

# Ethics & Artificial Intelligence in Finance



## AUTHOR:

Alex LaPlante, Managing Director, Research, Global Risk Institute



GLOBAL  
RISK  
INSTITUTE

## Introduction

We are standing on the cusp of the fourth industrial revolution—the rise of the “intelligent machine.” At the heart of this revolution is Artificial Intelligence (AI), algorithms that allow machines to mimic human cognitive functions like learning, problem-solving, and decision-making. AI offers a plethora of benefits, including increased speed and efficiency, reductions in labour and resource costs, improved customer experience, and enhanced security. Numerous sectors, including automotive, healthcare, retail, and defense have already witnessed the game-changing impact of AI, and the financial services sector is no exception. AI has begun to replace or augment human decision makers in business lines across financial institutions, from predictive tasks like fraud detection and risk management to customer interactions like loan approval and wealth advisory. As technological capabilities continue to improve, data collection grows in scale and scope, and competitive pressure from non-traditional financial institutions increases, the use of AI in finance will only become more widespread.

With the advancement and adoption of AI set to continue to increase for the foreseeable future,

there have been growing concerns around the broader implications of such a transformative technology.<sup>1</sup> Unsurprisingly, this has led to calls for government regulation of AI development and restrictions on the use of AI technologies. Perhaps more telling, however, is that tech leaders themselves are voicing concerns. Elon Musk, for one, has widely claimed that artificial intelligence is the biggest existential threat to humans and has called for regulatory oversight at both the national and international levels.<sup>2</sup> Other notable tech figures like Bill Gates and Steve Wozniak, as well as leading minds like Stephen Hawking, have voiced similar concerns around the long-term risks of AI.<sup>(3, 4)</sup>

While there may be some merit to the notion that AI could cause the eventual demise of the human race, there is an active debate about a number of more immediate ethical concerns surrounding the pervasive deployment of the technology.<sup>5</sup> These concerns range from growing fears that automation will lead to mass unemployment and extreme income inequality, triggering widespread social unrest, to more philosophical questions like the infamous “trolley problem” in the context of driverless cars.<sup>6</sup> However, there are even more fundamental concerns around the

<sup>1</sup> Kaushal, M., Nolan, S., (2015) “[Understanding Artificial Intelligence](#)”, Brookings Institute

<sup>2</sup> Domonoske, C., (2017) “[Elon Musk Warns Governors: Artificial Intelligence Poses ‘Existential Risk’](#)”, NPR

<sup>3</sup> Mack, E., (2015) “[Bill Gates Says You Should Worry About Artificial Intelligence](#)”, Forbes

<sup>4</sup> Kharpal, A., (2017) “[Stephen Hawking says A.I. could be ‘worst event in the history of our civilization’](#)”, CNBC

<sup>5</sup> Bossman, J., (2016) “[Top 9 ethical issues in artificial intelligence](#)”, World Economic Forum

<sup>6</sup> The trolley problem is a thought experiment in ethics first introduced by philosopher Philippa Foot in 1967. It explores the moral dilemma that one would face should they find themselves in the following situation: There is a runaway trolley moving toward five tied-up people lying on the tracks. You are standing next to a lever that controls a switch. If the lever is pulled, the trolley will be redirected onto a side track, saving the five people on the main track. However, there is a single person lying on the side track. From an ethical standpoint, what should you do?

algorithms themselves and the data we use to train them.

Interpretability of a model's results and transparency about how those results were generated are critical in ensuring the alignment of the model with the problem at hand. AI algorithms, which are often referred to as black-box approaches, suffer from a lack of transparency and interpretability, making it difficult to parse out how and why these algorithms come to particular conclusions. As a result, the identification of model bias or discriminatory behaviour can be challenging. In fact, even world leaders in AI like Google have had significant missteps with unintentional bias. In 2015, for example, Google found itself in hot water when it employed an image recognition algorithm that classified people of colour as gorillas. Even after this incident, there has been no shortage of high-profile cases of discriminatory AI algorithms.

Adding an additional layer of complexity is the use of non-traditional data sources like social media and IoT (Internet of Things) technologies. For one, the use of certain types of data such as race, disability status, and religious affiliation may be seen as unethical in and of itself, and may raise questions around a customer's data usage and privacy rights. Moreover, unrepresentative or systematically inaccurate data sets can be another key source of bias.

From a financial institution's perspective, AI offers a wide range of potential benefits, but one must ensure that the implementation of this technology is both prudent and ethical. To aid in the understanding of how this can be achieved, this report will introduce AI and detail issues of bias, interpretability, and data security and privacy as they relate to the ethical use of AI algorithms by financial institutions. It will also discuss the key risks that can arise from the

unethical use of AI and the considerations that should be made in order to manage these risks throughout the development and implementation stages.

## What is AI?

From Hephaestus' golden robots in ancient Greek mythology to the droid C-3PO in George Lucas' *Star Wars*, the concept of artificial intelligence is a recurring theme throughout history. Recently, however, AI has moved beyond the realm of the mythical and into the realm of the practical, with applications emerging across society. AI, which generally refers to a branch of computer science that aims to create intelligent machines, focuses on programming computers for capabilities like pattern recognition, self-directed learning, problem-solving, and decision-making.

Today, we interact with Narrow AI or Artificial Narrow Intelligence (ANI), which is programmed to perform a pre-determined task or set of tasks. ANI systems receive input data from pre-specified sources and can learn and perform their programmed tasks in real-time. They are not, however, able to learn or perform tasks that fall beyond those for which they were specifically designed, or seek out new sources of input data. For example, an algorithm designed for high-frequency stock trading cannot teach itself how to make targeted content recommendations on Netflix. ANI lacks the fluidity and flexibility of human intelligence, is not conscious or sentient, and operates within a pre-defined range. In other words, ANI cannot think for itself.

Even today's most sophisticated AI systems that are able, in some ways, to interact with their environments, like Amazon's Alexa or driverless cars, are nowhere close to having human-like intelligence. Nevertheless, impressive advancements in ANI continue to be made,

providing the likely building blocks necessary to one day enter the next technological paradigm—General AI.<sup>7</sup> Artificial General Intelligence (AGI), which is often showcased in Hollywood sci-fi movies (e.g. *Ex Machina* or *Her*), refers to AI that is capable of human-level intelligence. AGI systems are those that have self-awareness, sentience, and consciousness, and are able to reason, problem-solve, innovate, and create flexibly across various fields. Unsurprisingly, expert opinions vary widely when it comes to predicting when AGI will become a reality. Some commentators, like futurist Ray Kurzweil, believe General AI is only a decade away, while many others believe the leap from ANI to AGI will take several decades or more to achieve.<sup>(8, 9)</sup>

And then there is the “Singularity”, the point in time when humans are no longer the most intelligent beings on earth. This point will be defined by the achievement of Artificial Super Intelligence (ASI), where, according to philosopher Nick Bostrom, “machine intellect will greatly exceed the cognitive performance of humans in virtually all domains of interest.”<sup>10</sup> ASI is what many prominent thinkers believe could lead to the extinction of the human race.

Although there is a long way to go before either AGI or ASI are realized, the past decade has seen many noteworthy advancements in AI’s underlying technologies. On top of vast improvements in computer processing and data storage, significant progress has been made in hardware for operations like sensing, perception, and object recognition. This has led to important developments in several fields of AI including computer vision, natural language processing, and robotics.<sup>11</sup> Perhaps most notable, however,

has been the evolution of machine learning, which seeks to enable computers to learn from data—and act based on what they learn—without following pre-programmed rules. With recent developments in a subset of machine learning known as deep learning, the utility of the field has reached even greater heights.<sup>12</sup>

Many believe that AI will completely transform most industries, just as electricity did 100 years ago.<sup>13</sup> In fact, applications of AI are already playing an active role in our everyday lives, from traffic predictions on Google Maps and targeted web-based advertising, to intelligent customer relationship management (CRM) platforms and in-hospital treatment recommendation systems.

The financial services sector has also proven fertile ground for AI applications. One commonly-cited example is algorithmic trading in capital markets, which uses AI to track real-time trade data and enter and exit positions automatically, but AI has been applied to many other aspects of the financial services industry as well. Process automation, which uses AI to perform or facilitate repetitive manual tasks, is being used to automate customer service, employee training, and document processing. JPMorgan Chase’s Contract Intelligence platform (COiN), for example, processes commercial credit agreements and other legal documents several orders of magnitude more quickly than human workers. Security and fraud detection is another promising arena for AI in finance. Many banks, as well as alternative payment processing systems like PayPal and Stripe, now use AI to monitor credit transactions in real time and detect anomalous transactions with high precision. And robo-advisory, which uses AI to automate asset

<sup>7</sup> Saenz, A., (2016) “[How Today’s Jungle of Artificial Intelligence Will Spawn Sentience](#)”, Singularity Hub

<sup>8</sup> Reedy, C., (2017) “[Kurzweil Claims That the Singularity Will Happen by 2045](#)”, Futurism

<sup>9</sup> Urban, T., (2015) “[The AI Revolution: Our Immortality or Extinction](#)”

<sup>10</sup> Bloom, N., (2014) “Superintelligence: paths, dangers, strategies”, Oxford University Press

<sup>11</sup> Stanford, (2016) “[AI Research Trends](#)”

<sup>12</sup> Parloff, R., (2016) “[Why Deep Learning is Suddenly Changing Your Life](#)”, Fortune

<sup>13</sup> Lynch, S., (2017) “[Andrew Ng: Why AI Is the New Electricity](#)”, Insights, Stanford Business

management and the recommendation of financial products, can deliver personalized portfolio allocations at a fraction of the cost of human financial advisors.

The financial services sector is deeply intertwined with every segment of our economy, and thus AI innovations have the potential to bring wide-ranging benefits. At the same time, however, the unique role that financial services play in our society and in our everyday lives means that particular care must be taken in this sector to avoid the ethical pitfalls that can arise from real-world applications of AI.

## Ethics and AI

AI systems are not only leveraged to assist with relatively superficial decision-making, like what Netflix show to watch or what stock to pick, but are increasingly being used to make critical life decisions, like what healthcare treatment we receive or whether we are extended a mortgage loan. Given the importance of these decisions and the growing pervasiveness of AI, these technologies can quickly reinforce or alter power structures on an unprecedented scale and to the detriment of human rights. There is a growing body of evidence that AI, given its relative opacity, can contribute to unethical and discriminatory practices if developed and employed without taking the necessary precautions.<sup>14</sup>

There is, unfortunately, no shortage of examples of ethically-questionable applications of AI technologies around the globe. China's Social

Credit System, which is intended to standardise the "assessment" of citizens' or businesses' economic and social reputations, has been the focus of much consternation and debate.<sup>15</sup> The system, which is due to be fully implemented by 2020, employs mass surveillance and big data analysis to provide citizens and businesses with social credit scores which are then used to determine a set of rewards or punishments. Those with high scores have easier access to loans and jobs and have priority in bureaucratic paperwork. While those with low scores or those associated with other low scoring individuals or entities may be subject to disadvantages like slow internet connections, exclusion from prestigious work or schools, travel bans, and registration on a public blacklist.<sup>16</sup> Whatever the intention behind its implementation, this system has clear potential to significantly increase China's already-growing inequality and has far-reaching implications for both personal and economic freedoms.<sup>17</sup>

There are also a number of examples much closer to home. Google's Project Maven, a collaboration with the U.S. Department of Defense (DoD) that sought to use AI to improve drone strike targeting, was cancelled after thousands of Google employees signed an open letter to the company's management condemning the project due to ethical concerns around weaponizing AI. The letter stated "[t]his contract puts Google's reputation at risk and stands in direct opposition to our core values."<sup>18</sup> Although Google terminated the project, the DoD has since expressed its firm desire to continue researching applications of AI for defense purposes.<sup>19</sup> AI

<sup>14</sup> Amnesty International, Access Now, (2018) "[The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems](#)"

<sup>15</sup> General Office of the State Administration of Taxation: The People's Republic of China, (2014) "[SAT Boosted the Construction of Credit System and Practiced Reward and Punishment Based on Two Measures](#)"

<sup>16</sup> Chin, J., (2016) "[China's New Tool for Social Control: A Credit Rating for Everything](#)", Wall Street Journal

<sup>17</sup> Wang, M., (2017) "[China's Chilling 'Social Credit' Blacklist](#)", Human Rights Watch

<sup>18</sup> Shane, S., Wakabayashi, D., (2018) "['The Business of War': Google Employees Protest Work for the Pentagon](#)", The New York Times

<sup>19</sup> Fryer-Biggs, Z., (2018) "[Inside the Pentagon's Plan to Win Over Silicon Valley's AI Experts](#)", Wired

technologies have also made their way into police departments and judicial systems in the US and the UK.<sup>20</sup> Once such system that has come under great scrutiny is COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), a tool built to forecast a criminal defendant's propensity for recidivism. It was found that COMPAS' estimates of recidivism rates are racially biased; the system over-predicts recidivism rates for black offenders and under-predicts for white offenders.<sup>21</sup>

Although there have yet to be any widely-publicized cases of unethical applications of AI by financial institutions, concerns have been raised around a number of problems that could arise when the sector incorporates the technology.<sup>22</sup> Biased AI systems could lead to unequal access to financial services, unfair lending practices, and discriminatory product pricing. Furthermore, the use of alternative data by AI systems creates serious data security issues and may violate consumer privacy rights. As is the case for firms in other sectors of the economy, there are also material concerns around responsibility for negligent decisions made by AI systems and accountability for the resultant consequences.

## Drivers of Ethical Concerns in Financial Institutions

The ethical challenges in applying AI to finance stem from the inherent characteristics of AI algorithms and the troves of data on which they rely. For one, the opacity of AI systems makes the detection and elimination of unwanted bias far more challenging than when using traditional statistical methods. For another, the alternative data sources on which these systems thrive are prone to issues like misrepresentation, historical

bias, and intentional manipulation, and pose unique challenges for data privacy and security.

The complexity and flexibility of AI systems make them significantly more opaque than traditional analytical frameworks. While the latter follow pre-determined rules to estimate specific forms of statistical relationships, the former can seek out virtually any kind of relationship between features in the data on which they are trained. An AI system's lack of interpretability is further increased by its adaptability. Because these systems continue to learn as their training datasets grow, the statistical relationships they uncover can grow more complex over time; the smartest AI systems are often the least transparent. Moreover, the outputs generated by an AI system from a given set of inputs can change over time, so understanding the drivers of that system's decisions is a constant battle.

AI systems' lack of interpretability exacerbates the ethical issues surrounding their use in financial institutions, hindering analysts from detecting discrimination against certain classes of people and from determining whether personal data is used in ways that violate ethical boundaries. Not only do financial institutions have to meet stringent regulations often requiring high levels of transparency, but they owe their customers the assurance of fair and equitable treatment. Consequently, financial institutions have an obligation to understand the decisions of their AI-based systems and the potential adverse effects that these decisions may have on their customers.

AI algorithms can exploit a variety of alternative data sources that traditional analytical techniques are not well-equipped to handle. Financial institutions, having access to a rich set of alternative data sources like credit and debit

<sup>20</sup> Hao, K., (2018) "[Establishing an AI code of ethics will be harder than people think](#)", MIT Technology Review

<sup>21</sup> Larson, J., Mattu, S., Kirchner, L., Angwin, J., (2016) "[How We Analyzed the COMPAS Recidivism Algorithm](#)", ProPublica

<sup>22</sup> The Financial Innovation Lab (2018) "[Briefing: Ethical Use of AI in Finance](#)"

transaction histories, are well placed to use these methods. That being said, there are serious questions about the ethics of employing sensitive personal information in these models. Should consumers have the right to obtain confirmation that their personal data has been used in automated decision-making? Should consumers be able to opt out of these sorts of decisions? Regulators around the world have already begun to grapple with these issues, and some jurisdictions like the European Union have already begun to enact laws explicitly enumerating these sorts of rights. Furthermore, institutions that use sensitive personal information are naturally responsible for securing these data during all stages of processing, analysis, and storage. Financial institutions that use these data sources must also ensure that the data are accurate, particularly when they have been gleaned from external sources or provided directly by customers, like social media posts or surveys.

The complexity of these technologies and the broad array of data they exploit can enable them to make more accurate predictions than traditional analytical techniques, but these same traits also make these technologies particularly susceptible to bias. Algorithmic bias does not arise spontaneously; rather, it is a product of biased training data or biased programming. Training data may be biased when they are incomplete, unrepresentative, reflect historical discrimination, or contain underlying societal biases. For example, in 2014, Amazon developed an algorithm to automate its recruiting process, but found that it was biased against women because of the predominately male resume pool on which the algorithm was trained.<sup>23</sup> The lack of female representation in Amazon's training data was likely due, at least in part, to historical male dominance in the tech sector. If left unchecked,

algorithmic bias of this sort can perpetuate, or even exacerbate, underlying historical biases. Developers of AI systems can also introduce bias when processing training data, specifying objective functions, and designing learning behaviours. If, during any of these stages, a developer's choices are influenced by his or her own biases, those biases may inadvertently influence the predictions his or her systems make.

When biased algorithms are used to determine how a financial institution interacts with its customers, significant ethical problems can arise. For example, consider a lender that employs an AI system that uses natural language processing to evaluate loan application forms and make credit decisions. If the system found that people who make grammatical errors or spelling mistakes in their applications are more likely to default, it would deny credit to these individuals even though these habits have no causal effect on loan repayment. Because these habits are likely to be correlated with an individual's level of education, this system's actions could expose the lender to accusations of discrimination on the basis of education. Even more problematic, if the system found that the usage of words or phrases common among ethnic minorities was correlated with default, the lender could be held liable for racial discrimination.

## Ethics-Related Risks

The ethical concerns surrounding machine learning and artificial intelligence highlighted above interact to create a unique set of reputational, legal, and operational risks that merit careful consideration by any financial institution seeking to use these technologies to enhance their capabilities.

---

<sup>23</sup> <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>



Algorithmic bias poses significant reputational risks for institutions that rely on machine learning, AI, and related analytical tools. The public perception of a company can decline sharply when ethically dubious analytics are exposed. Microsoft, for example, has received backlash for building facial recognition software for U.S. Immigration and Customs Enforcement (ICE) amidst outrage over the agency's policy of separating undocumented immigrant children from their parents. These reputational risks can have significant pecuniary side effects, but algorithmic bias also poses legal risks that have direct financial consequences.

In the United States, the Federal Trade Commission (FTC), the Department of the Treasury, and the White House each recently published reports about legal issues stemming from algorithmic bias, particularly for financial institutions that use algorithms to make credit determinations. These reports explicitly outline scenarios in which institutions could find themselves liable for unintended, but nevertheless illegal, discrimination caused by their algorithms' behavior.<sup>24</sup> The Office of the Comptroller of the Currency (OCC) has also raised concerns about algorithm-driven discrimination in lending among Fintech companies.<sup>25</sup>

Additionally, the disparate-impact theory of liability has recently gained favour in U.S. courts. Under this theory, companies that discriminate, even unknowingly or unintentionally, against protected classes can be held liable for their conduct. In other words, effect, not intention, is what matters. In *Texas Department of Housing and Community Affairs v. Inclusive Communities Project*, for example, in which a non-profit sued a Texas agency for perpetuating racial bias in access to housing, the U.S. Supreme Court held

that plaintiffs needed to show only that the agency's policies had discriminatory effects, not that the agency intended to cause this discrimination.<sup>26</sup> Similar outcomes could befall banks that rely on biased analytics to deny mortgage loans more often to some groups of people than others. In this context, algorithms that use data that reflect historical discrimination towards certain groups may be particularly problematic.

More broadly, policy makers around the globe are beginning to take active measures to protect consumers from inappropriate data usage and the risks of AI, posing new regulatory risks for financial institutions. The General Data Protection Regulation (GDPR), a comprehensive data protection and privacy law enforced in the European Union, imposes a number of obligations on financial institutions and other firms that employ AI tools to make customer-related decisions. For example, prospective borrowers in the E.U. can request an explanation of the factors that drive automated lending decisions; banks and other creditors are obligated to explain the determinants of a loan denial.<sup>27</sup> As was showcased by the recent €50 million fine issued to Google by CNIL, France's data protection regulator, failure to follow these regulations carries with it the real potential for financial loss.<sup>28</sup>

In addition to the reputational, regulatory and legal risks that stem directly from the ethical concerns surrounding AI, bias and lack of interpretability also pose significant model-related operational risks. Model risk is not unique to AI, of course, but the opacity of this technology makes validating models that use it particularly challenging; detecting and correcting over- or under-fitting, for example, is significantly more

<sup>24</sup> Federal Trade Commission, (2016) "[Big Data: A Tool for Inclusion or exclusion?](#)"

<sup>25</sup> Office of the Comptroller of the Currency, (2016) "[Exploring Special Purpose national Bank Charters for Fintech Companies](#)"

<sup>26</sup> Petrasic, K., Saul, B., (2017) "[Algorithms and bias: What lenders need to know](#)", White&Case

<sup>27</sup> Kaminski, M., (2019) "[The Right to Explanation, Explained](#)", Berkeley Technology Law Journal, Vol. 34, No. 1

<sup>28</sup> Porter, J., (2019) "[Google fined €50 million for GDPR violation in France](#)", The Verge

difficult for AI-based models than for traditional statistical models.<sup>29</sup> Moreover, AI-based models evolve constantly as their training datasets grow, and so these models must be validated continuously in order to ensure their accuracy.<sup>30</sup>

## Engagement, Management, and Mitigation

In response to the numerous ethical concerns about AI, governments, universities, and even financial institutions have convened working groups to study these issues and determine best practices for the responsible use of this technology. Working with Harvard's Kennedy School of Government, Bank of America has formed the Council on the Responsible Use of Artificial Intelligence, which will bring together world-renowned experts to determine how to ensure privacy, equality, and transparency without detracting from the benefits of AI.<sup>31</sup> A number of prominent tech companies have also proposed best practices for the responsible use of this technology. Google, for example, recommends a human-centred design approach, the use of multiple metrics to assess training and monitoring, and extensive examination of training data to detect possible sources of bias and to understand the limitations of any AI systems that learn from it.<sup>32</sup> Canadian institutions, too, are working to ensure that the use of AI is guided by ethics. The Canadian Treasury Board has recently issued a directive that outlines the responsibilities of federal institutions that use AI to assure quality, transparency, and public accountability, while the Toronto Declaration, created by Access Now

in collaboration with Amnesty International, articulates measures that companies can take to ensure human rights are respected in the development and use of AI.<sup>33</sup>

The research community has also been working to provide methodologies that can help reduce bias in AI systems and make them more transparent. Explainable AI (XAI), also known as Interpretable AI or Transparent AI, aims to produce AI models that human users can understand and trust while still delivering accurate predictions. An explainable classification algorithm for example might, in addition to simply determining the probability that an object belongs to a particular group, articulate the features the object shares with other members of that group. Further, a number of research groups have developed methodologies to pre-process training data to reduce bias and, in turn, reduce discrimination by AI systems that learn from that data. IBM Research's approach, for example, ensures that people have the same chance to receive favorable treatment irrespective of their membership in protected classes, while disrupting each individual's data as little as possible and still enabling the AI to learn appropriately.<sup>34</sup>

It is critical that financial institutions that are looking to implement AI have a clear set of guidelines and safeguards in place to ensure the responsible development and use of the technology. This, of course, should encompass more than just fairness and explainability, but should also consider the robustness and lineage of the algorithm. To begin, financial institutions must ensure that employees who are tasked with

<sup>29</sup> LaPlante, A., Rubtsov, A., (2019) "[Artificial Neural Networks in Financial Modelling](#)", Global Risk Institute

<sup>30</sup> Osborn, T., (2019) "[Model risk chiefs warn of machine learning bias](#)," Risk.net

<sup>31</sup> Harvard Kennedy Center (2018) "[Harvard Kennedy School and Bank of America Announce The Council on the Responsible Use of Artificial Intelligence](#)"

<sup>32</sup> Google, "[Responsible AI Practices](#)"

<sup>33</sup> Government of Canada (2019) "[Directive on Automated Decision-Making](#)"

<sup>34</sup> Varshney, K., (2018) "[Reducing Bias in AI with New Methodology](#)", IBM Research Blog



either processing the data or developing, validating, and overseeing AI models have sufficient qualifications and experience to competently execute their responsibilities. Additional training on ethical considerations and social risks may also be warranted. Next, there must be clear guidelines for data management and usage practices as they relate to AI. There should be efforts made to understand what social and historical biases may exist in the data and how to adequately correct for them. Ultimately, the industry should commit to ongoing measures that improve the quality and consistency of their data and analytics.

When it comes to the actual development and use of AI technologies, financial institutions must have the capacity to evaluate the impact of automated decision-making systems, assess procedural fairness, and weigh the trade-offs between complexity, accuracy, and interpretability. In service of these prerequisites, financial institutions can deploy questionnaires or factsheets for completion by data engineers and AI developers that provide increased transparency and more effective oversight. These sheets should include information such as system operation, training data, underlying algorithms, performance benchmarks, fairness and robustness checks, intended uses, and maintenance and re-training instructions, and should be revised every time the AI system is modified.<sup>35</sup> Moreover, financial institutions must have internal policies and governance mechanisms in place to ensure that algorithmic monitoring and risk mitigation procedures are adequate and transparent, and are regularly reviewed and updated as the field of AI progresses.

## Conclusion

As AI continues to evolve and proliferate, there is no doubt that financial institutions will seek to capture the associated benefits of this transformative technology. However, given the critical role they play, both as a pillar of the economy and a keystone in our everyday lives, financial institutions have a duty to ensure that their use of AI is fair, equitable, and non-discriminatory. Though the use of AI poses real ethical concerns for financial institutions, as well as legal, regulatory, model and reputational risks, many institutions around the globe have already begun working on frameworks and best practices to ensure the responsible implementation of the technology. Indeed, financial institutions must act prudently and proceed with caution in this budding era of AI, but few, if any, can afford to sit back and ignore the technology all together.

---

© 2019 Global Risk Institute in Financial Services (GRI). This "Ethics & Artificial Intelligence in Finance" is a publication of GRI. This "Ethics & Artificial Intelligence in Finance" is available at [www.globalriskinstitute.org](http://www.globalriskinstitute.org). Permission is hereby granted to reprint the "Ethics & Artificial Intelligence in Finance" on the following conditions: the content is not altered or edited in any way and proper attribution of the author and GRI is displayed in any reproduction. All other rights reserved.

---

<sup>35</sup> Mojsilovic, A., (2018) "[Factsheets for AI Services](#)", IBM Research Blog