing a broader set of XAI objectives will require both greater awareness of their existence and a shift in incentives for accomplishing them. XAI standards and policy guidelines should explicitly include the needs of users, stakeholders, and impacted communities to incentivize this shift. Explainability case studies are one pedagogical tool that can help practitioners and educators understand and develop more holistic explainability strategies. Diverse organizational guidance documents, recommendations, and high-level frameworks can also help guide an organizations' executives and/or developers through key questions to support explainability that is useful and relevant to different stakeholders.

Establish XAI metrics. While there has been some work done to evaluate AI explanations, most attempts are either computationally expensive or only focus on a small subset of what constitutes a "good explanation" and fail to capture other dimensions. Measuring effectiveness more holistically *likely requires combining a comprehensive overview of XAI approaches, a review of the different forms of opacity, and the development of standardized metrics. In particular, the evaluation of explanations will need to take into account the specific contexts, needs, and norms in a given case, and use both quantitative and qualitative measures. Further work in this space will help hold organizations accountable and promote successful AI deployment.*

Mitigate risks. Explainability entails risks. Explanations may be misleading, deceptive, or be exploited by nefarious actors. Explanations can pose privacy risks, as they can be used to infer information about the model or training data. Explainability may also make it easier for proprietary models to be replicated, opening up research to competitors. Methods for both documenting and mitigating these risks are needed and emerging standards and policy guidelines should include practical measures to do so. For some high-stakes decisions, it may be better to forgo deep learning models and the need for explainability techniques.

Prioritize user needs. So far,explainability has primarily served the interests of AI developers and companies by helping to debug and improve AI systems, rather than opening them to oversight or making the systems understandable to users. Prioritizing user needs has received some research attention, but user needs remain neglected. Key considerations in providing explanations to users include understanding the context of an explanation, communicating uncertainty associated with model predictions, and enabling user interaction with the explanation. Other user concerns include design practices for user experiences and accessibility. The field might incorporate decades of experience from the theory of risk communication. For example, this roadmap for risk communication with users developed by the Center for Long-Term Cybersecurity provides insights into the needs for two-way communication, accessible choice architecture, and protection for whistleblowers, among other mechanisms that help promote user interests.

Explainability isn't enough. Although explainability may be necessary to achieve trust in AI models, it is unlikely to be sufficient. Simply having a better understanding of how a biased AI model arrived at a result will do little to achieve trust. When students in England recently learned that they had been assigned standardized test scores based on a simple algorithm that had ascribed weight to schools' historic performance and, thus, advantaged rich schools, they were outraged, sparking protests in cities around the country. Explainability will only result in trust alongside testing, evaluation, and accountability measures that go the extra step to not only uncover, but also mitigate exposed problems.

And while explainability techniques will highlight elements of how a model works, users should not be expected to determine if that process is sufficient or to force changes when it is not. The precedent set by the 2017 Loomis v. Wisconsin case, in which the lack of explainability and potential racial bias in a criminal risk assessment algorithm were not seen as violating due process, underscores the gaps in accountability. https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1 Independent auditing and updated liability regimes, among other accountability measures, will also be needed to promote lasting trust.

Explainability is seen as a central pillar of trustworthy AI because, in an ideal world, it provides understanding about how a model behaves and where its use is appropriate. The prevalence of bias and vulnerabilities in AI models means that trust is unwarranted without sufficient understanding of how a system works. Currently, there is a significant discrepancy between the vision of explainability as a principle that reaches across domains and works for diverse stakeholders, and how it is being incorporated in practice. Bridging that gap requires greater transparency about the goals being optimized, and further work to ensure those goals align with the needs of users and the benefit of society at large.

Without clear articulation of the objectives of explainability from different communities, AI is more likely to serve the interests of the powerful. AI companies should clarify how they are using XAI techniques, to what end, and why, and make full explanations as transparent as possible. The entities currently developing XAI standards and regulations, including the National Institute of Standards and Technology, should take note of current limitations of XAI in practice and seek out diverse expertise about how to better align incentives and governance with a full picture of XAI objectives. It is only with the active involvement of many stakeholders, from the social sciences, computer science, civil society, and industry, that we may realize the goals of understandable, trustworthy, and controllable AI in practice.

Jessica Newman is a research fellow at the UC Berkeley Center for Long-Term Cybersecurity
For more about the practices and challenges of implementing AI principles, see the CLTC report, Decision Points in AI Governance: Three Case Studies Explore Efforts to Operationalize AI Principles.
https://cltc.berkeley.edu/wp-content/uploads/2020/05/Decision_Points_AI_Governance.pdf

3.0 Cybersecurity: Whether institutions have identified particular cybersecurity risks related to AI and the types of controls that can be employed.

Dynamic updating (where the AI has the capacity to update on its own): How institutions manage risks related to dynamic updating, particularly validation, monitoring and tracking.

3.1 COMPLIANCE FOR A DIGITAL WORLD: BSA/AML

BSA/AML compliance programs in the United States are often associated with high costs and inefficiencies. At the core of the problem is the lack of standardization across institutions and BSA/AML compliance processes such as know your customer (KYC), customer due diligence/enhanced due diligence, and transaction monitoring and alerting. This has resulted in burdensome compliance costs for financial institutions and has also created tense relationships between such institutions and their customers.

Artificial Intelligence (AI) and blockchain could help financial institutions tackle this fragmented compliance landscape by making BSA/AML compliance more efficient. The machine-driven processes of AI, combined with the inherently secure and collaborative properties of blockchain, can facilitate a windfall in cost reduction while optimizing regulatory compliance. It is therefore incumbent upon legislators, regulators, developers and the general public alike to understand the potential in combining these two promising technologies into novel enterprise-ready solutions for BSA/AML compliance.

BIOMETRICS AND ARTIFICIAL INTELLIGENCE

Biometrics – working in concert with a combination of artificial intelligence (AI) and machine learning (ML) to scan, analyze and then to create what could be a varied biometric identity database capable of verifying and storing fingerprints, facial features, even voice and device data – could allow for not only tougher, more meticulous identity security, but also a deeper understanding of a financial institute’s customer profile – giving banks and other fintech a truer way to “know your customer”.
https://techwireasia.com/2020/08/does-the-combo-of-ai-biometrics-hold-the-key-to-stopping-identity-theft-and-money-laundering/

The FINTECH industry is currently developing at a huge pace, changing the financial infrastructure and even approaches to doing business. At the same time, there are only two key requirements for this industry: improving the efficiency of the financial system or individual business, and security. Biometric technologies meet both of these requirements and are perfectly combined with the latest achievements of FINTECH. The most obvious area of biometrics use is rapid and reliable identification of the client at different steps and in different scenarios of financial interactions. The second area is ensuring security when working with personal information and financial data. This is also true for all kinds of payment and transfer systems, banking and personal Finance, lending, asset and investment management, and, finally, insurance. The simplest biometrics is used in fingerprint scanners of modern smartphones and tablets, and they can be used to access the Google Pay and Apple Pay systems. Many large Russian banks already use biometrics in their mobile apps both for logging in to the app and for confirming transactions (usually instead of confirming via SMS). But this is also difficult to call a decent protection-smartphone manufacturers focus on the speed of the fingerprint sensor, which does not affect the accuracy of recognition in the best way. The sensor usually reads only part of the fingerprint. Both a recent study by scientists from Michigan and the creation of the so-called MasterPrint (a kind of «arithmetic mean» fingerprint) confirm this.

The international payment system MasterCard has launched in 12 European countries the ability to confirm online purchases using selfies. Amazon, Uber, and even some government organizations in the United States are moving in the same direction. Of course, this technology is convenient, but it is not a panacea, especially since you can change the password stolen by hackers, but it will be more difficult with the face. There is also 3D Secure 2.0 – a new version of the Protocol that received a modified verification procedure. The payment confirmation itself is implemented using various biometric parameters – facial contours, fingerprints, palm veins, and so on. At the same time, the internal risk assessment system is responsible for up to 95% of total transactions, and only in the case of the remaining 5%, the system will request a verification code. HSBC Corporation announced the launch of a voice payment confirmation service, which also cannot be a 100% reliable method of identification. Ubiquitous mobility has become one of the essential trends for FINTECH, and it is difficult to imagine any new financial application without using biometrics, so either sensor manufacturers will have to catch up with FINTECH, or start using additional identification systems other than fingerprints. Only a multi – modal approach-authentication using several biometric indicators at once-can provide the proper level of protection while maintaining comfort and speed of use. The weakest and most vulnerable modality is the voice, which is highly dependent on ambient noise and is easily intercepted by third-party technical means. Similar problems arise when choosing video identification as the only way – the quality of lighting, weather, and minor changes in appearance greatly complicate the process and affect the result. Much better is the case with identification by drawing the veins of the palm, a three-dimensional model of the face, a photo taken in the IR range, or the iris, especially when they are superposed in order to control compromise and manage risks. The Id-Me platform, for example, allows you to choose authentication based on different indicators. In Russia, the use of biometrics became particularly active after the FINOPOLIS forum (October 2016, Kazan) – the largest market participants (the Bank of Russia, Rosfinmonitoring and the Ministry of communications) announced the launch of a pilot project in a number of Russian banks. An undoubted plus for the banks themselves is additional expansion into the regions, because if there is an adequate identification and authorization system, the client will be able to fully use the Bank's services from any city or even country. If, after testing in several banks, the project is considered successful, the Ministry of communications will be able to create an NBP (national biometric platform), and the Central Bank will be able to significantly expand the range of financial services provided using methods of biometric verification of the credit institution's client in remote service channels. Here, more than anywhere else, it is important to ensure

data security and reliable authentication at all stages of interaction.
https://recfaces.com/articles/biometricheskie-tehnologii-v-fintehe-i-bankinge

The New ABC's: *Artificial Intelligence, Blockchain and How Each Complements the Other*
see Dario de Martino, Marc-Alain Galeazzi, Vivian L. Hanson, and Lee Adam Nisson available at https://www.mofo.com/resources/insights/200316-compliance-digital-world-bsa-aml-ai-blockchain.html?utm_source=publications&utm_medium=email

BSA/AML compliance programs in the United States are often associated with high costs and inefficiencies. At the core of the problem is the lack of standardization across institutions and BSA/AML compliance processes such as know your customer (KYC), customer due diligence/enhanced due diligence, and transaction monitoring and alerting. This has resulted in burdensome compliance costs for financial institutions and has also created tense relationships between such institutions and their customers.

Artificial Intelligence (AI) and blockchain could help financial institutions tackle this fragmented compliance landscape by making BSA/AML compliance more efficient. The machine-driven processes of AI, combined with the inherently secure and collaborative properties of blockchain, can facilitate a windfall in cost reduction while optimizing regulatory compliance. It is therefore incumbent upon legislators, regulators, developers and the general public alike to understand the potential in combining these two promising technologies into novel enterprise-ready solutions for BSA/AML compliance.

Under the Bank Secrecy Act of 1970 (BSA), buoyed by the USA PATRIOT Act of 2001, the United States instituted a compliance regime where financial institutions are required to collaborate with the government in order to prevent the occurrence of financial crimes, including money laundering and terrorist financing. The onus of these requirements falls on the financial institutions, which are responsible for setting up appropriate safeguards and for reporting any suspicious activities, usually referred to as BSA/anti-money laundering (AML) compliance programs.

BSA/AML compliance programs in the United States are often associated with high costs and inefficiencies. At the core of the problem is the lack of standardization across institutions and BSA/AML compliance processes, such as know your customer (KYC), customer due diligence/enhanced due diligence, and transaction monitoring and alerting. The BSA does not affirmatively set forth types of information that must be collected or parameters of what constitutes suspicious activity. Instead, BSA/AML compliance programs require financial institutions to have a risk-based compliance program, rather than a rule-based one. Each institution must establish and implement an adequate compliance program commensurate with such institution’s risk profile. Thus, no two BSA/AML compliance programs are the same.

This fragmented BSA/AML landscape has resulted in burdensome compliance costs for financial institutions and has also created tense relationships between such institutions and their customers. For example, costs relating to governance, risk, and compliance account for approximately 15-20% of the total “run in the bank” costs for most major banks.[1] Even though financial institutions with $10 billion or more in revenue each spent approximately $150 million in 2017 compared to $142 million in 2016 on BSA/AML, the average customer onboarding period increased to 26 days from 24 days in 2016.[2] In addition, BSA/AML still remains fallible and institutions are penalized for flaws, adding more costs as a result.[3] However, Artificial Intelligence (AI) and blockchain could help financial institutions tackle such issues by making BSA/AML compliance more efficient.

BSA/AML still remains a process largely driven by manual input and human analysis. Applying AI more broadly could drive down costs through workflow automation and through greater precision and speed in the analysis of large amounts of structured and unstructured data.

A prime example of where AI could add value is the suspicious activity report (SAR). According to a recent study, “over 95 percent of system-generated alerts are closed as ‘false positives’ in the first phase of review.”[4] All in all, such efforts lead to the financial industry wasting billions of dollars in investigations because a vast majority of all alerts never culminate in valid SARs.[5] Deploying an AI system that can “learn” as it encounters more data could result in more effectively weeding out false positives and fine tuning transaction monitoring scenarios to alert only on transactions that have a higher probability of resulting in an SAR filing.

Further, when a financial institution collects BSA/AML-related information, such information is largely siloed away from other financial institutions or even from other departments at the same financial institution. This means that, after Bank A completes BSA/AML diligence on potential customer Jane Doe, if she wishes to open an additional account at Bank B, Bank B must separately conduct its own process. This duplication is burdensome not only for Bank B but also for Jane. This is where blockchain could offer a solution to address the information portability problems of BSA/AML. Verified customer information could be placed on a permissioned blockchain once consensus is reached with respect to the accuracy of such information. The information recorded on such blockchain-enabled system would be tamper resistant given the cryptographically hashed nature of blockchain data entries. Financial institutions would be able to share access to such secure, transparent, and immutable BSA/AML information and would no longer need to duplicate the collection of BSA/AML-relevant information. However, since BSA/AML compliance programs must be commensurate with a financial institution’s specific risk profile, each institution would still have to conduct its own risk assessments with regard to its customers and their transactions.

AI + Blockchain, Beyond Compliance

The machine-driven processes of AI, combined with the inherently secure and collaborative properties of blockchain, can facilitate a windfall in cost reduction while optimizing regulatory compliance. It is incumbent upon legislators, regulators, developers, and the general public alike to understand this potential.

Do so, and we may craft laws that enable the synergies to be gained, find new means to combine these technologies more efficiently, and take full advantage of the power the unique combination of AI and blockchain technology has to offer us

[1] See Matthias Memminger, Mike Baxter and Edmund Lin, Banking Regtechs to the Rescue?, Bain & Company (Sept. 18, 2016), available at: https://www.bain.com/insights/banking-regtechs-to-the-rescue/.

[2] See Thomson Reuters 2017 Global KYC Surveys Attest to Even Greater Compliance Pain Points, Reuters (Oct. 26, 2017), available at: https://www.thomsonreuters.com/en/press-releases/2017/october/thomson-reuters-2017-global-kyc-surveys-attest-to-even-greater-compliance-pain-points.html.

[3] See Joshua Fruth, Anti-Money Laundering Controls Failing to Detect Terrorists, Cartels, and Sanctioned States, Reuters (Mar. 14, 2018), available at: https://www.reuters.com/article/bc-finreg-laundering-detecting/anti-money-laundering-controls-failing-to-detect-terrorists-cartels-and-

sanctioned-states-idUSKCN1GP2NV; see also Deepak Amirtha Raj, Spotlight on the Remarkable Potential of AI in KYC, LinkedIn Pulse (June 13, 2017), available at: https://www.linkedin.com/pulse/spotlight-remarkable-potential-ai-kyc-deepak-amirtha-raj.

[4] See Fruth, supra note 13.

[5] See Fruth, supra note 13.

3.3 **Cybersecurity Landscape Snippets**

According to data from the Deloitte Center for Financial Services, 64 percent of executives at financial firms expect to increase their spend on cybersecurity in the future, and 60 percent anticipate increasing investments in cloud storage and computing. This is unsurprising, as banks pursue increased integrations with third-party partners that present potentially higher security risks; therefore, ensuring cutting-edge cybersecurity solutions is vital. Current trends point to the possibility of a single cybersecurity centre for the entire market and the introduction of mandatory response procedures and reporting standards for all players. With a focus on cybersecurity practically guaranteed throughout this year, it remains to be seen what path will be taken, but we look forward to exploring how the industry can change and adapt.

Digital currencies

Digital currencies and central bank digital currencies (CBDCs) have dominated news headlines for some time. Digital currencies can deliver transparency in tracking the movement of funds, suitable for targeted allocations of funds. Currently, a complex system of monitoring by fiscal and regulatory authorities is used for this purpose, and digital currencies will provide us with the opportunity to reduce or even remove this burden.

Already, we have heard 2021 being referred to as the year of digital currencies, and so far, this seems to be holding true. However, with so many competitors in the field, the “winner” of the CBDC race will be the one who can bring its stablecoin as close as possible to the role of a universal currency. So far, the players are primarily hindered by the lack of uniform security standards for cross-border transfers.

Payments without intermediaries

Our final trend is the development of integrated-payment channels, where payments are not served by traditional financial providers. This is a long-term trend that we expect to continue through 2021 and beyond. We will see the development of solutions that allow customers to operate without Visa, Mastercard or other acquiring banks. They will be implemented by both fintechs (financial-technology firms) and bank consortia. For example, VTB is actively involved in forming a single-payment space with the countries of the Eurasian Economic Union (EAEU).

Running through all of these trends is a clear thread, the importance of a customer-centric approach to products and services. Customers are at the core of all financial products. In 2021, this must be prioritised even further, and technology can deliver this. By introducing technology to their business processes, products and services, financial institutions can promptly respond to shifting market needs.

Cutting-edge technology has become a key differentiator for competition within the Russian banking sector, and keeping up is now a matter of survival for financial-services providers. **A commitment to**

investing in innovative technology will be the key to success in the banking market. COVID-19 has shown us the importance of being agile and flexible, providing vital digital solutions for our customers. Now, in 2021, we must move another step forward, take this agility and combine it with digital transformation and innovation. By doing this, we will truly change what banking looks like for years to come.

3.4 COMPARATIVE CONSIDERATIONS

Implications of Data Privacy for Financial Technology in Asia

https://www.lw.com/thoughtLeadership/implications-data-privacy-financial-technology-asia-asifma

Impact/Considerations of Data Protection Regulations on Fintech for Personal Data

Data protection rules have a fundamental impact on the uptake and development of financial technologies. These financial technologies are being developed by companies ranging from start-ups through to existing, global financial institutions. Without exception, these companies rely on some measure of processing of personal information and this can be core to the relevant product. Different interpretations of data protection definitions and rules can cause confusion, as these rules directly impact the innovative technologies and products being developed, and the confidence of the business to make a product available across markets in Asia.

Some key areas where different regulations across the region have complicated the adoption or expansion of Fintech include the definition of ‘personal data,’ unclear and disparate consent and notice regimes, and restrictions on the sharing or export of data to third parties.

Defining ‘Personal Data’

The central tenet of data privacy regulation is that such laws govern the practices and activities relating to ‘personal data’ or ‘personal information’. However, the definition of personal information is not consistent across Asia and the interpretation of such terms also varies across the regulatory landscape.

In addition to this, certain jurisdictions go a step further and include additional rules for ‘sensitive’ data (such as biometric data, which can be a desired method of authentication as it can be considered more secure than traditional methods or which may be combined with another authentication method, such as a password, for increased security). However, not all jurisdictions have a subset of rules for sensitive data.

As a result, the use of biometric data, for example, by a company in one jurisdiction may not be practical or even permitted in another, even though it may be a more secure method of authenticating a user.

Whether encrypted data and IP addresses are personal data has been hotly debated.

In certain jurisdictions such as Australia, it can be argued that encrypted data is sufficiently obfuscated that it no longer constitutes ‘personal information’, while in others, such as in the European Economic Area (even before the entry into force of the General Data Protection Regulation), it remains personal data in its encrypted form. Encryption is at the heart of many emerging financial technologies and not knowing whether data is subject to regulation even when it is encrypted can hamper a firm's ability in many jurisdictions to take a dataset and obfuscate in a way that allows it to be useful while still complying with privacy requirements. A harmonised approach to the status of encrypted data and IP addresses would be a boost to innovation and the growth of Fintech ecosystems.

Consent and Notice Regimes

Jurisdictions across Asia have varying rules governing consent, with some common themes but no single, fixed rule that a firm can leverage to treat Asia as a whole. In certain jurisdictions, express consent, or 'opt in' is required for personal data, whereas in others, an implied consent will suffice.

Indeed, certain jurisdictions, such as Cambodia, have no explicit rules at all, and therefore firms must look to obscure legislation or general international treaties to determine what rules might apply to obtain consent for the use of personal data in that jurisdiction. If firms are subject to financial regulation, additional consent requirements may apply to personal and non-personal data (taking as an example, banking secrecy consents under the Banking Act in Singapore).

As such, there is no "one size fits all" approach that can be taken by companies looking to comply with data protection rules, meaning they and their partners need to have a tailored approach to consents: determining where they are actually relevant and meaningful and drafting up different consent clauses for different jurisdictions. This presents an additional costly barrier to entry to new markets and can especially factor into decisions on the value proposition whether of entering a smaller market or rolling out a new product in an existing market.

Some firms have resorted to asking for explicit consent from consumers whenever they collect personal data, under the misconception that doing so is the simplest way to ensure compliance with data protection regulations. This is an erroneous view, for multiple reasons.

In many instances, consent is not a meaningful basis for processing, as the processing is necessary or required in order to provide the goods or services (for example, AML or KYC processing, or consent from employees to outsource payroll functions to a third party where there is no alternative internal function). Furthermore, individuals must be able to withdraw their consent at any point. This means that companies must have mechanisms to remove personal data from their database upon an individual's request. For companies that use personal data collected for machine-learning, this might create complex problems such as ensuring that the algorithm can "reverse" any learning from a specific data set at any point in time. In addition, sometimes the very act of requesting consent through the use of personal data like email addresses is still a violation of data privacy regulations.

The multitude of laws across Asia that deal with the requirements for notice has also led to wide variation in privacy policies in the market. Start-ups favour simple, template-based notices while more sophisticated firms generally adopt more precise and detailed privacy policies. Privacy notices also vary in detail, from policies that provide a high-level overview to others that go into more granular detail (the latter driven by regimes such as those in Korea and Europe). The problem of regulatory fragmentation is compounded for multinational companies, whose products and services are offered in multiple jurisdictions. In those scenarios, companies usually adopt lengthy, and often convoluted, privacy policies with multiple annexes that only serve to confuse, rather than inform, users.

To add to the confusion, regulators across Asia have historically not proactively policed privacy notices, so until something goes wrong (for example, in the event of a data breach), an entity generally does not know if its privacy notice is considered compliant or appropriate by the regulator.

While there has not been definitive guidance on what is required to satisfy notice requirements, Asia has a great opportunity to learn from the experiences in the US and EU where privacy notices have become long and complex and are rarely read or understood by users, thereby inhibiting the user experience and creating an unnecessary bureaucratic burden. Instead, notices should be brief, to the point, and relevant to the type and risk of processing, and the means of delivery of the service (e.g. mobile phones versus websites versus paper etc.).

Sharing Data & Export Controls

The storage of data, and many processing operations, are in the modern environment outsourced to a third-party service provider that specialises in and may be better suited to securing such data (notwithstanding that institutions maintain control of the data even where the infrastructure is managed by a third party).

The ‘controller’, ‘processor’, ‘co-controller’ and ‘joint controller’ distinctions (to the extent these European-centric concepts have commensurate meanings in Asia), and the definition of what constitutes a ‘transfer’ of data, are difficult to align with the models of service delivery that are emerging. In traditional models of service delivery for software or infrastructure, it was reasonably clear who had responsibility for what personal information. This is not the case with hybrid solutions that mix, for example, Platform/Software as a Service with DLT.

Where privacy laws have created specific roles for entities, the laws are not as amenable to accommodate hybrid models of service delivery. The result is that either: (i) the parties to the transaction artificially assign themselves roles required by law that are not reflective of the nature of the service; or (ii) the parties have to try and backwards engineer the service to reflect the law (such as adding special permissions in a permissioned blockchain to allow an administrative account to have master control of the ledger – which is the antithesis of the security and integrity afforded by the technology). Such efforts to accommodate prior concepts and roles further have the adverse effect of impeding innovation and the ability to improve efficiencies.

Vagaries in interpretation of data protection laws also has an impact on the increasing volume of collaborations and joint ventures between established financial institutions and start-up firms. The aim of many of these relationships is that the parties can share information and pool resources to better service existing customers and attract new clients, and in doing so develop innovative new financial technologies and products. However, the challenges outlined above, including what conditions should be satisfied before data can be shared, whether such data can be encrypted or anonymised and therefore be used flexibly, and what role each party has given the traditional privacy frameworks, presents a significant barrier to rapid development and adoption of Fintech solutions by financial institutions, and consequently the growth of Fintech ecosystems.

Impacts and Limitations on Innovative Technology

Data privacy rules are playing a key role in influencing future innovations such as digital ID, big data, artificial intelligence and cloud-based systems, all of which play a significant role in the development of new financial technology and will be critical to the growth of Fintech firms and applications in the Asia-Pacific region. Policymakers should take care that their rules are technology neutral, such that the requirements do not unduly impede technologies that could otherwise present great opportunities in meeting the financial needs of their citizens.

Distributed Ledger Technology

A pertinent example is blockchain and DLT, which do not pair well with privacy regimes that mandate specific roles for entities that process data, such as the regimes in Korea and Europe. A permissionless blockchain stores (personal) information in multiple locations with permanency. Although it is possible to find creative arguments to align the technology with the law – for example, arguing that obfuscating the data is adequate to satisfy ‘deletion’ requirements; or to twist the technology to comply with the law (such as allowing a ‘super user’ on a permissioned blockchain to delete data; or having the personal data stored outside of the blockchain itself with the blocks storing mere reference files) – each of the arguments or technical solutions runs contrary to the intention of the technology itself, as they increase the overall complexity of fetching and storing data on a blockchain, and in some cases actually increases the risk to the integrity and security of the data. In particular, the benefit of transparency with blockchain is reduced since, by storing data outside the blockchain, there is no way of knowing who accessed the data, and who has access to the data.

As noted above, the way DLT processes personal data is challenging the established paradigms that underpin many data protection regimes, such as the notion of a ‘processor’ and a ‘controller’. The difficulties presented by the underlying conceptual framework, along with inconsistencies in data export, transfer and localisation laws across the region (for example Hong Kong has ‘soft’ export restrictions that are not yet in force but considered “guidance,” while localisation restrictions are already in place in Indonesia and Vietnam) makes it difficult for financial institutions to address Asia as a whole and reduces their ability to use promising technologies like DLT. This is compounded by issues in managing data subject rights (e.g., How do you delete information that is un-deletable? How can data be corrected if the ledger is immutable?).

Artificial Intelligence

Another example of new technology that does not fit neatly into the existing data privacy framework is artificial intelligence and machine learning. This includes robo-advisors and robo-solutions where systems are automated and the owner of the platform (i.e., the technology firm or financial institution) does not have interaction-by-interaction control over the system. Technology firms are offering up these types of platforms to financial institutions to use in test environments and in some cases with dummy customer data. It is unclear in these cases who would be the data controller and who would be the data processor. While robo-advisors or robo-solutions typically require user input of personal information in order to generate results, should a user decide not to proceed, data may be deleted, and therefore, may not technically be used or controlled by the entity that offers the platform. This makes it problematic for the parties involved with the platform to determine what their obligations are to users under data privacy rules. AI solutions are also challenging one of the founding notions of most transparency requirements under data privacy regimes: telling users what their data will be used for. The core benefit of AI is that it may have the ability to perform tasks and offer products and services to a customer that are perhaps not contemplated by a human being tasked with the same role. Does this mean that the AI is using personal data in a way that is not technically articulated as the purpose for which data was collected in the relevant privacy notice?

Big Data

Further, use of large databases of customer data with data sourced from different places (i.e., ‘big data’) is a helpful tool for start-ups seeking to experiment with products and train products to better address customer needs. However, the use of big data, particularly enriching data from multiple sources, presents a significant challenge: users may not be aware that certain data may be mixed with data sourced from a third party in order to create custom products.

Cloud Computing

Use of cloud solutions presents its own challenges for regulated financial institutions. Multiple regimes with varying degrees of expectation with regards to control, access, audit and transfer, make it difficult to economically leverage cloud services. For example, Korea, Singapore, and Europe each require some form of contractual control for data export, whether or not data is accessed by a third party. This means that use of a cloud service provider, even where the master encryption key may be held by the customer itself and not the service provider, is treated no differently than a traditional software-supply set-up. To contrast, a jurisdiction such as Australia arguably allows an entity to leverage cloud services and the law instructs that provided certain controls are in place such use of cloud is treated as if it is the clients’ own infrastructure (i.e., as a ‘use’ rather than a disclosure’ for the purposes of the Australian Privacy principles).

The varied nature of data protection laws across Asia, together with a lack of clarity on how such laws fit with emerging technologies results in significant uncertainty for innovative firms, large and small, developing financial technologies for the benefit of consumers. Compared to other regions, Asia would disproportionately benefit from harmonisation of approach and interpretation of data protection laws in order to boost innovation and strengthen its position as a centre for developing and bettering Fintech solutions.

III. Importance of a Principles-Based Approach

The Asia Securities Industry and Financial Markets Association (ASIFMA) recommends the adoption of a principles-based approach to regulation as a guide to develop tools for ensuring privacy of personal data. Broadly, a principles-based approach means moving away from reliance on detailed, prescriptive rules and instead designing highlevel rules or principles that are results-oriented and focuses on "what" rather than "how". https://www.lw.com/thoughtLeadership/implications-data-privacy-financial-technology-asia-asifma

Regulation should define the desired outcomes (i.e. results) rather than setting out an exhaustive and prescriptive list of the means that must be taken by a data controller or a data processor to achieve the desired outcomes (i.e. technical details). Such an approach should also be technology-neutral, to allow principles to preserve their relevance and applicability in the context of continually changing and emerging technologies, particularly in the Fintech space.

Principles-based regulation generally contains terms that are more qualitative than quantitative in nature, and uses evaluative terms (e.g. fair, reasonable, suitable) rather than bright line rules. This enables a risk-based approach to compliance, which is particularly important in the Fintech space as it does not stifle innovation with rigid requirements while still meeting regulatory needs in an area of rapid change and growth.

Promoting Innovation

A principles-based approach affords firms the flexibility and space to innovate, recognising the shortcomings of a one-size-fits all model. It gives firms, whether they are large, well established institutions or new start-ups, the flexibility to take a risk-based approach to compliance that is tailored to their own business models, needs and practices. For start-ups with limited resources, this flexibility is particularly critical as it permits them to focus more on their products and services rather than diverting precious, limited resources towards complying with prescriptive rules unfit for their risk profiles.

Following principles rather than prescriptive rules also makes it easier for firms to operate confidently across borders and enter new markets, a consideration that is especially important if the disparate markets of the Asia-Pacific region are to benefit from Fintech.

Harmonised principles can apply effectively throughout the region in ways that let firms access consumer bases internationally. When firms are more comfortable working across borders, they are more likely to confidently enter new markets, encouraging beneficial competition among firms that ultimately results in better solutions for clients and consumers.

Meeting Regulatory Needs

Regulators applying prescriptive rules to the industry face significant headwinds in a time of unprecedented innovation. Rules are just best guesses as to the future, and new technologies and applications make it likely that regulators will encounter unexpected situations where previously drafted rules cannot be effectively applied. Innovation is simply moving too quickly for backward-looking regulation and prescriptive rules serve only to stifle the potential of innovation. Principles-based and technology-neutral regulations, on the other hand, allow for a greater degree of “future-proofing” where the regulatory regime does not need to be updated or amended with every new technology or application. They can also continue to apply well even in situations where regulators have not yet been able to understand a new technology, lessening the risk of a delayed (or rushed) regulatory response that could bring about vulnerabilities in the regulatory framework.

A principles-based approach can also provide a solid basis for open and ongoing dialogues between regulators and firms using new technology solutions. This dialogue can in turn facilitate a more cooperative and educative approach to supervision, and industry buy-in to the flexibility of a principles-based approach can end a "tick-the-box" mindset to compliance that ultimately yields more substantive compliance and better results.

Principles that focus on operational results rather than technical details can be more effective because firms and their management are better placed than regulators to determine what processes and actions are required to best achieve regulatory objectives. Rather than prescribing specific processes or actions, regulators can simply define desired outcomes and check for compliance through normal supervisory mechanisms.

IV. Fundamentals of a Principles-Based Approach

That policymakers take the following into consideration when developing their own principles based regulatory frameworks for personal data privacy:

1. Focus on outcomes rather than processes
The definition of desired outcomes by regulators, rather than prescription of "one size fits all" processes for all firms and business models, allows firms to apply a risk-based approach to compliance that is best-tailored to specific business models and activities. Firms are better positioned to understand their own systems and vulnerabilities than regulators and should be empowered to leverage this understanding to make informed decisions on how to best meet regulatory objectives.

2. Ensure technology neutrality of regulation

Principles afford a degree of technology neutrality that is necessary in the ever-changing Fintech space. A regulatory framework, even with broadly principles-based, can be undermined if it does not account for the possibility, and indeed likelihood, of innovation and new technologies. Policies with specific technology requirements are inherently reactive to threat environments and become quickly outdated.

3. Ensure consistency with existing international best practices

As a means of ensuring regulatory harmonisation governments can turn to both international and regional frameworks to guide their efforts. This approach will provide a foundation for globally synchronised regulatory approaches and mutual recognition which would facilitate similar protections and also facilitate better access for individuals to protect their data rights.
The OECD Guidelines on the "Protection of Privacy and Transborder Flows of Personal Data" were developed by OECD member states over four decades ago through engagement with a wide cross section of stakeholders and focus on personal data. Notably, the OECD's guidelines specifically warn against restrictions on cross-border data flows that could "cause serious disruption in important sectors of the economy, such as banking and insurance."
The APEC Privacy Framework, which is partly based on the OECD Guidelines, is also an important resource for Asian governments. All of the member governments of APEC have signed the Privacy Framework which "promotes a flexible approach to information privacy protection across APEC member economies, while avoiding the creation of unnecessary barriers to information flows." It includes a set of principles and implementation guidelines that form the basis of the APEC Cross-Border Privacy Rules System (CBPR), an international framework that could be used to help harmonise the requirements for cross-border transfers. Currently, one of the key limitations of CBPR is that certification does not in itself mean that personal data can be transferred from any other APEC economy. The law in each other economy must permit such transfers. Currently, no laws in APEC economies clearly provide

that exports to APEC CBPR-compliant companies are allowed. This is an area that the Asian governments and privacy regulators may want to work on.

4. Focus on how data is stored, not the geographic location

There is a trend in the region of governments instituting data localisation and cybersecurity laws that require data be stored onshore to ensure its accessibility to local authorities. However, the geographic location of data is less important than how it is processed, as regional data centres can more effectively monitor and react to threats than can localised national data centres that segregate data. Differing requirements across jurisdictions also cause legal tensions that undermine the coordinated multijurisdictional approaches necessary to make cybersecurity, sanctions and AML enforcement effective. Governments and regulators should instead focus on how data is processed to ensure minimum protections are met regardless of the physical location of data centres.

5. Allow for mechanisms to facilitate cross-border sharing

The regulatory regimes that govern the financial sector must allow for mechanisms to facilitate the cross-border sharing of specific information that allows the private and public sectors to work together more effectively to ensure investor protection and combat financial crime. One such mechanism that has seen success is the global network of Financial Information Sharing Partnerships (FISPs). More than 20 countries have committed to developing public–private financial information-sharing partnerships (FISPs) that bring law enforcement and other public agencies together with groups of major financial institutions to tackle money-laundering and terrorist-financing risks more effectively. It would be helpful for local privacy laws to contain an exception to allow information sharing and cross-border transfer under such mechanisms.

In 2017, three new FISPs were launched in the Asia-Pacific. This includes the Fintel Alliance in Australia, the Anti-Money Laundering and Countering the Financing of Terrorism Industry Partnership (ACIP) in Singapore; and the Fraud and Money Laundering Intelligence Taskforce (FMLIT) in Hong Kong. In the first four months of FMLIT's operation, public–private information sharing through FMLIT is credited with contributing to the arrest of 65 persons and the restraint of HK$1.9 million worth of assets. FISPs must act within existing laws, however, and laws that develop barriers to information sharing threaten the development of these tools to fight financial crime more effectively. Further, there are many firms, large and small, using technology to come up with innovative solutions in the area of financial crime. Barriers on the movement of data across borders limits the efficacy of such solutions and therefore discourages innovation.

6. Preserve ability of firms to outsource functions to third-parties

As innovative firms in the financial sector grow, they often need to rely on a network of other firms to provide services that they are unable to accomplish effectively in-house or to leverage third-party expertise. These might include customer service, KYC screening or even some back-end IT functions. Through outsourcing, firms can achieve greater consistency of approach, leverage established expertise and reduce operational costs while maintaining high levels of efficiency. We recognise that an effective data protection regime must include requirements related to outsourcing and the engagement of personal data processors, but to avoid unnecessary complications or barriers and make local companies less competitive than those located outside, regulators should be mindful of how requirements might limit access

to critical third-party providers. https://www.lw.com/thoughtLeadership/implications-data-privacy-financial-technology-asia-asifma

3.5 NEWS on COMPARATIVE SYSTEMS

(Reuters) - China should introduce a regulatory framework for artificial intelligence in the finance industry, and enhance technology used by regulators to strengthen industry-wide supervision, policy advisers at a leading think tank said on Sunday.

"We should not deify artificial intelligence as it could go wrong just like any other technology," said the former chief of China's securities regulator, Xiao Gang, who is now a senior researcher at the China Finance 40 Forum.

"The point is how we make sure it is safe for use and include it with proper supervision," Xiao told a forum in Qingdao on China's east coast.

Technology to regulate intelligent finance - referring to banking, securities and other financial products that employ technology such as facial recognition and big-data analysis to improve sales and investment returns - has largely lagged development, showed a report from the China Finance 40 Forum.

Evaluation of emerging technologies and industry-wide contingency plans should be fully considered, while authorities should draft laws and regulations on privacy protection and data security, the report showed.

Lessons should be learned from the boom and bust of the online peer-to-peer (P2P) lending sector where regulations were not introduced quickly enough, said economics professor Huang Yiping at the National School of Development of Peking University.

China's P2P industry was once widely seen as an important source of credit, but has lately been undermined by pyramid-scheme scandals and absent bosses, sparking public anger as well as a broader government crackdown.

"Changes have to be made among policy makers," said Zhang Chenghui, chief of the finance research bureau at the Development Research Institute of the State Council.

"We suggest regulation on intelligent finance to be written in to the 14th five-year plan of the country's development, and each financial regulator - including the central bank, banking and insurance regulators and the securities watchdog - should appoint its own chief technology officer to enhance supervision of the sector."

Zhang also suggested the government brings together the data platforms of each financial regulatory body to better monitor potential risk and act quickly as problems arise.

(Reporting by Cheng Leng in Qingdao, China, and Ryan Woo in Beijing; Editing by Christopher Cushing)

With regard to intelligent customer service, at the end of 2015, the Bank of Communications launched China's first smart service robot, Jiao Jiao. Jiao integrates many cutting-edge AI technologies, including speech recognition, speech synthesis, natural language processing, faceprints, and voiceprints.

Currently, Jiao Jiao is deployed in Shanghai, Jiangsu, Guangdong, Chongqing, and other provinces. In April 2018, China Construction Bank launched China's first "unmanned bank," which solely uses artificial intelligence to handle business.

In mid-2019, Ping An Bank highlighted that with the continuous accumulation of application scenarios, the replacement rate of Ping An Bank's voice customer service has reached more than 80%. In contrast, the volume of customer service has increased by 2-3 times. As a result, the labor cost of customer service has decreased by 40%.

Over recent years, Robo-advisors have evolved from being an experimental technology to becoming mainstream. At the end of 2016, China Merchants Bank's mobile phone app launched China banks' first Robo-advisor "Machine Gene Investment." Since then, Robo-advisors have expanded rapidly in China. The big four commercial banks in China (BOC, ABC, ICBC, CCB) have successively deployed similar services.

Other joint-stock banks besides CMB, such as Shanghai Pudong Development Bank, CITIC Bank, Xingye Bank, Ping An Bank, and Guangfa Bank, have also built their own Robo-advisors. Among city commercial banks, Jiangsu Bank released "Alpha" as early as August 2017. Securities companies, such as Guangfa Securities, Everbright Securities, and Huabao Securities also developed Robo-advisors service. With the growth and development of China's AI technology, the financial industry is becoming increasingly more intelligent. As Yonglin Xie, the chairman of Ping An Bank, stated, "No matter if it is basic retail or consumer finance, private banking, or wealth management, all of it should be fully AI-based."Although banks are busy finding talent for their new AI facilities, integrating resources to build data platforms, and structuring and securing data storage, there are limits to what they can do with AI technologies. At present most AI systems are not sufficiently robust for all situations — for example intelligent customer service bots may fail at voice recognition in noisy environments, as may facial recognition in extremely bright or dark environments.

It is estimated that applying AI systems in the banking industry takes about one year from procurement to deployment. From the start, there tend to be system integration issues particular to different banking systems, and compatibility issues due to multiparty software and hardware suppliers.

Regulatory uncertainties may also also hinder AI innovation in banking and financial services industries, mainly regarding data use standards, privacy protection of user information, financial license procurements and so forth.

In 2018, Chinese financial institutions invested about CN¥160.4 billion (US$23 billion) in technology, an increase of 10 percent over 2017, of which AI hardware and software-related investment accounted for 10.4 percent. The banking industry was the biggest investor in AI-related applications, accounting for 70 percent of all market purchases.

A wide range of software programs, including those tasked with precision marketing and intelligent risk control platforms, accounted for two thirds of Chinese banks' AI purchases. AI-powered cameras and document identification machines and other hardware accounted for the balance. According to statistics from China's Banking and Insurance Regulatory Commission, the total investment made by

Bank of China in technology increased by 13 percent from 2017 to 2018; while relevant staff hires also increased by 10 percent.

Driving the investments is the belief that AI technologies will help banks become more flexible and specialized. The step beyond traditional banks and online banking, smart banks use data-driven methods and state-of-the-art technologies to redefine existing services, products, operations and business models. These banks of the future favour economies of scale and offer improved efficiency and reduced costs

Source: Synced China

By 2022, banks will be spending as much as $12.3 billion on AI and cognitive technologies with the race underway to integrate the latest capabilities into financial services

Globally, finance is believed to be outpacing all other industries when it comes to introducing AI, with Chinese banks and fintechs leading the way.

There is a strong trend for banks entering joint ventures to make the most of AI, but they learn from the mistakes of the big tech unicorns and communicate their progression.

While a lot of exciting innovation is occurring, regulation surrounding data protection and privacy, and an increased focus on liability will present ongoing concerns for financial institutions
In an era of low interest rates and wavering market confidence, putting in place the right efficiency measures is as critical as ever to banks' bottom lines. Add to that the ongoing disruption from fintechs, as well as the opportunity to gain insight into the future of workforces and everything points to tech solutions in artificial intelligence (AI), machine learning and even robotics.
Chinese banks leading the way

The view in the global technology community is that Chinese banks and fintechs are leading the way when it comes to machine learning and building a data science workforce. At the same time, many European and US banks are slightly ahead in terms of deployment — but Asia is moving fast and expected to overtake the West in the coming years.

Leading the way is China CITIC Bank, which developed its 'brain platform' in conjunction with Tsinghua University. The project has some 15 machine learning models applied in different parts of the bank, marketing, automation, AML and anti-fraud businesses. The project includes construction of an artificial intelligence platform and a blockchain-based trade finance business model in partnership with Bank of China and China Minsheng Bank.

For CITIC, the opportunities are in targeting individual needs of different operations within the bank. "The platform aims to intelligently empower each business line," a spokesperson explained about the platform's self-developed architecture.

"The platform implements intelligent full-process services from artificial intelligence model training to application deployment. The platform is easy to use and tailored to fit the specific needs of banking business, provides real-time interfaces and batch processing interfaces," the spokesperson added.

With CITIC's AI operations firmly in place, it is turning its attention to aiBank, its digital banking joint venture with Baidu. aiBank spent 2019 seeking strategic investment partners for as much as $1 billion

(RMB 7 billion), as well as announcing a partnership with credit card company 51 Credit Card to create a “state-level innovation trial” fintech ecosystem. aiBank will reportedly focus on smart risk control and big data applications, as well as finding solutions in more traditional fields such as consumer finance, credit payments, escrow operations and fintech.

ICBC, on the other hand, has its smart banking construction scheme focused on improving services for its more than one billion retail customers. The bank is concentrating on intelligent customer service and building an operation support system that integrates all channels and prioritises customer experience, using tools such as voice bots, seamless connection across AI and manual platforms, and scenario-embedded smart Q&A.

Making rapid advancements while cutting through hype

The most advanced players are already way ahead in their vision of the tech future.

None more so than China’s Ping An Group, which is forging a distinct path with a comprehensive vision. Chief innovation officer Jonathan Larsen recently revealed that in spite of the $1 billion innovation fund under his watch, the bank has “no illusion that we have any monopoly” on transformational technology.

“We have found that the capabilities it has acquired in areas like visual artificial intelligence, natural language processing and the ability to create integrated cloud services that can be sold as verticals to other financial institutions,” he said.

While Ping An is investing in tech integrations, connect technologies and building a proprietary tech arm that is a provider to other banks, part of Larsen’s remit is to use the fund to scout businesses who have technology that is ahead of the game. Larsen shared that they are finding broad applicability to these concepts, which are also have significant horizontal scale.

“Smart city is one of our five ecosystems that we’re focused on. We provide a single app that allows every citizen in Shenzhen to access pretty much every government service with the same ease of use as you can access an online service,” Larsen explained. “What we’re finding is that a lot of our analytics, AI and blockchain solutions can be used for government recordkeeping, property registers, traffic management and pollution management.”

In 2018, Ping An’s IT capital expenditure was up by 82% year-on-year, and it increased its technology staff by over 44%. But Larsen stressed that a big part of his remit is applying a healthy skepticism to cut through the hype to find real capabilities and technologies with a unique business mode.

It can be argued that China is leading the way in AI in finance due to an early appetite to merge tech and finance, such as Tencent-led WeBank and Ant Financial’s Alipay, which dominate China’s third-party mobile payments sector, estimated to be worth around $7.17 trillion (RMB 50 trillion). The two firms specialise in building full psychographic profiles of customers through personal, social, financial and commercial data.

WeBank recently announced the establishment of Retail & AI Joint Laboratory with retailer BKK Group, to “break through the difficulties of traditional retail industries to help the increase of the economic growth.” The firm will focus on three core pillars: smart labor, smart operation analysis, popular products forecast, with bold aims such as to “improve 50% of efficiency in the cashier position”.

The rest of Asia is taking advantage of late arrival

Other Asian nations, which are not yet as advanced as China, are starting to pick up on robotic process automation, chatbots and machine learning in credit analytics.

Some firms, such as Myanmar’s Yoma Bank, have showed that there are advantages from entering late into the market and leaping ahead, taking decision analytics platform with Experion and building a leading credit analysis by jumping on the latest advances in machine learning.

Deepak Sharma, chief digital officer at India’s Kotak Bank, said that his firm began its AI journey two years ago. The brief is to reduce cost and improve efficiency, not simply in man hours (although they’ve saved around 15,000 in 2019) but through reducing error rate and decreasing turn-around time.

Kotak can be viewed as a fast follower of international trends, looking at productivity, personalisation, and fraud detection like most leading banks. But they are also tackling unique challenges, such as the diversity of languages in India. Kotak is the first Indian bank to do voicebot on interactive voice response and in two languages, so far, that cover 70-80% of the population, with answers in real time.

“We have a couple of million customers communicating with us via WhatsApp, and we are building full integration into that,” Sharma added. “In terms of voice and vernacular, the challenge of being in a country with so many languages is that we cannot train employees to deal with them all, so we are building automated video generation and language support, which will pick up a lot more in the next 12 months.”

Sharma noted that talent in the field of AI and machine learning is still at a premium, but banks in India have been successful in attracting and retaining top talent in the AI field, second only to the big tech firms. This is due to their large and rich data sets which are getting used in relevant and impactful use cases. He added that the challenges in keeping pace are modernisation of infrastructure and maintaining the skills and knowledge to put together the tools for the future.

“We think that the next generation of AI will be looking at things including roboadvisory, overall customer engagement level, hyper-personalisation, both direct to customer and employees who serve customers in real time. In terms of risk management, there are inroads to be made in alternate credit score, underwriting and creating new products, and risk pricing using alternative data that is non-bureau,” he added.

How artificial intelligence is changing the face of banking in India?

India is the second-largest country in terms of population in the world after China. The country’s economy largely relies on robust financial infrastructure contributing to the growth of each sector in the nation. Since the financial industryis sufficiently capitalised and well-regulated, Indian banks are actively capitalizing on advanced technologies. No wonder artificial intelligence is leading the way into the country’s financial institutions, spotting atypical human behaviour, lessening operational costs and improving efficiency. AI is already having a vital impact on human life, transforming everything from how we live and work.

With soaring customer expectations and with an aim of delivering a better customer experience, financial services providers are implementing AI technology into banking operations. Artificial intelligence has the potential to detect frauds, mitigate uncertain risks, and help manage regulatory compliance. This article will let you go through the top Indian banks that are using AIand how are they benefiting from it.

State Bank of India

State Bank of India (SBI) is the largest public sector banking services provider in the country. To deliver effective banking services, the bank capitalizes on artificial intelligence. SBI Intelligent Assistant (SIA), an AI-powered smart chat assistant, addresses customer enquiries instantly and helps them with everyday banking tasks like a human does. Developed by an AI banking platform Payjo, this smart chat assistant is equipped to handle nearly 10,000 enquiries per second or 864 million in a day, which is almost 25% of the queries are processed by Google each day, reports noted.

HDFC

Headquartered in Mumbai, HDFC is another Indian banking and financial services firm that uses AI. The bank's smart chatbot called 'Eva' works with Google Assistant on millions of Android devices to solve customers' queries and provides them with better services. Built by Bengaluru-based Senseforth AI Research, Eva has reportedly claimed to have answered over five million user queries with more than 85% accuracy. HDFC also has an AI-enabled chatbot, OnChat, which launched on Facebook Messenger in 2016.

ICICI

ICICI Bank, a leading private sector bank in India, has applied software robotics in over 200 business processes across diverse functions of the company. Through this, the bank became the first in the country to deploy an AI system at a large scale in various processes. According to the report, ICICI bank has scaled its RPA initiative to over 750 software robotics handling nearly 2 million transactions daily, which is 20% of the transaction volumes.

Axis

Axis bank allows its customers to talk about their banking issues anytime, anywhere through an AI-powered bot. India's third-largest private sector bank in July 2020 unveiled a conversational interactive voice response (IVR) system, called AXAA. As a next-generation multilingual voice bot, AXAA assists customers to traverse through the IVR and addresses their queries and requests, without the need for any human intervention in most cases. The private lender also has an innovation lab called 'Though Factory' that aimed to expedite the development of innovative AI technology solutions for the banking sector.

Bank of Baroda

Bank of Baroda is another public sector lender advancing banking services and reducing the cost of managing accounts while focusing on improving customer service through AI. The bank uses advanced gadgets like an artificial intelligence robot named Baroda Brainy and Digital Lab with free Wi-Fi services. It also has a chatbot named ADI (Assisted Digital Interaction). In 2018, Bank of Baroda partnered with IBM and Accenture to power a state-of-the-art IT Center of Excellence (ITCoE) and Analytics Center of Excellence (ACoE).

Andhra Bank

Andhra bank is a medium-sized public sector bank of India that merged with Union Bank of India in April 2020. As the bank has a network of branches, with numerous satellite offices in the country, it has adopted advanced technology. The bank uses an AI interactive assistant named “ABHi” to address customer queries immediately and effectively. This AI chatbot, developed by Floatbot, is integrated with Core Banking Servers (CBS) of Andhra Bank and will automate customer support for five crore account holders of the bank.

Kotak Mahindra Bank
Kotak Mahindra uses a smart AI-enabled chatbot to power millions of Kotak customers with quick and available to answer banking queries round the clock. The chatbot, named Keya, is a bilingual voicebot that comes integrated with Kotak’s phone-banking helpline and will augment the traditional interactive voice response (IVR) system. The bank launched Keya 2.0 voicebot with new features in 2019.

The banking system in India far superior and well-regulated. As per the report, the asset of public sector banks stood at US$1.52 trillion in the fiscal year 2020. Furthermore, bank credit grew at a CAGR of 3.57% during FY16-FY20. As of FY2020, total credit extended reached US$ 1,698.97 billion. As the country’s banking ecosystem is relentlessly growing, the adoption of artificial intelligence will continue to evolve, enabling a digital banking infrastructure.

The Chinese central bank has issued a new set of assessment standards for artificial intelligence (AI)-based fintech applications.

The People’s Bank of China (PBOC) issued the “Evaluation Specification of Artificial Intelligence Algorithm in Financial Application” (人工智能算法金融应用评价规范) on 26 March, which came into effect immediately.

The Evaluation Specification is an “artificial intelligence financial application algorithm assessment framework” which “systemically outlines basic requirements, algorithms assessment measures and judgement rules for the application of AI to the financial sector,” and is applicable to financial institutions, algorithm providers and third party security assessment organisations.

With regard to AI algorithm assessment methods, the Specification highlights the use of report inspections, system inspections, personnel interviews, systems testing, shock testing, and algorithm testing, while also providing specific assessment content and judgement standards for security, interpretability and functionality.

Conclusion

I recommend that the Comptroller of the Currency, the Federal Reserve System, the Federal Deposit Insurance Corporation, the Consumer Financial Protection Bureau, and the National Credit Union Administration take steps to ensure compliance in FinTech with the OECD AI Principles, and the OSTP/OMB Guidance on Regulation of Artificial Intelligence Applications. I specifically recommend that the industry limit the scope of defenses for negligent and fraudulent parties whose actions have a legal or significant effect on an individual and discredit the commitment toward trustworthy AI.

Respectfully Submitted,

Susan von Struensee, JD, MPH