

**Deloitte.**



**Navigating uncharted waters**

Understanding the strategic and regulatory risks of artificial intelligence in financial services


## More from the Forum and Deloitte

The future of financial services (2015)

The future of financial infrastructure: An ambitious look at how blockchain can reshape financial services (2016)


Disruptive innovation in financial services: A blueprint for digital identity (2016)

Beyond Fintech: A pragmatic assessment of disruptive potential in financial services (2017)

The new physics of financial services: How artificial intelligence is transforming the financial ecosystem (2018)

The next generation of data-sharing in financial services: Using privacy enhancing techniques to unlock new value (2019)

Navigating uncharted waters: A roadmap to responsible innovation with AI in financial services (2019)



# Introduction

Introduction	4
The AI challenge	5
AI explainability	7
Systemic risk and AI	9
Bias and fairness	12
Other uncertainties	14
The algorithmic fiduciary	14
Algorithmic collusion	14
Concluding thoughts	16

# Introduction

Dear colleagues,

Since 2015, Deloitte Consulting LLP (Deloitte) has worked with The World Economic Forum (the Forum) to gauge the forces of change in financial services. Eventually—or maybe inevitably—this endeavor led the Forum to examine the [role of artificial intelligence \(AI\)](#) in financial services.

At the same time, we wondered: How can financial institutions reap the benefits of AI while effectively navigating the risks it introduces? That question formed the basis of the Forum report, [Navigating uncharted waters: A roadmap to responsible innovation with AI in financial services](#).

Our search for insight lasted over ten months. We spoke with more than 250 subject matter experts around the world, including executives of many well-known financial institutions. We also held a number of in-person workshops around the world.

Throughout these activities, three common concerns emerged:

- What do the various stakeholders need to know about how AI systems make decisions?
- Could algorithms destabilize the financial system?
- How can institutions keep bias out of their AI solutions?

To understand these and other concerns better, we explored the underlying risks and uncertainties, the opportunities to be captured by managing them, and leading examples of how institutions are addressing them today. We hope the Forum report, summarized here, sheds some light on the responsible use of AI in financial services and helps to steady your organization's course through the uncertainties ahead.

Sincerely,

**Bob Contri**

Financial Services Industry Leader,  
Deloitte Global  
[bcontri@deloitte.com](mailto:bcontri@deloitte.com)

**Rob Galaski**

Global Leader, Banking & Capital Markets,  
Deloitte Consulting  
[rgalaski@deloitte.ca](mailto:rgalaski@deloitte.ca)

# The AI challenge

The rise of AI has brought financial services into uncharted waters. In this new expanse, early navigators may realize outsized gains. But the voyage is marked with risks and uncertainties.

A smooth passage into an AI-powered future depends on the industry's ability to:

**Responsibly deploy AI systems today.** AI systems can learn on their own and 'think' in a way that's deeply foreign to humans. This inscrutability can lead to unfair treatment of customers, systemic instability, and other problematic outcomes—especially as AI systems interact with one another at greater frequency and speed.

In response, the financial sector may have to shift to new forms of governance. This could include safeguards that detect and prevent AI accidents, mechanisms to remediate accidents when they do happen, and an ethics process that provides for other intervention as needed. Beyond that, the systems themselves may need some form of auditability specific to each use case.

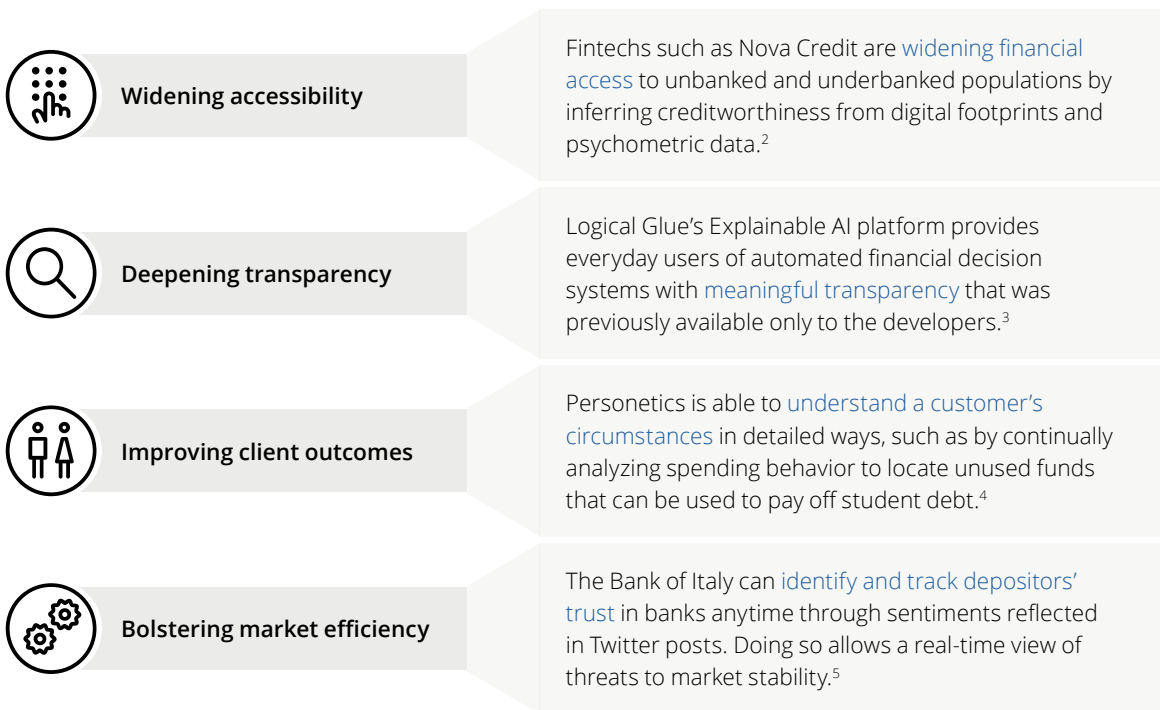
**Responsibly scale the AI-ubiquitous financial systems of tomorrow.** The use of AI is spurring a cross-industry re-examination of competition policy, privacy rights, and operational resilience. The common thread through this activity? Data. Institutions are increasingly trading in data. Large data stores are contributing to a winner-take-all environment for data-hungry AI systems. Among customers, meanwhile, there's a growing awareness of how valuable their data can be.

How these dynamics play out will shape the ways financial institutions can deploy AI and, more broadly, the strategic choices at their disposal. The potential effects extend beyond the use of data to include changes in market participants, allowable business models, regulations, and more. If policy decisions are to appropriately reflect the implications for financial institutions, the proactive engagement of industry leaders might be in order.

**Harness the potential of a financial system built on responsible AI.** The responsible use of AI isn't just about avoiding harm. It's also an opportunity to raise the ethical bar for the financial system as a whole. Some organizations already are using AI to create financial offerings that serve clients and society in better ways (Figure 1).

The highly regulated nature of financial services could prove a silver lining where it comes to AI. Consumers are paying closer attention to their data rights and digital sovereignty, putting technology giants on the defensive and giving financial institutions a chance to seize the advantage from having [higher rates of customer trust](#).<sup>1</sup> The result could be an entirely new value proposition for financial services customers.

Figure 1: How to raise the ethical bar for the financial system with AI



So what will it take to accomplish all this? Among other things, industry leaders likely will have to work with regulators and policymakers to address pressing concerns about the use of AI in financial services. Of these concerns, one of the most prominent is a lack of visibility into how AI systems produce their results. Another is AI's potential to destabilize the financial system. There's also the possibility that AI could exacerbate unfair bias in financial decision-making. Added to these are two less commonly-discussed uncertainties: the algorithmic fiduciary and algorithmic collusion.

These are some of the key barriers to widespread AI implementation in the financial service industry today. In order to overcome them, however, we first need to understand them. In that spirit, let's unpack each of them in turn.

**We heard five primary concerns about the use of AI in the financial sector:**

 <p><b>AI explainability</b></p> <p>How does business context shape what we need to know about our AI deployment?</p>	 <p><b>Systemic risk and AI</b></p> <p>Could algorithms destabilize the financial system?</p>	 <p><b>Bias and fairness</b></p> <p>How can institutions ensure their systems do not discriminate against a specific group?</p>	 <p><b>The algorithmic fiduciary</b></p> <p>Could AI be trusted as a fiduciary?</p>	 <p><b>Algorithmic collusion</b></p> <p>How can we manage AI systems that learn to engage in anti-competitive behavior?</p>
--	--	--	--	--

# AI explainability

With deep learning and other advanced approaches, AI is becoming significantly more complex. Partly this is because an AI system can consider a high volume of inputs, and it isn't always obvious which ones have the most influence on the output. Beyond that, today's systems often have hundreds of processing steps between data inputs (e.g., name, income, and credit score) and the resulting outputs (e.g., offer a \$5,000 loan at 12% interest). This can make it difficult to follow the logic of an AI-generated decision.

Other challenges are more specific to the nature of cognitive technology. Since AI systems adapt themselves over time, past decisions aren't necessarily predictive of future decisions. And this one is key: AI doesn't operate the same way as human intelligence. The judgment calls people make may have little bearing on an AI-generated decision, making it that much harder to understand why the system behaves the way it does.

## Informed trust

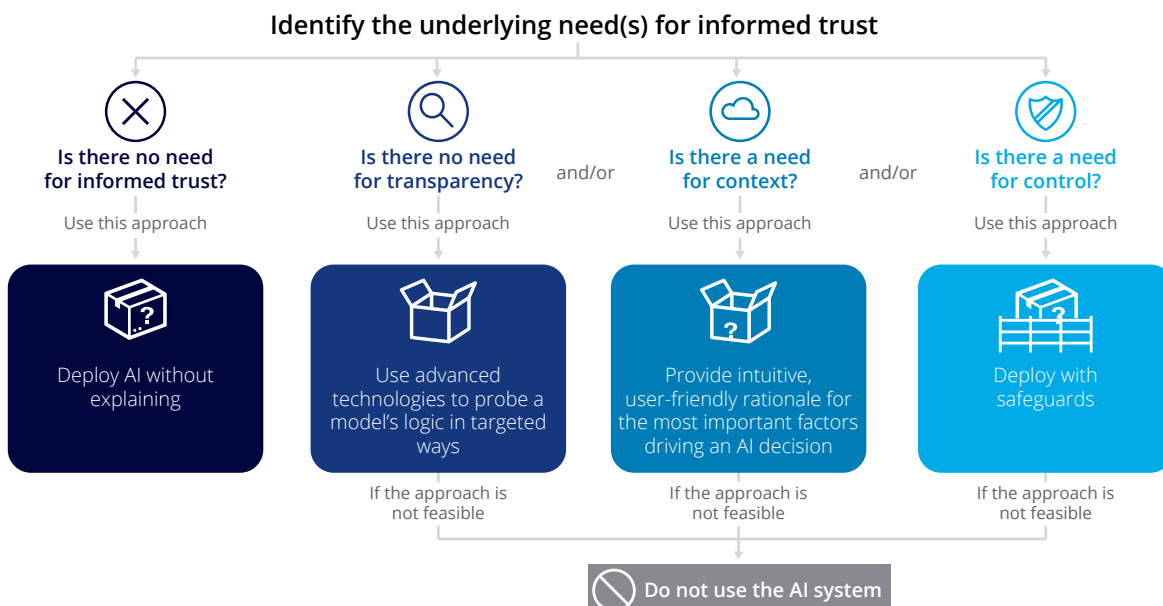
All of this can turn an AI system into a "black box," even to its developers. This has left financial institutions and regulators searching for middle ground—a place somewhere between blind faith in the AI system (which is too risky) and dismissal of all it has to offer (which is not innovative). Equilibrium, it seems, is a function of providing users with enough information about an AI system that they can trust its decisions.

What would it take to create this level of informed trust? It depends on the situation. In some cases—think personalized marketing promotions—no explanation may be necessary. Other cases may require:

- **Transparency** so that users understand how the AI system arrives at its decision
- **Context** so that users understand why the system makes a specific choice
- **Control** so that users know an AI-generated decision is in line with norms, laws, and business requirements

In other words, there are different reasons why an AI system may need to be explainable to users, and each reason requires a different approach to the underlying technology (Figure 2).

Figure 2: Approaches to managing explainability in an AI system



It's important to note that explainability isn't a one-and-done achievement. AI technology is changing too fast for that. Instead, financial institutions will have to periodically re-evaluate their AI strategy over time. It's already possible to imagine the changes ahead. Currently, regulatory acceptance is the main impetus for explainability. But as more AI solutions draw on sensitive data such as facial or speech recognition, explainability will become a condition for consumer acceptance and a differentiator for conveying improved consumer outcomes.

### Key implications for executives



AI systems will need to explain themselves in ways that enable meaningful risk management and accountability across interconnected institutions.



Depending on the context, only specific types of informed trust may be necessary.



Industry dialogue is important to establish thresholds for how much explanation is enough.



Gaining customer or employee trust via AI explainability could become a strategic choice, requiring deeper investment from financial institutions.



Explainability won't necessarily prevent real-world bias from becoming codified in algorithms. (More on this later.)



# Systemic risk and AI

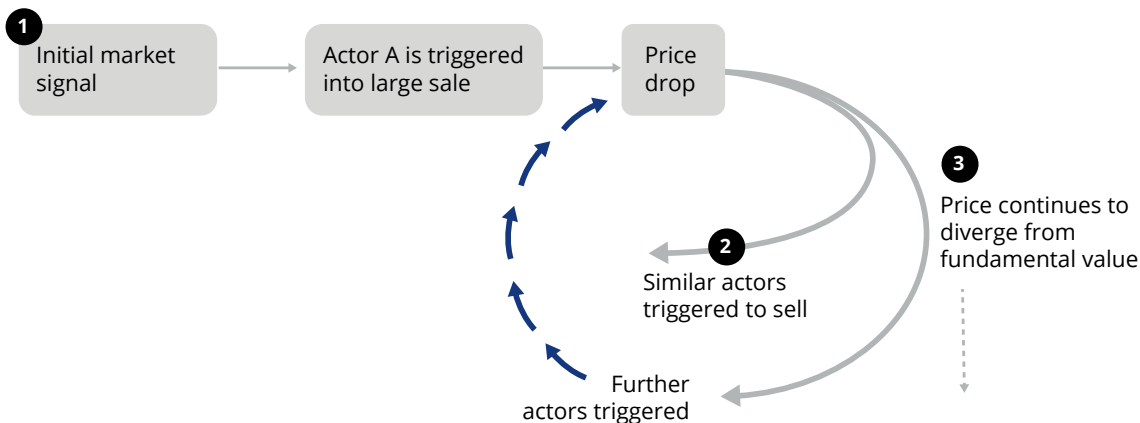
Before the 1970s, investors had a relatively level playing field in the financial markets. Computerized trading changed all that, giving the advantage to large, specialized players that could afford direct access to automated exchanges. Their lightning-fast trades often exacerbated market volatility, plunging market prices in minutes before other investors could even react.

## New sources of risk

Today, AI is poised to transform the financial markets once again. AI can be applied to a wide variety of use cases across every subsector of financial services, causing humans and machines to interact more frequently. At the same time, AI's black-box effect can make it harder for investors to interpret changing market dynamics and identify or understand emerging risks. Together, these conditions could amplify systemwide risks via:

**Herding.** Herding occurs when financial actors interpret a market signal in a similar way. For instance, they could all react to a news event by selling assets. The mass selloff causes asset prices to drop, which in turn drives further selling (Figure 3).

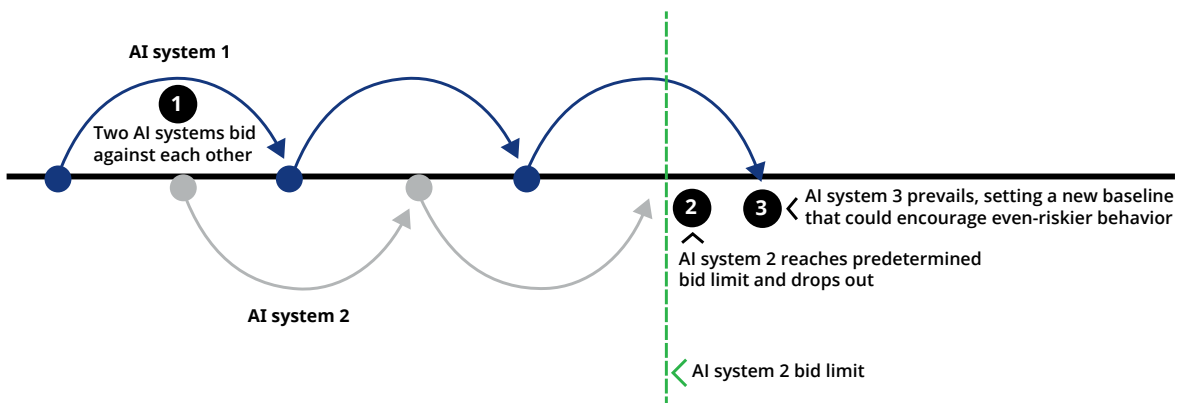
Figure 3: Herding behavior



Mechanisms like circuit breakers, which let participants pause and separate market signals from algorithmic noise, can mitigate AI-based herding. Diversity in AI design can also help, although this likely will require the industry to organize and enforce a robust set of AI development standards.

**Algorithmic competition.** Suppose two AI systems continuously bid against each other, optimizing their actions to achieve the highest market price. As they compete, the average market price continues to rise until one of the systems is forced to drop out due to profitability constraints. Sounds efficient? It is—but it could also threaten balance sheets by encouraging participants to either engage in riskier behavior to achieve profitability or stay out of the market altogether (Figure 4).

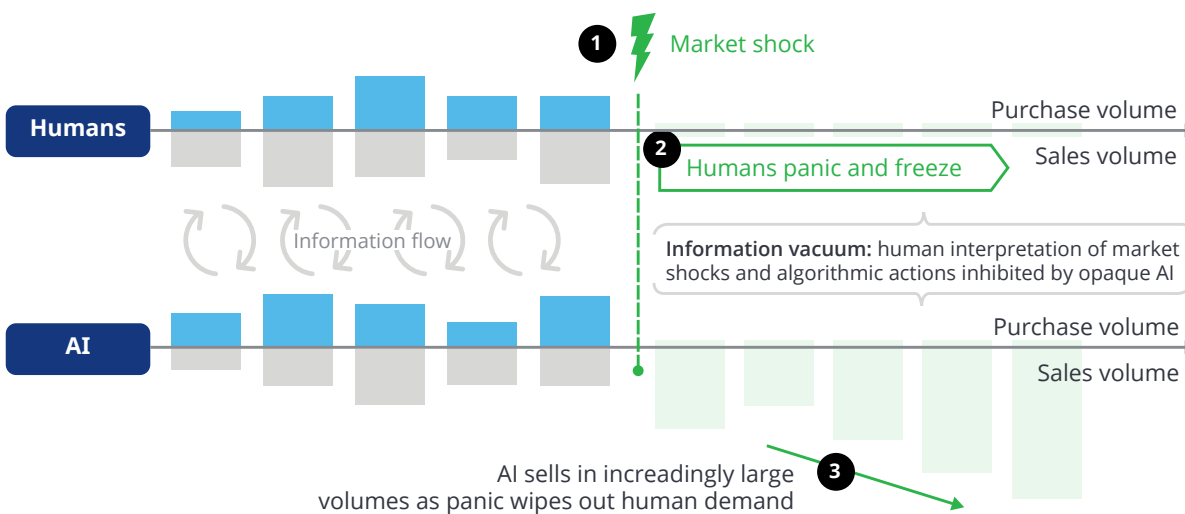
Figure 4: Algorithmic competition



Scenario modeling could help identify where AI's destabilizing behaviors are likely to occur. But managing the behaviors may have to involve a baseline level of diversification in allocations or limits on how fast self-driving AI can reallocate assets.

Information vacuums. A typical scenario is when automated systems are triggered into a mass selloff. As prices go into free fall, human investors struggle to interpret them and—in the absence of information—freeze. Algorithms interpret their uncertainty as disinterest and an indication to sell at lower prices. Markets remain out of equilibrium as long as these informational asymmetries persist (Figure 5).

Figure 5: Human panic in an information vacuum



Circuit breakers can interrupt feedback loops rooted in algorithmic noise by pausing all activity so that participants have time to verify their information sources. That said, a significant near-term priority likely will be to design news and sentiment analysis AI for resiliency against misread or falsified information.

### Erosion of existing defenses

Even as AI introduces new sources of risk, the traditional bulwarks of financial stability could fall away. For instance, today the financial system relies on human judgment to detect and prevent crises. By relying on AI, people may gradually lose their crisis-prevention skills. On top of that, if AI causes more small-scale disturbances, people could learn to view them as normal, and as a result not anticipate or prevent major destabilizing accidents. There's also the possibility that some AI systems may intentionally or unintentionally game regulatory rules or predictable supervisory systems.

To preserve the safeguards that are currently in place, financial institutions can:

- Arm people with key information on evolving market and consumer behaviors
- Train employees with interactive and explainable AI systems
- Scan for emerging systemic risks
- Systematically analyze the shocks that occur
- Stress-test the effects of AI-based gaming in a crisis
- Run discretionary “sanity checks” on top of institutional rules

However, these and other strategies are subject to change because the financial system itself is changing. Upstart players are entering the financial services market. Incumbent institutions are trying out new business models. Amid these structural shifts, system resilience may soon become a moving target.

### Key implications for executives



To manage the risk of market panic in an AI world, older tools (such as circuit breakers) will be paired with new ones.



Explainable AI can help to preserve systemwide guardrails in the long run.



The dynamic risks of a machine-integrated financial system will require regulators to take on new roles and forge deeper, mutually beneficial partnerships with financial institutions.

## The supervisor of the future

In a fast-changing, increasingly complex financial sector, supervisors may find themselves taking on new responsibilities such as:



Simulations and other analyses of the potential effects of AI



Reporting of insights on important new activities within financial services



Dissemination of application-specific best practices in AI design



Issuing pre-emptive alerts in exchange for institutional data



Separating fundamental movement from algorithmic noise during market crashes



Providing a sandbox to test new AI solutions for the financial industry

# Bias and fairness

To price their products and services appropriately, financial institutions must evaluate their customers for risk. But historically, financial decisions have sometimes incorporated factors such as race and gender that have no relationship to risk. The result can keep customers from accruing wealth and the economy from gaining new productivity. It can also subject financial institutions to greater regulatory scrutiny and public backlash.

## The role of AI

When it comes to unfair biases, AI's potential is mixed. On the one hand, AI can create opportunities to:

- Remotely provide high-quality, low-cost offerings to the unbanked
- Serve underbanked and unbanked customers by drawing on social, telematic, and other alternative data
- Offer personalized products, such as small loans or micro-duration policies, on demand

On the other hand, AI can also:

- Derive unintentional biases from new data sources
- Make it more difficult to identify unintended biases
- Magnify the effect of bias by spreading it faster and more widely across systems

Bias can come up within the financial system in a number of ways.

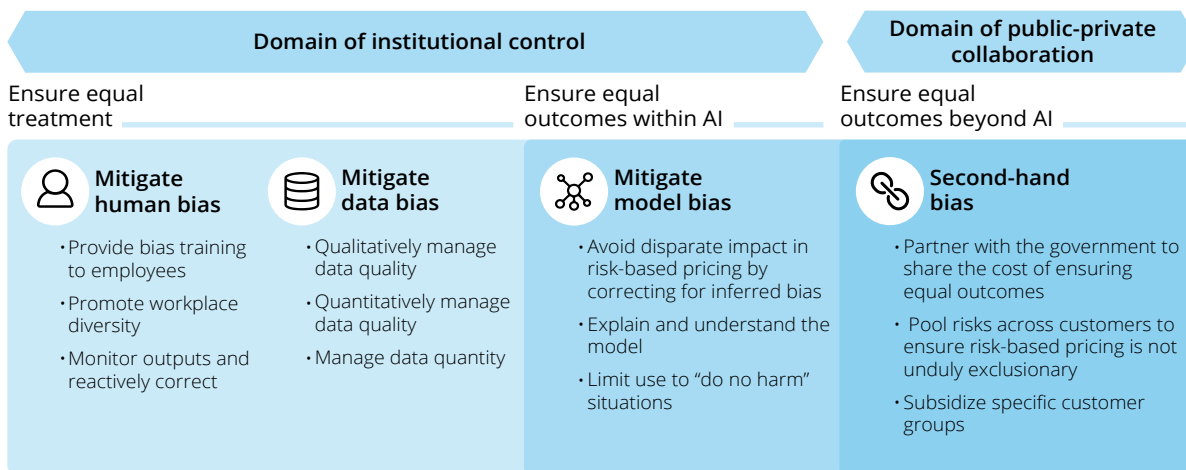
**Human bias.** Systematic errors in human thinking affect the decisions and judgments people make. These biases can be intentional or unintentional. AI can introduce human bias via system design and data collection practices. Supervised learning and model application can also bring human bias into an AI system.

**Data bias.** If inaccurate, nonrepresentative, or otherwise biased data is used to support the development or training of a decision-making system, it can produce distorted outcomes for specific populations.


**Model bias.** Limitations in computational power, system design, or programming logic—not to mention user error—can result in biased output from an AI system. Examples of model bias include the confusion of correlation for causation and unintentionally proxying for protected classes.


**Second-hand bias.** Unfair discrimination is unfortunately present in the world today across a variety of facets of life (e.g., education, employment, legal proceedings). This genuinely increases the financial risks that individuals subject to the discrimination can face. Unlike other biases, second-hand bias is not directly controllable by those who build or use a given model; mitigating it may require institutions to forgo some statistical accuracy. AI can introduce this bias through incorporating a greater breadth and depth of underwriting data (e.g., social media), including data that comes from third parties.


Figure 5: Minimizing bias in AI requires making choices about which type of “fairness” is acceptable and what an institution’s responsibility for “second-hand bias” will be




**Key implications for executives**

- 

AI creates new opportunities for financial services to support financial inclusion. At the same time, AI could increase the risk of unfair and biased financial decisions.
- 

Although bias in financial decisions isn’t new, AI raises fresh concerns by creating new ways for bias to enter into decisions, spread faster, and become harder to detect.
- 

There are existing tools and systems to help identify and mitigate bias from people, data, and algorithms. But AI can introduce complications—especially when it comes to second-hand bias (where the source of discrimination lies outside the direct control or influence of the institution).
- 

Mitigating bias (especially when it’s second-hand) may require choosing to forgo a purely risk-based approach. Navigating this challenge will require financial institutions and policymakers to arrive at a shared definition of fairness, which could involve cost-sharing between public and private sectors.

# Other uncertainties

When it comes to AI in the financial sector, explainability, systemic risk, and bias and fairness are the three priority concerns. However, two others bear mentioning. One is whether AI systems can meet the obligations of a fiduciary—that is, act only in the best interest of the customer. The other uncertainty has to do with algorithmic collusion, where AI systems learn to engage in anti-competitive behavior.

## The algorithmic fiduciary

AI systems have started to take on financial tasks that were once the sole province of humans. For example, they can:

- Select securities and manage them for above-market returns
- Recommend asset allocations that align with customers' long-term goals
- Provide holistic, day-to-day financial management across a broad set of products and services (such as insurance, retirement, and tax or estate planning)

All this raises important questions about fiduciary duty. These questions have taken on a new urgency in the United States, where AI algorithms are already in use among registered investment advisors. The good news is that it's possible for automated systems to meet the same fiduciary standard as traditional firms. This is at least partly due to the efficient way AI can search, optimize, and aggregate data related to each client's financial situation.

However, there's the explainability issue we discussed earlier. A fiduciary should have—and communicate—a reasonable basis for its decisions and recommendations. But as AI systems incorporate a broader set of information for a greater variety of products and services, it becomes increasingly difficult to understand the logic behind a specific decision.

Separately, there's uncertainty over what happens if an AI system falls short of its fiduciary responsibilities. Who should be responsible: developers, institutions, management, or someone else? Alternatively, should indemnity insurance take over? Whatever the approach, regulators and institutions likely will need to bolster AI systems with other methods that make affected parties whole, as quickly and seamlessly as possible.

## Algorithmic collusion

AI-enabled systems can complicate the monitoring and governance of collusive behavior in the market. Why? Consider that collusion today refers to explicit communication and cooperation between institutions to achieve specific outcomes, or change competitive dynamics, in their favor.

Under this definition, collusion among AI systems may be hard to prove. For instance, it may not be clear whether decisions are the result of independent and valid analysis or collusive behavior. Another dilemma springs from AI's ability to incorporate vast amounts of structured and unstructured data from external sources in real time. That can make it easier to understand signals from other market participants without any explicit communication.

As a result, plausible forms of AI collusion could be more difficult to detect, prove, and prosecute under existing regulations. This raises uncertainties over what should be considered illegal and what should not.

Even so, financial institutions aren't powerless to mitigate these risks. They can restrict AI systems and models to:

- Communications that have specific, explicitly justifiable business purposes
- Decisions informed by valid, legal business reasons
- Human oversight to validate the business case behind AI decisions

Besides all that, algorithmic collusion is hardly a given. For it to happen, financial products and services would need to be relatively homogenous. Institutions would need to have similar operating margins as well, and buyer power would need to be significantly limited.

### Key implications for executives



Addressing the concerns around fiduciary duty and algorithmic collusion involve bridging the gap between technical AI expertise and policymakers or legal expertise.



It remains to be seen whether a single framework for human and machine actors will suffice, or if separate frameworks designed specifically for machine agents need to be defined.



By heading off collusion and adopting the obligations of fiduciary duty, organizations can garner trust and differentiate themselves from firms with less stringent standards.

# Concluding thoughts

AI represents a fundamentally different paradigm in computing. Unlocking its benefits will require the financial sector to navigate the challenges AI brings to explainability, systemic risk, and fairness. In the end, this will require the financial sector to:

- **Responsibly deploy AI in the financial ecosystem of today.** AI requires the financial industry to develop—and become comfortable with using—completely new tools to safeguard the financial system.
- **Responsibly scale the AI-ubiquitous financial ecosystem of tomorrow.** Over time, AI will alter the way financial institutions operate and go to market. Institutions may want to get involved now if they hope to influence the policy changes ahead, many of which involve challenging questions that fundamentally alter the strategic avenues available to financial institutions.
- **Harness the potential of a financial ecosystem built on responsible AI.** AI systems can have a larger, quicker impact than traditional systems. As a result, they should be held to a higher standard.

Above all, it's important to recognize that AI doesn't exist in a vacuum. It's tightly integrated with technologies like cloud and IoT already. Quantum computing, 5G, and other emerging technologies will join them soon enough. It's against this shifting backdrop that the financial industry will need to consider the impact of AI.

And what will that impact look like? A shift in strategy, for one. The proliferation of AI (along with other emerging technologies) is prompting policy dialogues that will meaningfully shape how these technologies can be deployed. Institutions will have to proactively engage all stakeholders if they want an active role in shaping their future strategic options. The same approach extends to managing the operational, strategic, and human capital implications of policy changes.

Next is a reordering of priorities. With AI, governance is job one—for heeding risk warnings as well as for staying in line with customers' interests. A focus on governance not only offers financial institutions a chance at differentiation, it can support data-driven policymaking by creating a culture of collaboration with regulators and other players in the financial services ecosystem.

The upshot is that AI may not require a new set of ethics in financial services after all. What it does demand is an openness to new approaches that are better suited to the needs of complex AI-enabled systems. However innovative those approaches may be, they will continue to be weighed against the standards of fairness, stability, transparency, and accessibility that have set the financial sector apart through many earlier eras of technological upheaval.



# Contacts

## Global contacts

### **Bob Contri**

Financial Services Industry Leader,  
Deloitte Global  
[bcontri@deloitte.com](mailto:bcontri@deloitte.com)

### **Rob Galaski**

Global Leader, Banking & Capital Markets,  
Deloitte Consulting  
[rgalaski@deloitte.ca](mailto:rgalaski@deloitte.ca)

## Author

### **Ishani Majumdar**

Senior Consultant, Omnia AI, Deloitte Canada  
[ismajumdar@deloitte.ca](mailto:ismajumdar@deloitte.ca)

We would also like to thank Courtney Kidd Chubb and Denizhan Uykur from Deloitte Canada for their contributions to this report.

# End Notes

1. BaFin, Big Data meets artificial intelligence, 16 July 2018, [https://www.bafin.de/SharedDocs/Downloads/EN/dl\\_bdai\\_studie\\_en.html?sessionId=DF0654DBF78C8FA8AE5E4AA58148CB44.1\\_cid381?nn=9866146](https://www.bafin.de/SharedDocs/Downloads/EN/dl_bdai_studie_en.html?sessionId=DF0654DBF78C8FA8AE5E4AA58148CB44.1_cid381?nn=9866146), accessed 5 November 2019.
2. "New Scoring Models For The Underserved Gaining International Attention," PYMNTS, 3 July 2018, <https://www.pymnts.com/news/banking/2018/alternative-credit-scoring-underbanked-unbanked-nova/>, accessed 5 November 2019.
3. Sudipto Ghosh, "Temenos Acquires a SaaS-based, Patented, Explainable AI (XAI) Platform," AiThORITY, 18 July 2019, <https://www.aithority.com/natural-language/information-extraction/temenos-acquires-a-saas-based-patented-explainable-ai-xai-platform/>, accessed 5 November 2019.
4. "Personetics leverages AI to help chip away at student loan crisis," FinTech Futures, 30 October 2017, <https://www.fintechfutures.com/2017/10/personetics-leverages-ai-to-help-chip-away-at-student-loan-crisis/>, accessed 5 November 2019.
5. Financial Stability Board, Artificial intelligence and machine learning in financial services: Market developments and financial stability implications, 1 November 2017, <https://www.fsb.org/wp-content/uploads/P011117.pdf>, accessed 5 November 2019.



# Deloitte.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited ("DTTL"), its global network of member firms, and their related entities. DTTL (also referred to as "Deloitte Global") and each of its member firms are legally separate and independent entities. DTTL does not provide services to clients. Please see [www.deloitte.com/about](http://www.deloitte.com/about) to learn more.

Deloitte is a leading global provider of audit and assurance, consulting, financial advisory, risk advisory, tax and related services. Our network of member firms in more than 150 countries and territories serves four out of five Fortune Global 500® companies. Learn how Deloitte's approximately 286,000 people make an impact that matters at [www.deloitte.com](http://www.deloitte.com).

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms or their related entities (collectively, the "Deloitte network") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. No entity in the Deloitte network shall be responsible for any loss whatsoever sustained by any person who relies on this communication.

© 2019. For information, contact Deloitte Touche Tohmatsu Limited.