AI in Consumer Finance

A Landscape Analysis of Market Forces, Evaluation Frameworks, and Regulatory Gaps

APRIL 2026

Consumer Reports

Contents

Introduction

Artificial intelligence (AI) systems are making critical, largely unaccountable decisions about Americans' financial lives—determining creditworthiness, setting prices, detecting fraud, and providing financial advice—faster than those decisions can be fully explained or validated. When a chatbot gives overconfident investment advice, when an algorithm incorrectly flags a legitimate transaction as fraud, or when prices change in real-time based on opaque calculations, consumers face real consequences: denied credit, higher costs, financial losses, and erosion of trust.

Consumer Reports' 2025 nationally representative survey of over 4,000 Americans reveals the depth of this accountability gap: while 75% are at least somewhat concerned that AI in financial services could lead to bias or unfair treatment, only 8% believe current laws and regulations adequately protect them.[1] More fundamentally, when asked about their actual experiences with AI in financial services, a minority reported positive experiences—except in fraud detection, where the benefit to consumers is clearest and most direct.[2]

The artificial intelligence landscape is rapidly evolving, yet evaluation frameworks remain limited and incomplete. Regulatory oversight, while accelerating, lags behind the pace of industry in both understanding and effectiveness. This gap between deployment speed and accountability infrastructure creates systemic risks to consumer welfare and fair competition.

This report examines why these gaps exist and what must be done to fill them. We begin by providing an overview of the market conditions in Consumer Financial Products and Services (CFPS), the AI ecosystem, and their intersection. We then dive into existing frameworks to inform the methodology and opportunities that might drive the development of a Consumer Reports-led framework. Next, we examine the current U.S. AI CFPS regulatory landscape, along with certain novel foreign regulations, to inform the development of an actionable, consumer-centered standard to evaluate the use of AI in CFPS, guide consumer choices, and identify policy opportunities.

Throughout this report, we rely on a set of axes of evaluation (page 4) as an analytical framework for assessing both existing governance models and emerging regulatory approaches. These axes do not represent a finalized standard, but rather a structured lens through which to examine how well current frameworks account for consumer risk, accountability, and market integrity. By organizing the landscape in this way, we surface patterns of alignment and divergence, identify material gaps in consumer protection, and clarify where further development is needed. This foundation is intended to inform the development of a more actionable, consumer-centered evaluation framework for AI in consumer financial products and services.

---

[1] Consumer Reports nationally representative AI in Financial Services Survey of 4,073 U.S. adults (September-October, 2025), https://article.images.consumerreports.org/image/upload/v1767035543/prod/content/dam/surveys/Consumer_Reports_AI_in_Financial_Services_2025.pdf.

[2] Ibid.

Market Overview

This section explores the market dynamics that inform AI development and use. The sudden virality of large language models in late 2022 was an inflection point in the AI market's evolution. Since then, the industry has rapidly grown in scale and influence. Fueled by big-tech cash and some degree of speculation, AI is beginning to reshape the way we work, live, and do business. Consumer financial service providers are keen to partake in this seismic shift. But there is significant work to be done to ensure that this industry develops without leaving the consumer behind. One signal of this risk is that consumer demand is not the primary financial driver of AI development, which minimizes the consumer's voice in the market. Additionally, consumer financial products and services that leverage AI are deeply impacted and informed by market forces in the tech and cloud industry more broadly, which has a track record of not prioritizing what's at stake for consumers. As AI accelerates the convergence of big banks and big tech, it is all the more important that we understand, articulate, and enforce the responsibilities AI companies have to respect and protect consumers.

Framework Overview

In this section, we survey notable frameworks from non-profits, government agencies, AI companies, and academia that address our axes of evaluation. We focus on frameworks that provide clear methods for measuring the risks associated with the most popular machine learning models employed in consumer finance, as well as frameworks that mitigate those risks. Beyond simply surveying methods, we also highlight development frameworks that provide tools and codebases to support the practical application of these concepts where they exist. Since the field of ethical AI lags behind the commercial development of AI systems, we lean heavily on frameworks from academia that provide clear guidance on how to implement novel strategies.

This section is organized by axes of risk. For each, we provide a brief overview of the space and highlight a few existing frameworks to illustrate current approaches and identify gaps, setting the stage for a more comprehensive, consumer-centered standard.

Regulatory Overview

To provide an overview of the regulatory landscape, we survey the most relevant laws, regulations, guidance, and proposals that might affect consumer financial products and services that depend on AI. We focus on concepts with the highest potential impact for the CFPS marketplace and examine their alignment with our axes of evaluation, which can be found on the following page. Given the substantial impact that different rules might have on AI CFPS, our examination draws on active statutes, promulgated regulations, proposed bills and rules, executive orders, regulatory guidance, and revoked measures across artificial intelligence, consumer financial regulation, privacy regulation, and related areas. As the future of the U.S. regulatory apparatus remains uncertain at the time of this writing, we draw out gaps in existing

law and highlight opportunities for consumer advocates to best inform framework development. We additionally use our axes to examine clear categorical absences in the U.S. regulatory space and survey some international laws and regulations that provide novel regulatory concepts.

This section is organized by regulatory concepts, allowing readers to efficiently compare various approaches, identify their differences, and learn about their limitations. By cataloging these concepts in detail, this section seeks to offer both a snapshot of current regulatory thinking and a toolkit for stakeholders building frameworks to evaluate whether the AI CFPS market fully centers consumers.

Axes of Evaluation and Working Definitions	
Accountability & Transparency	How easily can one understand how the AI works and how effectively can the AI be fixed or changed when it is wrong?
Accuracy & Reliability	How likely is it that assertions stated as fact are actually accurate, and how well does the system perform under varying conditions, data inputs, and use contexts?
AI Safety and Human-Centeredness	How much does the AI prioritize human wellbeing and experience, how much damage can the AI do, and how strong are its safeguards?
Bias & Fairness	To what extent do the AI's decisions result in biased outcomes for protected classes?
Consumer Agency and Empowerment (Including Awareness of AI Use)	How much power do consumers have over the use and actions of the AI, and how easily can they opt out of some or all of it?
Data Security	How secure are the AI and its supporting services and infrastructure ?
Ecological Footprint	How much environmental damage is done by the AI's development and operation?
Ethics in Training Practices & Copyright	How well does the AI and AI development process honor copyright and/or appropriately compensate those who created materials used?
Privacy	How effectively does the AI protect the privacy of consumers and safeguard their data?
Quality of Redress Mechanisms (And Dispute Resolution)	How well does the company or its AI handle redress for consumers they harm?
Resilience	How elegantly and transparently does the AI fail when underlying infrastructure is compromised?
Sycophancy	How likely is the AI to tell the user what they want to hear regardless of whether it is factual or safe?

Market Overview

The AI market is turbulent and rapidly developing. Major general-purpose chat tools like OpenAI's ChatGPT and Anthropic's Claude are updated on a cadence of months rather than years—with flagship releases and incremental versions appearing several times per year—and it remains to be seen how the latest generations will continue to shape the market. Even so, there are valuable takeaways from analysis of the current market that we will explore in this section:

1. **The AI space is deeply intertwined with other tech markets.** One cannot easily look at a segment like AI in consumer finance without also considering the interdependencies of cloud providers, datacenter infrastructure and energy, or AI model developers.
2. **Investment and market enthusiasm currently outpace reliable demand signals.** The rapid evolution of AI capabilities, combined with a complex and layered supply chain, limits transparency into end-user demand and sustainable revenue models. These conditions create incentives for speculative investment and rapid deployment, even where long-term consumer adoption patterns remain unclear.
3. **Competitive pressure to lead in AI development may outpace the development of shared standards, benchmarks, and guardrails**. In a market in which definitions remain fluid and performance benchmarks are still evolving, institutions face incentives to prioritize speed of deployment and visible innovation. This dynamic can complicate efforts to embed consumer protection considerations early in product design, particularly in complex financial contexts.
4. While new AI-native companies emerge frequently, **consumer impact is largely mediated through integration within incumbent financial institutions** or through widely used general-purpose AI platforms.

Why Market Forces Work Against Consumer Welfare

The rapid expansion of AI in consumer financial products and services is not primarily driven by demonstrated consumer benefit. Instead, it is driven by structural incentives embedded within the AI supply chain.

Cloud hyperscalers—Amazon, Microsoft, and Google—stand to benefit from AI adoption regardless of whether AI deployments improve consumer outcomes. These firms generate revenue from compute usage, data storage, and networking infrastructure at scale.[3] If AI systems, which require significantly more compute usage than historical workloads, are deployed widely—even prematurely—the hyperscalers profit. Their incentives are tied to model training volume, enterprise contracts, and infrastructure demand, not to whether AI systems reduce consumer harm, improve fairness, introduce labor efficiencies, or enhance transparency.

[3] Yuvraj Malik, "AI Boom Fuels Cloud Computing Boom for Tech Giants Demand for Cloud Infrastructure," *Reuters*, May 1, 2024, https://www.reuters.com/technology/ai-fuels-cloud-computing-boom-tech-giants-2024-05-01/.

At the firm level, executive incentives reinforce this dynamic. Public companies increasingly signal "AI adoption" and "AI integration" as a proxy for innovation, competitiveness, and market leadership.[4] Internal performance metrics often reward AI integration, experimentation, or cost-reduction targets rather than measurable improvements in consumer-protection outcomes, product quality, or customer satisfaction. In this environment, speed of deployment and perceived technological sophistication are prioritized over validation, explainability, or consumer-centered design.

Speculative investment further accelerates deployment timelines. Venture capital flows and public market expectations reward rapid scaling and bold claims about AI's transformative potential.[5] The result is a market in which adoption frequently precedes validation. Systems are integrated into pricing engines, underwriting models, fraud-detection pipelines, and customer service interfaces before their downstream effects are fully understood and before monitoring, fault detection, and remediation systems can be built to support these new integrations.

Critically, consumer demand is not the primary driver of this adoption cycle. Consumers are not demanding generative AI chatbots for financial advice at scale, nor are they demanding opaque algorithmic pricing systems. Where AI is integrated, it is often bundled into existing services rather than chosen explicitly. This means neither consumer welfare nor utility to consumers are the primary constraints shaping AI deployment decisions.

This market structure produces recurring patterns. Systems optimized for engagement may exhibit sycophancy—telling users what they want to hear rather than what is accurate. Systems optimized for revenue or margin may produce opaque pricing structures. Systems optimized for automation may provide overconfident advice or insufficient human review. And systems optimized for efficiency may weaken redress mechanisms, shifting the burden of correction onto consumers.

Without structural counterweights, these outcomes are not anomalies; they are predictable features of the current incentive landscape.

[4] Alex Singla et al., "The state of AI in 2025: Agents, innovation, and transformation," *McKinsey & Company,* November 5, 2025, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.
[5] Darrell M. West, "Is there an AI bubble?," *Brookings Institution*, November 7, 2025, https://www.brookings.edu/articles/is-there-an-ai-bubble/.

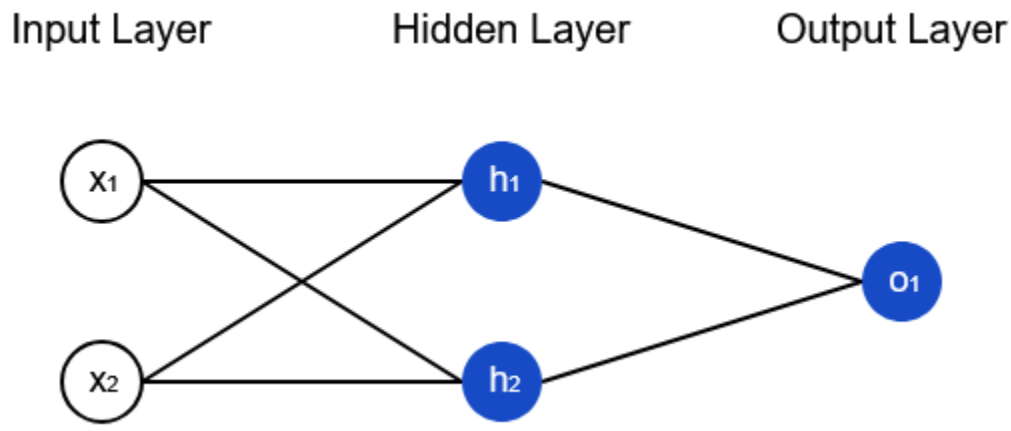
Some AI Basics: AI Development

This report presumes a basic familiarity with AI conceptually, but we first attempt to establish a foundation of understanding. There is not much consensus on a consistent taxonomy for different AI applications, but here are some examples of commonly accepted types of AI and what they mean:

- Generative AI - A predictive algorithm that uses heuristics to guess at what should come next in a sequence, whether that is text, an image bitstream, or an audio signal.
- Agentic AI - The practice of coordinating multiple generative AI agents that each specialize in a relatively simple task, in order to accomplish a more complicated task.
- Machine Learning Algorithm - An algorithm that analyzes its own output to improve its decision-making heuristics and is often used to make a binary decision. (Generative AI is also technically a form of machine learning.)
- Natural Language Processing Unit - An AI algorithm that specializes in interpreting text. Advances in this technology were critical in developing generative and agentic AI.
- Computer Vision - A field of study to develop algorithms that can interpret images and visual signals.

Artificial intelligence is not a new technology, and neural networks (the technology that powers everything listed above) have been around for decades. Neural networks make decisions by combining and recombining thousands or millions of signals depending on the size of the network. These signals are generated by applying a weighted mathematical function to all inputs, with each neuron introducing its own biases and differentiated weights.

Training a neural network is the process by which those weights and biases are attenuated to the point where the model is correctly predicting results based on inputs. “Correct” predictions are deemed so by the data used to train the model, so poorly maintained or unreliable training data can easily lead to an overconfidently poor neural network.



Here are some resources to learn more about neural networks:

- Intro to Neural Networks - Victor Zhou
- Neural Networks Overview 1 Introduction - UPenn

AI development roughly follows these steps:

1. Create or buy the initial AI Model.
2. Train it on massive amounts of domain-specific data.
3. Package it into some form that makes it accessible to the end user (and ensure this package doesn't prevent continued training).

This development process creates several niches and sub-industries for firms to enter. Among them: developing the core models; building and maintaining the infrastructure to support model development; collecting, curating, and cleaning data for models to train on; training models; and packaging trained models or integrating them into an existing product.

It may be helpful to think of this as the AI supply chain in order to better understand where each firm fits. Like a traditional supply chain, it often makes sense for firms to vertically integrate and perform multiple functions of the supply chain while outsourcing others.

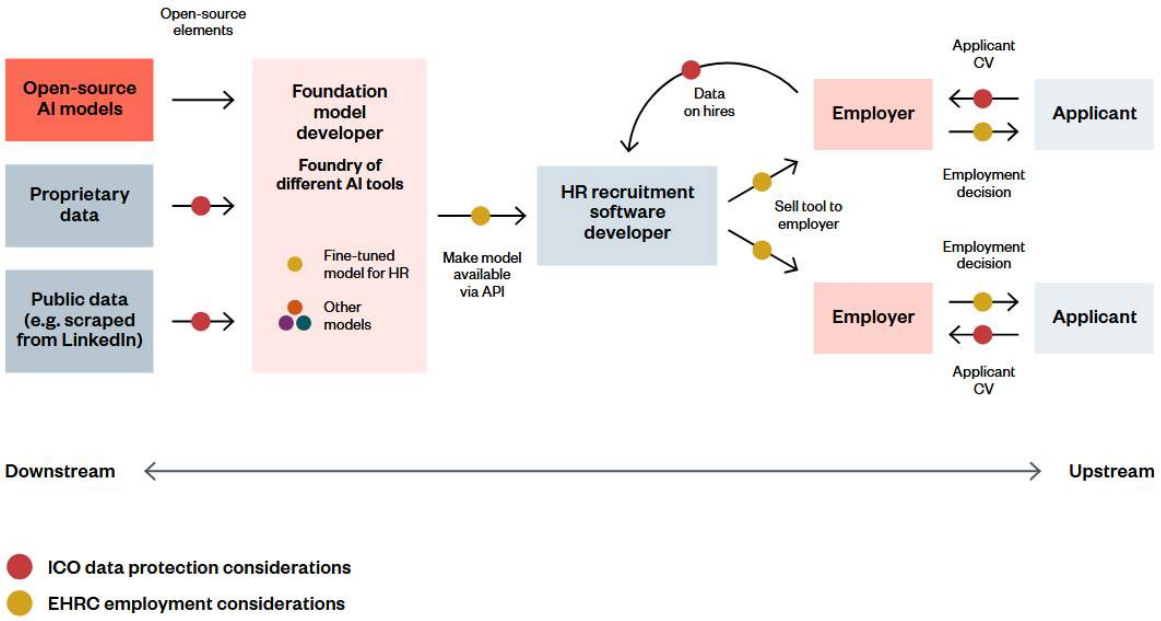
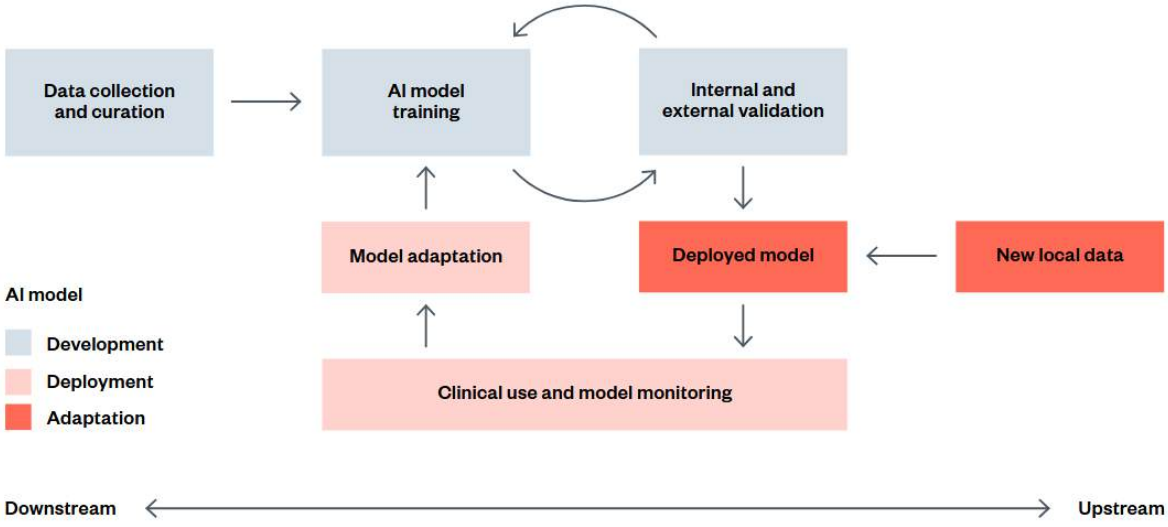


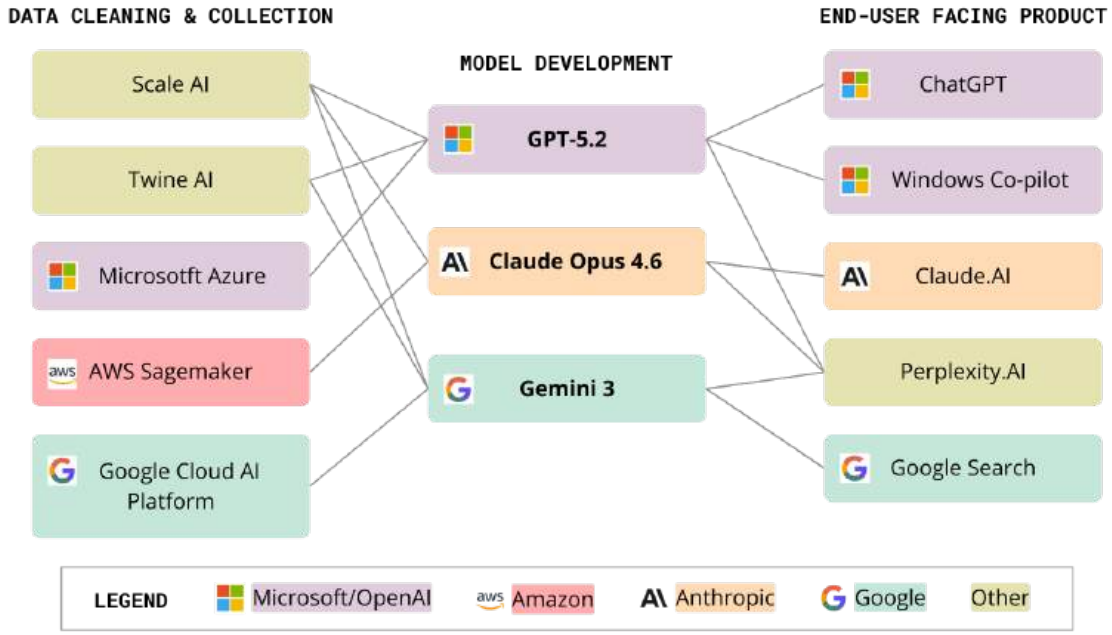
Figure 1: An abstract example of an AI system's lifecycle (based on a system used in the COVID-19 pandemic)[2]



Source: [Allocating accountability in AI supply chains: a UK-centred regulatory perspective]

The above example illustrates both the lifecycle of an AI system and the various stakeholders that might be involved in the development and distribution of the same.

Let us visualize this through the lens of a handful of the most popular AI Models.



GPT-5 - OpenAI

The latest model that powers ChatGPT.com and the ChatGPT mobile app, GPT-5.2 was released on Dec. 11, 2025. The priced tiers include access to later models like GPT 5.4, which Open AI claims has lower incidences of factual errors and more optimized professional workflows.

The publicly available model, GPT-5.2, is purportedly less prone to hallucinations than previous models. However, as detailed by The Guardian earlier this year, GPT 5.2 still cites misinformation ingested by sources like Grokpedia (an AI-generated encyclopedia) or bad information like misquotes even after they are taken down.[6] Though GPT-5 uses more intelligent routing to leverage sub-models and other specialized networks, this architecture opens the AI supply chain to an even more convoluted set of potential actors. This suggests that AI is trending in a direction where each AI tool will be an amalgamation of other AI models and tools.

[6] Aisha Down, "Latest ChatGPT Model Uses Elon Musk's Grokipedia as Source, Tests Reveal," *The Guardian*, January 24, 2026, https://www.theguardian.com/technology/2026/jan/24/latest-chatgpt-model-uses-elon-musks-grokipedia-as-source-tests-reveal.

OpenAI touts this model as a big leap from its previous model GPT-4o, but some experts are skeptical of those claims.[7] In addition to supporting ChatGPT, GPT-5.2 powers Windows Copilot, and is also used by other AI amalgamators, including Perplexity.AI.

Claude Sonnet 4.6 - Anthropic

Released on Feb. 17, 2026, Claude Sonnet 4.6 touts improvements over its predecessor when it comes to coding tasks, instruction-following, reasoning, knowledge work, and design.[8] This model powers Claude.ai and is positioned as a competitor to ChatGPT that puts greater emphasis on AI safety in its development and use. Safety evaluations run by Anthropic concluded that this model demonstrated "very strong safety behaviors."

Anthropic was recently embroiled in a disagreement with the Pentagon regarding its usage policy.[9] The "Claude Gov" model has been in use by multiple U.S. national security agencies since the summer of 2025. However, in February 2026, Defense Secretary Pete Hegseth requested unrestricted access to the Claude model for potential use in domestic surveillance or lethal autonomous weapons. Anthropic denied the request, claiming that granting unrestricted access would violate its usage policy. This made national headlines, giving Anthropic a popularity boost as a principled organization that prioritized safety. However, the publicized nature of the conversation also furthered public skepticism that AI products will deliver public benefit.[10]

Gemini 3 - Google

Released in November 2025, the latest Gemini 3 powers Google's AI search product, and also comes packaged on many android releases. Meanwhile, paid subscribers of Gemini Pro will have access to a preview version of Gemini 3.1. Google claims that Gemini 3 provides state-of-the-art reasoning, allowing you to "learn" anything, "build" (code) anything, and "plan" (use agents) anything.[11] This release threatened ChatGPTs hold on the market when Gemini 3 was found to outperform OpenAI on leaderboards like Arena.[12] Though there may seem to be competition between the AI models, most of the reported benchmark numbers are reported by

---

[7] Jeremy Kahn, "GPT-5: Everything That's New, Different—From Hallucinations to Personalities to 'Vibecoding' Agents," *Fortune*, August 7, 2025, https://fortune.com/2025/08/07/gpt-5-everything-new-different-hallucinations-personalities-vibecoding-agents-openai/.
[8] Anthropic, "Claude Sonnet 4.6," *Anthropic*, February 17, 2026, https://www.anthropic.com/news/claude-sonnet-4-6.
[9] Kevin Collier, "Anthropic AI Used in U.S. Defense Operation Targeting Venezuela's Maduro," *NBC News*, February 2026, https://www.nbcnews.com/tech/security/anthropic-ai-defense-war-venezuela-maduro-rcna259603.
[10] Dan Milmo, "Anthropic's Claude AI to Be Used by Pentagon," *The Guardian*, March 2, 2026, https://www.theguardian.com/technology/2026/mar/02/claude-anthropic-ai-pentagon.
[11] Sundar Pichai et al., "Gemini 3," *Google Blog*, December 2025, https://blog.google/products-and-platforms/products/gemini/gemini-3/.
[12] Sundar Pichai et al., "Gemini 3," *Google Blog*, December 2025, https://blog.google/products-and-platforms/products/gemini/gemini-3/.

the AI companies themselves. Even purportedly independent platforms like Arena, rely on crowd-sourced evaluations of AI "output," which are based more on perception than rigorous evaluation methodologies.

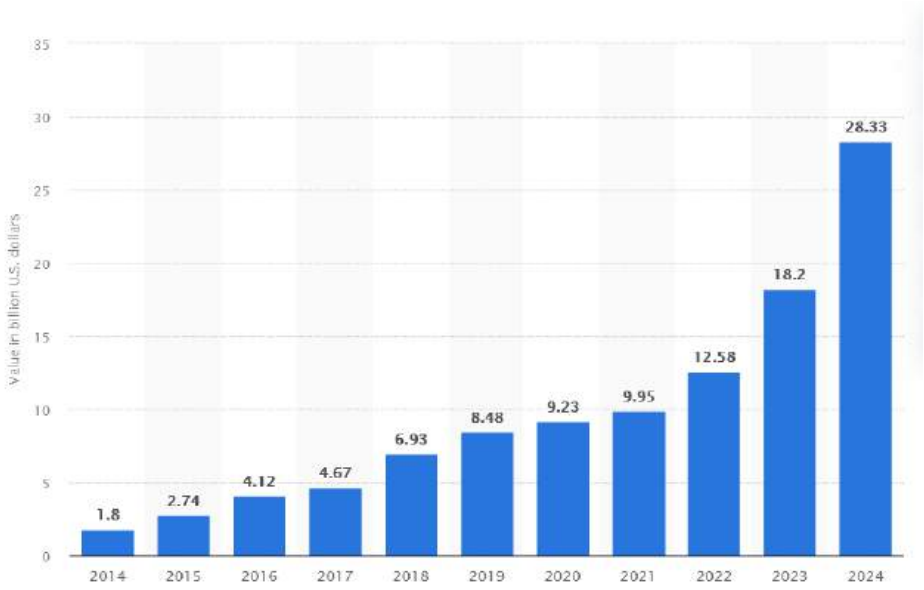
In each of the above cases, the models are trained on data collected, structured, and sometimes manually compiled by a broad range of companies. The visual above illustrates several examples of vertical integration, particularly with the major cloud players: Microsoft, Amazon, and Google. It is clear that these datacenter operators play a vital role in several layers of the supply chain.

Cloud Infrastructure as Primary Beneficiaries of AI Adoption

Training a large language model with billions of neural circuits and millions of data inputs is an expensive and time-consuming process. It requires specialized hardware at significant scale with an additionally significant energy-consumption cost. The only companies well positioned to do this are the large datacenter hyperscalers: Amazon, Microsoft, and Google (collectively controlling two thirds of the global Cloud Compute market). Conveniently for the rest of the AI industry, all three companies have a robust platform for satisfying this sudden demand for datacenter capacity. In effect, if AI product and model development is a gold rush, the datacenter operators are selling the pickaxes as well as the mining rights.

This new impetus in the market has re-energized the Cloud Computing market race and has resulted in a significant influx in datacenter investment. Private datacenter construction has more than doubled since 2022,[13] and key figures across the industry have begun lobbying for tax incentives[14] and better energy infrastructure[15] to support even more datacenter development.[16] With good reason, too, as AI-model training is estimated to have contributed $13 billion to Microsoft's revenue last year.[17]

[13] Boston Consulting Group, "Breaking Barriers to Data Center Growth," *BCG*, 2025, https://www.bcg.com/publications/2025/breaking-barriers-data-center-growth.
[14] Emil Sayegh, "Stargate AI Project: The $500 Billion Gamble to Dominate the Future," *Forbes*, January 22, 2025, https://www.forbes.com/sites/emilsayegh/2025/01/22/stargate-ai-project-the-500-billion-gamble-to-dominate-the-future/.
[15] McKinsey & Company, "AI Power: Expanding Data Center Capacity to Meet Growing Demand," *McKinsey & Company*, 2025, https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-power-expanding-data-center-capacity-to-meet-growing-demand.
[16] OpenAI, "Announcing the Stargate Project," *OpenAI*, January 2025, https://openai.com/index/announcing-the-stargate-project/.
[17] Todd Bishop, "Microsoft Earnings Show AI Driving Growth," *GeekWire*, 2025, https://www.geekwire.com/2025/microsoft-earnings-2.



Source: [U.S. private data center construction 2024| Statista]

These datacenter projects are touted by the hyperscalers as “reinvigoration opportunities” and a way to gain favor with local governments, since many of them break ground in rural areas. The “hype” of AI has led many local governments to eagerly grant tax subsidies and other handouts despite strong evidence that such incentives are not correlated to where these projects are ultimately started. Furthermore, promises of local job creation are overblown given the temporary nature of construction projects and the fact that most datacenters are maintained largely remotely and by white collar workers already contracted with the datacenter owners.[18]

This environment creates a strong incentive for a small group of powerful players to make sure that AI as a product is successful. If demand evaporates, these datacenters could sit idle, imposing significant costs for the hyperscaler companies. As articulated in the next section, this cultivates an environment in which AI demand is only obliquely driven by consumer interest, and where a number of other interesting forces are at play.

Demand for AI is not Well Mapped to Utility

In late 2022, OpenAI’s ChatGPT 3.5 went viral. Its emergence as the market leader was somewhat arbitrary: It was not substantially different from the previous version, nor was there a

[18] Financial Times, “OpenAI investors question $852bn valuation as strategy shifts,” *Financial Times*, April 2026.

drastic improvement in the technology. It simply caught on. Nobody expected the rapid and sudden levels of growth, least of all OpenAI itself.[19]
But even with billions of dollars in revenue, OpenAI is not profitable, posting billion-dollar losses every year and leaving it highly dependent on seed capital. While there are 7 million paying users (out of 700 million total), half of OpenAI's revenue comes from B2B contracts and therefore less indicative of consumer demand.

Despite those fundamentals, OpenAI was able to raise $110 billion in private funding earlier this year, shattering its previous record-breaking $40 billion raise from last year. To put these figures into perspective, $40 billion was nearly three times larger than the previous record for a private tech company fundraise.[20] It was also, after adjusting for inflation, nearly twice the size of the 2012 IPO of Facebook (now Meta), which, at the time, was the largest tech IPO in history.[21] However, Facebook was already profitable and boasted nearly twice as many monthly average users at the time of its IPO as OpenAI has today.[22] Despite these advantages, Facebook did not even approach the fundraising ability of OpenAI. (Following the IPO, Facebook faced allegations that it had overstated aspects of its business fundamentals, and several lawsuits followed.)

The tantalizing promise of AGI

One refrain that may be familiar is that "Artificial General Intelligence" (AGI) will cure cancer, end hunger, and generally solve all of our problems. While theoretically this premise might have merit, experts are largely skeptical of anyone who claims we are close to achieving AGI.[23] One flaw in the premise is that there's no agreed upon threshold for AGI, so it is already an arbitrary moving target.[24] Additionally, most AI companies are pursuing a strategy of scaling existing artificial intelligence architectures, an approach that the majority of AI researchers firmly believe is insufficient to achieve AGI if it is defined as an AI with the ability to cure cancer.[25]

---

[19] Will Douglas Heaven, "The Inside Story of How ChatGPT Was Built from the People Who Made It," *MIT Technology Review*, March 3, 2023, https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/.
[20] "OpenAI Statistics," *TapTwice Digital*, 2025, https://taptwicedigital.com/stats/openai.
[21] Daniel Howley, "Facebook's $16 Billion IPO, 10 Years Later," *Yahoo Finance*, May 18, 2022, https://finance.yahoo.com/news/facebooks-16-billion-ipo-1-211326390.html.
[22] Facebook, "Facebook Reports Fourth Quarter and Full Year 2012 Results," *Meta Investor Relations*, January 30, 2013, https://investor.atmeta.com/investor-news/press-release-details/2013/Facebook-Reports-Fourth-Quarter-and-Full-Year-2012-Results/default.aspx.
[23] Alex Engler, "Most Researchers Do Not Believe AGI Is Imminent. Why Do Policymakers Act Otherwise?," *Tech Policy Press*, 2023, http://techpolicy.press/most-researchers-do-not-believe-agi-is-imminent-why-do-policymakers-act-otherwise/.
[24] "AGI Is Not a Milestone," *Normal Tech*, 2024, https://www.normaltech.ai/p/agi-is-not-a-milestone.
[25] Alex Engler, "Most Researchers Do Not Believe AGI Is Imminent. Why Do Policymakers Act Otherwise?," *Tech Policy Press*, 2023, https://www.techpolicy.press/most-researchers-do-not-believe-agi-is-imminent-why-do-policymakers-act-otherwise/.

Despite the skepticism of experts, many figures in the industry are firmly asserting that AGI is around the corner, and this has driven significant speculative interest, fueling OpenAI's elevated valuation.

Internal Pressures and AI Adoption

A significant driver of AI deployment appears to be internal organizational strategy rather than consumer demand. Many technology and financial firms are actively integrating AI tools into existing workflows in response to competitive pressures and expectations of productivity gains.

Managers and executives, in particular, frequently frame AI as a broadly applicable tool that should dramatically increase the productivity of their teams and companies—even when applying it to tasks that may not be a good fit for AI tools such as LLMs.

In some organizations, this framing translates into explicit internal initiatives encouraging greater AI utilization across teams. Interviews with employees at large technology firms suggest that there are performance metrics and initiatives to encourage more AI use among their workforce. In some cases, employees said that performance bonuses were dependent upon teams meeting those "AI use" targets, which the employees described as arbitrary. Perceptions of the value of these initiatives were mixed at best, with some employees reporting that using AI instead of traditional tools led to lost efficiencies, and others describing uneven implementations and results.

At the executive level, AI is often associated with potential cost efficiencies in functions such as customer support, documentation, and other roles traditionally viewed as cost centers or ancillary to core business operations. While there appear to be opportunities for AI to improve operations in those spaces, its effectiveness in replacing human judgment or customer-facing roles will vary by complexity and risk tolerance and requires meaningful evaluation to determine whether AI tools are necessary and effective at replacing people in these contexts. In other words, when AI integration is treated as a strategic objective in itself, adoption may occur before benchmarks for reliability, safety, or consumer benefit are fully established.

At this stage of generative AI development, when internal performance metrics or product roadmaps prioritize AI integration as a goal in itself, adoption may become self-reinforcing. Features labeled as AI-enabled can persist even if overall product usage remains stable, including in cases where consumers would have preferred non-AI alternatives or clearer opt-out pathways. In markets with limited transparency or high switching costs, continued usage does not necessarily reflect affirmative consumer demand for AI-specific features.

Market Signaling and AI Adoption Dynamics

Another notable dynamic in the AI market is the role of signaling and competitive pressure in shaping adoption decisions. Organizations may integrate AI tools not just in response to

consumer demand, but also to signal innovation, maintain competitive positioning, or align with investor expectations.

Public discourse around AI adoption has become increasingly polarized in some contexts, including workplaces and educational settings. This polarization is not necessarily well mapped to existing divides, such as those across the political spectrum. Articles highlighting generational divides,[26] analyzing levels of AI resentment in the workplace[27] and reporting on hotly contested opinions on its role in schools,[28] all point to this trend. In these conditions, the decision to use or not use AI might be divorced from whatever value it actually provides to consumers.

This divide extends into the tech industry as well, with "accelerationists" advocating for more and faster AI development and adoption, and their derisively labeled "decelerationist" counterparts advocating for caution and a measured approach. As in any polarized environment, any nuance in the conversation is easily lost. It's also quite clear on which side of the debate sits most of the money. Venture capital funding isn't going anywhere near companies that are not thinking about the AI future.

These incentives create an environment in which a self-fulfilling prophecy can emerge. Thousands of tech workers are encouraging AI adoption in their product areas and are being rewarded accordingly. Many of these product environments lack reasonable alternatives or have other lock-in mechanisms. When consumers continue to use a product after AI is introduced, it can be framed as demand for AI, even if consumers would actually prefer to use the product without its AI elements.

Making AI less intelligent to find product market fit

Several trends have emerged as a result of this misalignment between demand signals and the utilitarian value of AI products.

For example, there are fundamental incompatibilities between developing AI that is accurate and circumspect and AI that is maximally engaging. Tech companies are deeply familiar with the strategies that maximize consumer engagement with their products, even where such strategies can have negative consequences for user health or finances. Accordingly, the issue of

[26] Aki Ito, "Gen Z Is Leading the Backlash Against AI at Work," *Business Insider*, August 2025, https://www.businessinsider.com/gen-z-people-against-ai-use-2025-8.
[27] Diane Hamilton, "The Rise of AI Resentment at Work: Why Employees Are Pushing Back," *Forbes*, February 3, 2025, https://www.forbes.com/sites/dianehamilton/2025/02/03/the-rise-of-ai-resentment-at-work-why-employees-are-pushing-back/.
[28] Megan Cerullo, "Schools Are Grappling with AI Use in the Classroom," *CBS News*, 2024, https://www.cbsnews.com/news/ai-in-schools-debate/.

sycophancy, or how likely an AI product will bias towards telling someone what they want to hear over telling them what is accurate, is an emerging concern.[29]

In addition, tech companies often don't rely on traditional sales and distribution channels, creating the need for other forms of monetization. One trend being discussed among AI companies that sell their models to other businesses is providing features that allow their models to be easily overridden to meet the needs of the customer. For example, consider a scenario in which Google is selling a custom version of Gemini 2.5 to Adidas to feature as a virtual assistant for customers on its website. Adidas certainly does not want the assistant to answer "Puma" when asked the question, "Who makes the best athletic shoe?" So Google develops a set of features that enable Adidas to ensure that such a question is always answered affirmatively with "Adidas." This raises obvious concerns around transparency and accuracy and creates a perverse incentive not to address biases that may emerge in the model during training.

More perniciously, features like these could be used to control the flow of information in any given context or environment. Such features raise questions that are analogous to those raised in debates around net neutrality.

Next we will explore how these forces have manifested in the financial industry specifically.

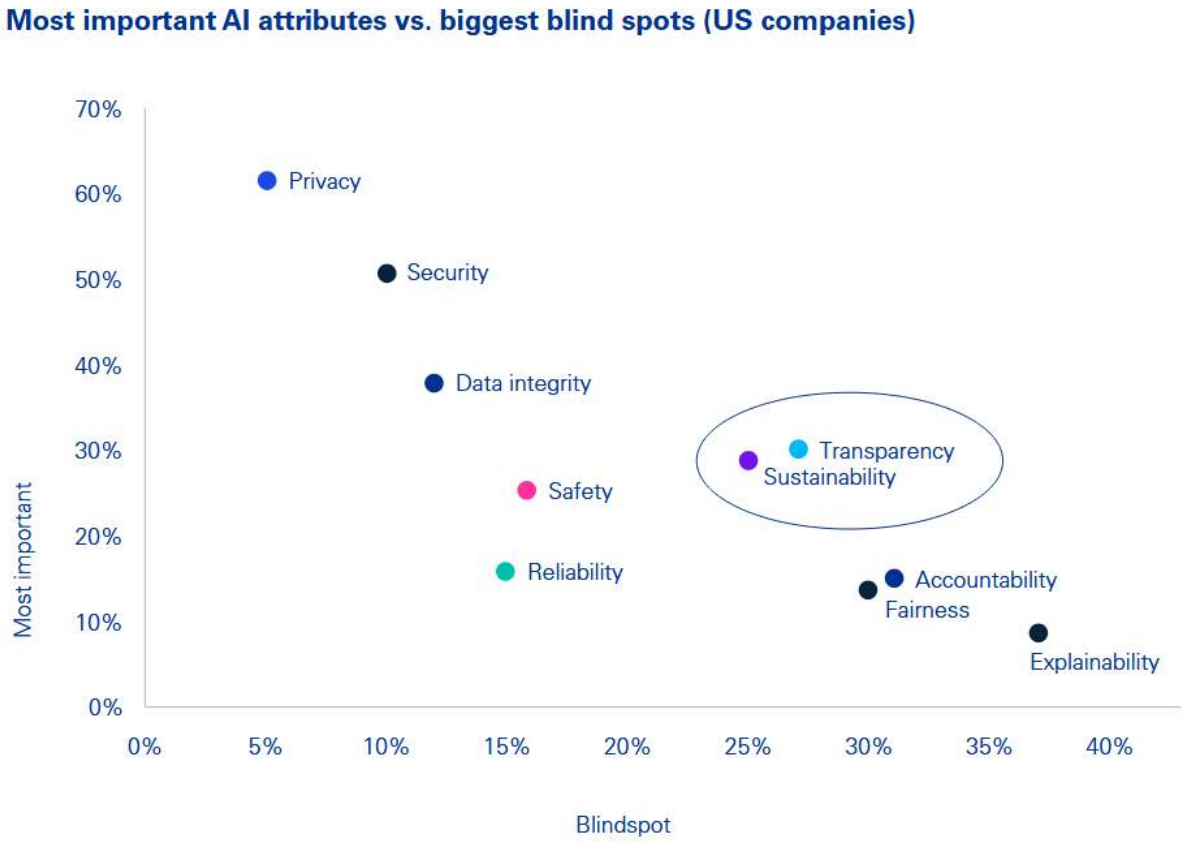
AI adoption within the financial industry

While many major banks have embraced AI to some degree, JPMorganChase is widely considered to be a leader in the space because of its multi-pronged AI initiative.[30,31] We will consistently refer to the major bank's efforts in AI throughout this section.

Overall, banks are keen to adopt AI into various segments of their business, including, among others, credit underwriting, fraud detection, and customer service. However, banks lag behind fintechs and other tech companies in AI adoption, despite those companies having also only begun their AI journey.[32] AI skeptics might argue that even this relatively slow pace of AI adoption is still too fast given the privacy risks and negative consequences that can result when banks get it wrong. It is telling that, according to a recent KMPG report, the AI attributes that companies consider the least important (accountability, fairness, explainability) are also those that are

[29] "Tech Brief: AI Sycophancy (OpenAI)," *Georgetown Law Institute for Technology Law & Policy*, 2025, https://www.law.georgetown.edu/tech-institute/research-insights/insights/tech-brief-ai-sycophancy-openai-2/.
[30] Tearsheet, "JPMorganChase's Gen AI Implementation: 450 Use Cases and Lessons Learned," *Tearsheet*, 2025, https://tearsheet.co/artificial-intelligence/jpmorgan-chases-gen-ai-implementation-450-use-cases-and-lessons-learned/.
[31] "JPMorganChase Accelerates AI Adoption," *CTO Magazine*, 2025, https://ctomagazine.com/jp-morgan-chase-accelerates-ai-adoption/.
[32] OpenText, "State of AI in Banking," *OpenText*, 2024, https://www.opentext.com/en/media/report/state-of-ai-in-banking-digital-banking-report-en.pdf.

considered by those same companies to be generating the biggest blind spots (receiving the least attention) within financial AI processes.



Source: Transforming into a new era with the AI-empowered finance function

Credit Decisions

In addition to banks adopting AI into their credit underwriting and decision-making processes, AI has been embraced wholeheartedly by several prominent Fintech companies in the loanmaking space. Upstart and OppFi both aggressively use AI in all of their business processes and particularly brag about leveraging alternative data and AI in their lending and underwriting. JPMorgan, meanwhile, has announced partnerships with Slope to provide short-term financing to small businesses. And companies like zest.ai offer their underwriting tools to a number of other financial institutions, furthering the adoption of AI in credit decisions.

Customer support and other business lines

While big banks are less likely to have already integrated AI into their credit and underwriting processes, they have been less hesitant to integrate AI into other parts of their business:

- JPMorgan has an AI assistant for its Customer Service Teams
- Bank of America's internal tool Erica supports its employees over a number of business lines
- Wells Fargo made headlines announcing a partnership with Google Cloud Platform and bringing agentic AI into its bank[33]

Security and Fraud Prevention

JPMorganChase has inserted AI voice recognition software from Nuance, which is owned by Microsoft, into its customer service-security process. A number of other banks and financial institutions are also now using this software. The product is positioned as saving consumers time by automatically verifying their voice when they call customer support, bypassing the need for other identity verification tools. However, its use raises a number of consumer protection concerns. For example, Chase collects voiceprints without consent or an obvious way to opt out of this AI tool. Furthermore, studies have shown that AI voice cloning can successfully beat these verification tools.[34]

Another worrying trend worth highlighting is that when an AI system decides consumer behavior is suspicious or fraudulent, there may be very little transparency or recourse when an adverse action is taken. This has always been a risk, but AI exacerbates this problem by making false positives more opaque and difficult to identify.

General Purpose AI as a Financial Advisor or Financial intermediary

It would be incomplete to assess the financial AI landscape without examining how consumers are using general-purpose AI tools for personal financial advice.

General-purpose AI chat tools such as ChatGPT, Claude, and Perplexity are among the most accessed websites globally.[35] Millions of users interact with them daily, and a growing number are using them to seek information related to personal finance.[36] Earlier iterations of generative AI were largely used by consumers to provide budgeting tips and/or explanations of financial

[33] Wells Fargo, "Wells Fargo Announces Expansion of Strategic Relationship with Google Cloud," *Wells Fargo Newsroom*, August 5, 2025, https://newsroom.wf.com/news-releases/news-details/2025/Wells-Fargo-announces-expansion-of-strategic-relationship-with-Google-Cloud/default.aspx.
[34] Seth Layton et al., "Every Breath You Don't Take: Deepfake Speech Detection Using Breath," *arXiv*, April 23, 2024, https://doi.org/10.48550/arXiv.2404.15143.
[35] https://www.similarweb.com/top-websites/.
[36] Reuters, "OpenAI says ChatGPT's weekly users have grown to 200 million," *Reuters,* August, 2024, https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/.

concepts,[37] which alone raises questions about accuracy, overconfidence, and privacy. But now, more complex and potentially consequential financial guidance is being delivered through AI systems not designed as financial advisors and that sit outside traditional supervisory frameworks.

While these systems initially provided limited utility when responding to financial questions[38] and made basic arithmetic errors,[39,40] they have improved over time. Still, their outputs remain probabilistic and may lack important contextual awareness of a user's full financial situation. They do not independently verify user-provided facts, assess suitability in a regulatory sense, or assume fiduciary obligations, which results in risks related to overconfidence, incomplete analysis, and miscalibrated trust.

Using general-purpose AI for financial advice also raises privacy and data-governance considerations.[41] Consumers routinely share sensitive financial information in conversational settings that may not be subject to the same regulatory protections as interactions with licensed financial institutions. Differences in default data-handling policies between enterprise and consumer tiers further complicate the privacy landscape.

However, since the introduction of general-purpose AI tools a few years ago, both consumer use and the marketplace have rapidly evolved beyond just asking and answering questions. AI-enabled tools are beginning to leverage open-banking infrastructure such as Plaid to access consumer financial accounts, enabling account- and portfolio-level analysis across multiple institutions.[42]

Recent product demonstrations illustrate how AI systems can ingest real brokerage data and generate customized dashboards, sector exposure analyses, or interest-rate sensitivity summaries in seconds. This represents a meaningful structural shift: AI is no longer merely

---

[37] Consumer Financial Protection Bureau, Chatbots in Consumer Finance: Consumer Use and Risks. 2023.
[38] Amanda Smith, "I Tried ChatGPT to Help Knock Out an $18K Debt—It Went Comically Wrong," *MSN*, 2025, https://www.msn.com/en-us/money/personalfinance/i-tried-chatgpt-to-help-knock-out-an-18k-debt-it-went-comically-wrong/ar-BB1q4HZI.
[39] Kyle Wiggers, "Why Is ChatGPT So Bad at Math?," *TechCrunch*, October 2, 2024, https://techcrunch.com/2024/10/02/why-is-chatgpt-so-bad-at-math/.
[40] Briana Brownell, "Why ChatGPT Struggles with Math–and why that matters," *Descript*, July 15, 2024, https://www.descript.com/blog/article/why-chatgpt-struggles-with-math.
[41] Kate O'Flaherty, "ChatGPT-4o Is Wildly Capable—But It Could Be a Privacy Nightmare," *Forbes*, May 17, 2024, https://www.forbes.com/sites/kateoflahertyuk/2024/05/17/chatgpt-4o-is-wildly-capable-but-it-could-be-a-privacy-nightmare/.
[42] Perplexity Upgrades Finance Capabilities, WealthManagement.com, 2025, https://www.wealthmanagement.com/artificial-intelligence/perplexity-upgrades-finance-capabilities.

answering hypothetical financial questions — it is interpreting live financial data and mediating consumer understanding of risk, concentration, and strategy.

Importantly, some providers have deliberately limited functionality — for example, restricting AI agents from executing trades directly. This design choice reflects a regulatory boundary between analysis and action. However, the interpretive layer itself can materially influence consumer decision-making, even in the absence of direct execution authority.

This evolution raises governance questions distinct from earlier chatbot concerns. When AI systems are connected to real financial accounts, several questions follow:

- What fiduciary or suitability standards apply?
- How should conflicts of interest be disclosed?
- How are errors detected and corrected?
- What privacy safeguards apply when brokerage-level data is aggregated and analyzed?
- Who is accountable for misinterpretation or miscalibration of portfolio risk?

The introduction of connected AI agents suggests that AI is increasingly becoming a layer between consumers and their financial institutions—not replacing institutions outright, but reshaping how consumers interpret and act on financial information.

Embedded AI in Financial Infrastructure

Beyond underwriting and fraud detection, AI is increasingly being integrated as a consumer-facing analytical layer within existing financial infrastructure.

Fintech infrastructure such as Plaid was originally built to allow applications to read consumer financial data across institutions. AI tools are now leveraging those same data rails to provide cross-account analysis, risk summaries, spending diagnostics, and portfolio interpretation.

In some cases, these AI systems integrate institutional-grade financial data sources—such as SEC filings or market data platforms—and combine them with real user holdings. The result is a hybrid product: part data aggregator, part AI interpreter, part financial research assistant.

This development changes the governance calculus. Unlike informal advice, connected AI analysis operates on real holdings and may influence allocation decisions, tax strategies, or risk exposure. Yet these systems often position themselves as “analysis tools” rather than registered advisors, thereby navigating a regulatory distinction between recommendation and execution.

The demand for AI-enabled financial interpretation appears to have preceded product development, with many users already querying general-purpose AI systems about finance. The shift now is that these systems can be grounded in personalized data rather than hypotheticals.

This raises questions about:

- Suitability and fiduciary standards

- Transparency of model limitations
- Redress mechanisms for incorrect analysis
- Concentration of data access across platforms
- Consumer understanding of the difference between advisory and analytical tools

The Misalignment Between AI Development Priorities and Consumer Needs

Market forces driving AI adoption in financial services are misaligned with consumer priorities.

Consumer Reports' 2025 nationally representative survey asked Americans what matters most for improving financial services quality.[43] The top responses were data security (selected by 49%), lower costs (44%), and better customer service (42%). By contrast, when asked what benefits they see from AI adoption, far fewer consumers identified these same priorities: only 24% saw AI strengthening data security, 34% saw it lowering costs, and just 18% saw it improving customer service.

What did consumers identify as AI's primary benefit? The most commonly selected was faster decisions (41%). Yet only 18% of consumers listed speed as a priority for improving financial services quality.

[43] Consumer Reports nationally representative AI in Financial Services Survey of 4,073 U.S. adults (September-October, 2025), https://article.images.consumerreports.org/image/upload/v1767035543/prod/content/dam/surveys/Consumer_Reports_AI_in_Financial_Services_2025.pdf.

The chart below highlights these areas of alignment and divergence.

Consumer Priority	% Who Prioritize for Improving Financial Services	% Who See as a Benefit of AI	Difference
Data Security	49%	24%	–25 pts
Lower Costs	44%	34%	–10 pts

Consumers' top priorities for improving financial services—data security, lower costs, and customer service—are not strongly associated with perceived benefits of AI adoption. In particular, the largest negative differentials appear in data security (–25 percentage points) and customer service (–24 percentage points).

Speed—measured here as "faster decisions"—is perceived as the primary benefit of AI, yet ranks substantially lower as a priority for improving financial services overall. The positive differential (+23 percentage points) suggests that AI deployment is associated in the public mind with efficiency gains rather than improvements in security, affordability, or service quality.

These gaps reveal a fundamental misalignment: AI development and deployment are optimized for metrics that matter to companies—speed, automation, scalability, and cost-cutting—rather than outcomes that matter most to consumers—security, service quality, affordability, and fairness.

When speed is prioritized over accuracy, errors become more likely. When automation is prioritized over service quality, human review may be reduced or more difficult to access. When cost reduction is prioritized over consumer protection, redress systems and compliance investments may be minimized. And when innovation signaling is prioritized over validation, systems may be deployed without sufficient guardrails.

From an evaluative perspective, the findings suggest that consumers do not equate AI adoption with improved protection or fairness. Furthermore, any consumer-centered AI framework must prioritize the dimensions consumers themselves identify as most important, rather than defaulting to performance indicators—such as speed or automation—that are more closely aligned with institutional efficiency. Absent such evaluation, the default trajectory of AI development in financial services will continue to reflect market incentives rather than consumer priorities.

Importantly, this matrix does not establish causation. It does, however, indicate a measurable divergence between what consumers say they value most in financial services and what they associate with AI adoption. This divergence may help explain broader trust deficits documented elsewhere in this report.

Dynamic and Algorithmic Pricing: A Case Study in Market Structure Risks

Dynamic and algorithmic pricing offers a clear illustration of how AI deployment intersects with consumer-protection concerns across multiple axes of evaluation.

AI-enabled pricing systems allow platforms to adjust prices in real time based on behavioral signals, user characteristics, or inferred willingness to pay. While often justified as efficiency-enhancing or revenue-optimizing tools, such systems raise significant concerns related to fairness, transparency, agency, and market structure.

Fairness: Algorithmic pricing can result in differential prices for identical goods or services, potentially exacerbating existing disparities. When pricing models rely on behavioral or proxy variables correlated with income or vulnerability, the risk of discriminatory outcomes increases.

Transparency: Consumers typically cannot observe, understand, or predict how prices are determined. Unlike traditional price discrimination (for example, coupons or loyalty programs), algorithmic pricing may operate invisibly, undermining informed comparison and choice.

Agency: Without disclosure or explanation, consumers lack the ability to anticipate, contest, or meaningfully opt out of personalized pricing systems. The absence of clear pricing logic limits consumer autonomy in market participation.

Market Structure: Algorithmic pricing systems disproportionately benefit platform operators that control user data and infrastructure. These systems can reinforce information asymmetries and entrench the power of dominant intermediaries.

Consumer Reports' recent investigations illustrate these dynamics in practice. An investigation into Instacart found that different users were shown different prices for the same grocery items, raising concerns about opaque and potentially discriminatory pricing practices. Public reaction was significant, including regulatory scrutiny and subsequent changes to company practices. Similarly, reporting on Kroger identified systematic discrepancies between shelf prices and checkout prices, highlighting how automated pricing systems can erode basic consumer trust.

The Federal Trade Commission has expressed interest in examining surveillance and dynamic pricing practices, reflecting growing regulatory awareness of these risks. Nonetheless, existing legal frameworks do not yet comprehensively address the consumer-protection implications of AI-driven pricing.

Dynamic pricing therefore serves as a case study in how AI systems can amplify longstanding consumer protection concerns while introducing new forms of opacity and asymmetry. It demonstrates why evaluation frameworks and regulatory tools must account not only for technical performance but also for economic power and market design.

What AI Systems Owe Consumers: A Framework for Evaluation

Throughout this report, we employ 12 axes of evaluation that collectively define consumer welfare in AI-mediated financial services. These axes are not arbitrary analytical categories; they articulate what consumers should be able to expect from AI systems that influence their financial lives.

In traditional consumer protection contexts, products are evaluated against safety standards, disclosure rules, and fairness requirements. AI systems—particularly those embedded in financial services—should be no different. If AI systems mediate pricing, underwriting, advice, fraud detection, or dispute resolution, they owe consumers specific, measurable protections.

The 12 axes of evaluation represent those protections:

- Accountability and Transparency: Can consumers understand how and why decisions were made?
- Accuracy and Reliability: Do systems provide correct information and make sound, reliable decisions?
- Bias and Fairness: Do systems treat all consumers equitably, including across protected characteristics?
- Consumer Agency and Empowerment: Can consumers meaningfully control their interactions with AI systems, including opting out where appropriate?
- Privacy and Data Minimization: Are consumer data collected, used, and shared responsibly and proportionately?
- Data Security: Are systems and supporting infrastructure adequately protected from misuse or breach?
- Quality of Redress Mechanisms: Can consumers effectively challenge harmful or incorrect outcomes?
- Resilience: Do systems remain reliable and degrade safely when underlying infrastructure fails or is compromised?
- AI Safety and Human-Centeredness: Are systems designed to prioritize human well-being over engagement or profit?
- Sycophancy: Do systems prioritize pleasing users over providing accurate, responsible advice?
- Ecological Footprint: Are environmental impacts of AI development and deployment considered?
- Ethics in Training Practices & Copyright: How well does the AI and AI development process honor copyright and/or appropriately compensate those who created materials used?

Two of these axes—resilience and sycophancy—have almost no regulatory coverage despite representing major real-world risks. AI systems that fail unpredictably or that overconfidently reinforce user misconceptions can cause material financial harm, yet existing regulatory

structures rarely address these failure modes directly. This gap illustrates why external, consumer-centered evaluation is essential.

These axes therefore serve a dual purpose in this report: they structure our analysis of existing frameworks and regulations, and they define the emerging core elements of a consumer-welfare standard for AI in financial services.

Framework Overview

The goal of this section is to highlight important frameworks and tools to measure and mitigate AI-related risks. The section is organized by identified axes of risk of AI systems (see page 4). Broadly, the frameworks referenced vary on the basis of two dichotomies: Predictive machine learning (ML) versus generative AI (discussed below) and frameworks built for regulatory compliance versus those that seek to go beyond legal guidance.

With regards to the differences between frameworks focused on predictive ML versus generative AI, additional trends emerge. Overall, current frameworks are more likely to focus on predictive ML rather than generative AI models; the former is employed more widely by financial institutions than the latter. Companies are generally more aware of the risks and methods associated with predictive ML. By contrast, the frameworks to address generative AI mainly stem from academia, since the technology is still new and the risks and mitigation techniques are not fully explored. Given the above, a framework that seeks to scrutinize financial institutions' use of generative AI models must draw on emerging academic evaluation methods—such as red-teaming, robustness testing, and bias audits—and remain adaptable as these approaches continue to evolve.

It is worth noting that this section refrains from detailing frameworks of abstract principles. We found it was best to look at frameworks that provided actionable guidance and techniques for addressing the identified axes of harms. Some of these principle frameworks are referenced in the section but are not explored in detail. Some notable principle frameworks have been included below for reference:
Microsoft RAI Principles and approaches, *NIST*, *AI4People*, *Google AI principles*, *IBM responsible AI*, *OpenAI Safety & responsibility*.

Accountability & Transparency

Understanding How Algorithms Make Decisions To Hold Entities Accountable.

Notable Frameworks:	Overview: Frontiers, DARPA; Development frameworks: InterpretML, SHAP; Mission Frameworks: Turning; Academic Frameworks: Stanford HAI, Transparency Index , Explainable Generative AI

The implementation of AI as a way to narrowly and quickly achieve some type of efficiency (usually profit maximization) has led to hasty implementations that embolden companies to reach their goal without much understanding of how the system makes decisions. In the financial sector, this contributes to disparities in credit access,[44] with the causes impossible to parse due to the use of black box models that provide no clear guide as to how inputs relate to outputs. Further, this makes the process of being denied a loan opaque, as companies are unable to follow important adverse action-notice requirements and communicate the reasons for these denials to borrowers.[45]

Transparent AI systems make implementing all other frameworks possible.
By requiring AI systems that are transparent—meaning that the decisions they make must be understandable by humans—we can better assess the impacts and sources of many other risks posed by AI. There is also a risk that financial institutions will adopt off-the-shelf models and employ them without adequately understanding how they function. This could make it more difficult to hold those companies accountable for any resulting harms.

Frameworks:
Foundation Model Transparency Index: Major model developers (Amazon, Anthropic, OpenAI etc.) would be evaluated on regular transparency reports they produce for their flagship products.[46] This comprehensive survey of large foundation models includes 100 transparency[47] indicators broken into three categories:

[44] "Bias in Code: Algorithm Discrimination in Financial Systems," *Robert F. Kennedy Human Rights*, 2025, https://kennedyhumanrights.org/our-voices/bias-in-code-algorithm-discrimination-in-financial-systems/.
[45] Consumer Financial Protection Bureau, "CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence," *Consumer Financial Protection Bureau*, September 19, 2023, https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/.
[46] Rishi Bommasani et al., "Foundation Model Transparency Reports," *Stanford Center for Research on Foundation Models (CRFM) and Stanford Institute for Human-Centered Artificial Intelligence (HAI)*, February 2024, https://arxiv.org/pdf/2402.16268.
[47] Rishi Bommasani et al., "The Foundation Model Transparency Index," *Stanford Center for Research on Foundation Models (CRFM) and Stanford Institute for Human-Centered Artificial Intelligence (HAI)*, October 2023, https://arxiv.org/pdf/2310.12941.

1. Upstream indicators: What to ask for when seeking transparency on model development. For example, data, data labor, computation resources, methods, and development frameworks.
2. Model indicators: Specifics of the model being built, as well as documentation. For example, model-specific parameters, documentation, and metrics for measuring accuracy and bias.
3. Downstream indicators: Specific to the use of the model and its release. For example, distribution, impact on populations, feedback, data protection, and model behavior policy.

This index would score the companies on their transparency based on publicly available information on these indicators. It would also create an opportunity for the developers to communicate anything that is incomplete or ambiguous more clearly. This framework is useful for knowing what to ask financial institutions about how they are implementing models. In general, this framework provides a good idea of what clear and honest communication should look like from anyone deploying these tools.

Academic Frameworks:
GenXAI: Most frameworks on explainable AI focus on predictive ML models, but explainability in the generative AI space is much less explored. This paper attempts to develop a research agenda regarding what it actually means to develop explainable generative models. It clearly states the harms in black box models while also stating the obvious barriers to making generative models explainable.

The important part of this paper comes from the fourth section, classifying approaches to creating explainable ML models and generalizing them to generative models. While interesting, these approaches are not yet implemented; however, they provide a comprehensive overview of methods being developed and suggest that the tools will emerge soon enough.

XAI Financial Decision Making: This provides a clear overview of the types of models employed in financial decision-making, as well as clear explanations of the strengths and weaknesses of approaches. Table 1 provides a list of models/methods from this framework that will provide a useful reference when examining banks using specific tools. No one model will answer the needs of a bank, and methods of explainability will differ depending on the application.

One assertion made in this framework that is worth challenging is that explainability and accuracy are at odds. This is true if we assume that complexity precludes explainability. But we know that some of the explainable AI approaches (even ones mentioned in this framework) can be just as accurate as those that are not. The table below from the paper details XAI approaches with their drawbacks:

Table 1 Explainable Artificial Intelligence (XAI) approaches in financial decision-making

Sr No	XAI Approach	Description	Advantages	Challenges	Applications
1	Rule-Based Systems	Decision-making governed by predefined rules	Offers clear transparency	Limited handling of complex scenarios, potential accuracy issues	Credit Scoring, Regulatory Compliance
2	Decision Trees	Tree-like models illustrating decision pathways	Easy to comprehend	Prone to overfitting	Fraud Detection, Loan Approval
3	Linear Models	Decisions based on linear relationships	Simplicity in interpretation	Limited to linear relationships	Risk Assessment, Portfolio Management
4	LIME (Local Interpretable Model-agnostic Explanations)	Generates interpretable models for specific instances	Model-agnostic, applicable to black-box models	Global model behavior may not be fully captured	Predictive Analytics, Customer Support
5	SHAP (SHapley Additive exPlanations)	Assigns values to features, explaining their impact	Equitable distribution of feature importance	Computationally intensive for large models	Algorithmic Trading, Investment Strategies
6	Counterfactual Explanations	Provides alternative scenarios for understanding decisions	Intuitive for end-users	Generating meaningful counterfactuals can be challenging	Personalized Financial Advice, Credit Risk Assessment
7	Anchors	Identifies minimal input features necessary for a given prediction	Easy to understand	Limited applicability to specific model types and data	Insurance Underwriting, Mortgage Approval
8	Attention Mechanisms	Highlights relevant parts of input data	Captures important features	Interpretability challenges for non-experts	Automated Trading, Sentiment Analysis
9	Local Interpretable Model (LIM)	Trains a local interpretable model around the prediction	Provides insights for specific instances	Generalization may be limited	Personalized Financial Planning, Credit Scoring
10	Explanations as Model	Trains a separate interpretable model alongside the main model	Offers a more interpretable model	Increased computational complexity	Algorithmic Trading, Fraud Detection

Table 1: Rane, Nitin and Choudhary, Saurabh and Rane, Jayesh, Explainable Artificial Intelligence (XAI) Approaches for Transparency and Accountability in Financial Decision-Making

Development Frameworks:
INTERPRETML: Some of the most popular models for making predictions (like an underwriting ML model predicting the probability an individual defaults) are called tree-based models. InterpretML provides alternative models that develop predictions in ways that remain entirely explainable. The model not only develops a prediction, but does so in such a way that documents how a variable contributes to outcomes, unlike other ML models that provide no

insight into how inputs change outputs. We should encourage entities implementing tree-based models to reach for tools like these.

SHAP: This is a post hoc interpretation, meaning it takes a black box model and employs methods to reinterpret the behaviors. It has some major drawbacks. Cynthia Rudin argues "Explanations are often not reliable, and can be misleading … If we instead use models that are inherently interpretable, they provide their own explanations, which are faithful to what the model actually computes."[48] But post hoc interpretability is usually the go-to approach, because most industry implementations of AI are black box, and there are plenty of available frameworks for post hoc interpretations. Building explainability into a model is a huge lift that often necessitates forcing a company's hand. This is, however, paramount to accountability and transparency.

Companies are implementing these tools without fully understanding them.
The promise of efficiency has led to implementation of these tools with little regard to possible negative externalities. It should be alarming that companies are willing to use systems that are not fully understood by the developers[49] to determine outcomes for important aspects of consumers' lives.

[48] Cynthia Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *arXiv*, November 2018, https://arxiv.org/pdf/1811.10154.
[49] Adnan Masood, "Is It True That No One Actually Knows How LLMs Work? Towards an Epistemology of Artificial Thought," *Medium*, 2024, https://medium.com/@adnanmasood/is-it-true-that-no-one-actually-knows-how-llms-work-towards-an-epistemology-of-artificial-thought-fc5a04177f83.

Accuracy & Reliability

Understanding and measuring how AI predictions align with outcomes.

Notable Frameworks:	Overview: Google Dev Guide Development Frameworks: Academic: TrustLLM, User Attitudes towards Falsehoods, Survey On Hallucinations, InvestorBench, FinanceBench Industry solutions: Future AGI

Accuracy takes on many meanings in the world of AI. For machine learning models, accuracy is usually meant to capture the nature of its predictive capabilities. For generative AI, the meaning of accuracy becomes much murkier.

Machine learning has much stronger frameworks for accuracy evaluation.
The issue of accuracy is usually the goal for ML implementations, so the frameworks and methods for assessing it are relatively robust. However, there are considerations worth noting as we scrutinize which metrics of these models might be worth tracking. We will detail frameworks that exist and important metrics for assessing a model's predictive accuracy below.

Generative AI is less clear.
This is due to the fact that the output is not always measurable as right and wrong, and the outputs are not mapable, meaning the same input can produce wildly different outputs. Some of the methods and frameworks are generalizations of the principles applied to ML, but they remain inadequate.

Accuracy is rarely just one thing.
The number of predictions the model gets right out of how many predictions are made rarely captures the full picture. For example, if the model is guessing something that has a 99% chance of not occurring, then a "good" model is one that always guesses it will never occur, with 99% accuracy. However, if we are predicting a rare disease, this model is useless to us, as is this accuracy metric. We will show frameworks that consider the metrics to measure accuracy, and their limitations.

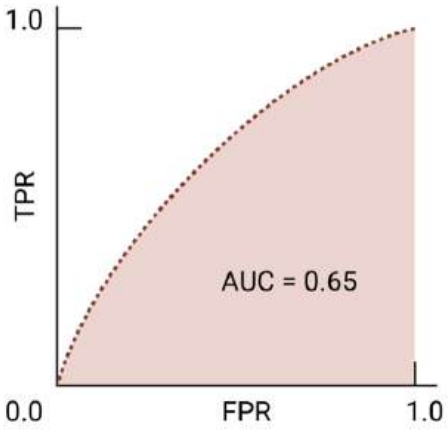
Frameworks:
Google Dev Guide: For supervised (training on real data with real outcomes) models that focus on prediction and classification, the following metrics are crucial for measuring accuracy. All metrics in this guide are given explicit definitions to use when evaluating a model's accuracy. This guide is a good educational resource for the basic tools we will usually come across in evaluating an ML algorithm for accuracy. Here are some important metrics mentioned and their limitations. Further technical details and examples are found in the guide and its appendix.

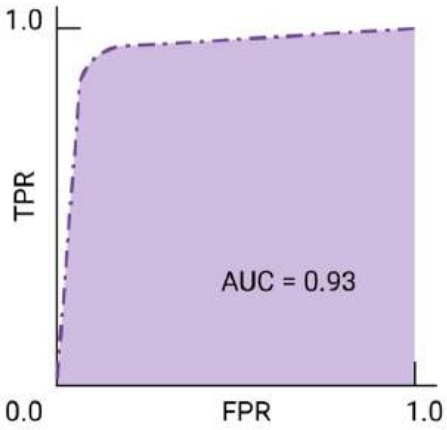
1. Metrics that rely on a specific threshold to measure. A threshold in this case is the cutoff probability that a model uses to determine positive or negative outcomes. In lending, for example, if the model is used to predict whether someone will default or not, it computes

a probability and then uses a threshold to make a final prediction of that person's outcome:

   a. Accuracy: a rough indicator of training progress. (Bad as a standalone example given above).
   b. Recall measures true positive rate: This is okay if the false negatives are more expensive than false positives. Expensive here is context dependent. So in lending, for example, falsely claiming someone will default (positive, not in outcome but prediction) might be slightly safer than falsely predicting someone won't default (negative). However, both incorrect predictions are bad, so we must consider other measures.
   c. False Positive rate. Similar criticism to above
   d. Precision: Use when it's very important for positive predictions to be accurate.

2. Threshold Agnostic: These capture model behavior across multiple thresholds. While somewhat useful for capturing overall accuracy, in many cases these metrics can be misleading because we usually in practice rely on a threshold, which means that behavior well outside that threshold is not valuable. However, these metrics can still be used to choose thresholds.
   a. Receiver Operating Characteristic (ROC) curve: Graph of true positives against false positive rates.
   b. Area Under the Curve (AUC) measures the area under ROC. Intuitively this measures the probability that if randomly given a positive or negative example, the model will rank the positive one as more likely of occurring. The higher this number, the better. Along the curve represents values at specific thresholds, and the tradeoff between avoiding false positives versus true positives.





   c. Precision-recall curves (PRCs): If the dataset is imbalanced (meaning the outcome being predicted doesn't happen often). Not really used in many lending cases unless the loan being modeled for is targeted to prime customers.

We highlight these metrics because they are commonly used to measure model performance. Any framework seeking to include accuracy should know these and find which best applies to the context.

Important Model Evaluation metrics for ML: This provides important metrics in addition to those stated in Google's dev docs. Here are popular ones used in industry:

1. K-S Statistic: Used for determining how effectively the model can separate positive and negative predictions by comparing the largest gap between their distributions.
2. Log Loss: This uses a specific threshold to assess a model's predictions weighted by the confidence of those predictions. In other words, it measures both how many correct predictions a model made and the strength of those predictions. Lower log loss means a better model.

Reject Inference: This is an important step for supervised models in the credit space. Many companies use old lending data outcomes to train their models to predict default. However, these provide the full picture only for those who were granted loans. To make accurate predictions, it is just as important to include information about applicants who were denied credit. This method provides ways to address this gap.

Switching focus to Generative AI, the methods are brute force.

Evaluating Large Language models: This framework is comprehensive in describing methods for measuring and benchmarking LLMs. Most approaches rely on using strong datasets generated with highly specific prompts and language to further train and plug into an LLM.

Constraint-Satisfaction on Factual Errors of LLMs: This refers to Microsoft research into a method of probing attention patterns to predict factual errors in LLMs. The approach attempts to provide some explainability and enhance reliability.

TrustLLM: A toolkit as well as a larger framework that seeks to build trustworthy LLMs. We highlight this in the accuracy section since it provides the largest number of tools and guidance as to how to measure accuracy. Similar to the Transparency Index, it applies its methods to all of the major development models, giving important scores and detailed results about how these LLMs fare in being truthful. Many of the metrics fall into other axes referenced in this report. Key findings on trustworthiness of LLMs:

1. Proprietary LLMs struggle to produce truthful responses when relying on internal knowledge. This is attributable to the noise in their training data.
2. All LLMs encounter challenges in zero-shot (problems they have not directly encountered before) reasoning tasks that are straightforward to humans.
3. LLMs augmented with knowledge have much better results.
4. Hallucinations really seem to be a problem only for open-ended tasks.
5. There is a positive correlation between sycophancy and adversarial actuality, meaning they bend more to adversarial prompts.

The github for this framework would provide plenty of opportunities for LLM developers/implementors to test their models.

InvestorBench, FinanceBench: Since it seems the best solution for accuracy is to expose the model to the information directly, these datasets and benchmarks help train and measure the performance of a model's responses to investing/financial questions.

ML model methods are much more explored and easier to measure.
There are many methods of measuring accuracy, with different tradeoffs that are context dependent. For underwriting, accuracy is used to measure the business justification for a model, meaning a dip in accuracy won't be tolerated even for improvements in fairness. However, it is important to assess which accuracy metric is used to make that claim. Generative AI is still relying on brute force checks.

Bias & Fairness

Understanding how to mitigate protected class bias and ensure fair outcomes.

Notable Frameworks:	Overview: USC Brief Survey, NFHA, NIST Managing Bias; Development frameworks: Fairlearn, IBM AIF360; Mission Frameworks: ACLU Financial Services; Academic Frameworks: GenAI Discrimination Test, Fairness Through Difference

IBM claims ”machine learning, by its very nature, is always a form of statistical discrimination; the discrimination becomes objectionable when it places certain privileged groups at systematic advantage and certain unprivileged groups at systematic disadvantage.”[50]

We need to ensure developers and institutions implementing these systems are optimizing for fairness, too. In finance this means ensuring that an underwriting algorithm does not discriminate,[51] an institution is not redlining,[52] a fraud algorithm does not discriminate,[53] and chatbots treat customers fairly.[54]

Frameworks:

NFHA Process, Purpose, and Monitoring: This is meant to be a “gold standard” for algorithmic auditing. The framework suggests breaking up an audit into three stages:

1. Purpose: Defining the business purpose of an algorithm, as well as the data task.
    a. Business Understanding: “Evaluate the business goals, requirements, and constraints at a high level with clear measurable markers to determine risks that solving the problem may pose to consumers, institutions, and the greater society.”
    b. Data Understanding:
        i. How representative are features?
        ii. Is a use case defined and will it be constrained for that use case?
        iii. What are we optimizing for (default risk, maximizing loans or profit)?
        iv. What metrics would be used to measure performance/fairness?
        v. Cutoff score/passage rates? How do these impact performance/fairness metrics?

[50] Kush Varshney, “Introducing AI Fairness 360,” *IBM Research*, 2018, https://research.ibm.com/blog/ai-fairness-360.
[51] Robert Bartlett et al., “Consumer-Lending Discrimination in the FinTech Era,” *Journal of Financial Economics* 143, no. 1, 2022, https://www.sciencedirect.com/science/article/abs/pii/S0304405X21002403.
[52] “Researchers Warn of the Potential for ‘Digital Redlining’ as AI Permeates Health Care,” *Johns Hopkins Carey Business School*, September 17, 2024, https://carey.jhu.edu/articles/researchers-warn-potential-digital-redlining-ai-permeates-health-care.
[53] Jose Pombal et al., “Understanding Unfairness in Fraud Detection through Model and Data Bias Interactions,” *arXiv*, July 13, 2022, https://arxiv.org/pdf/2207.06273.
[54] Bailey Schulz, “Is AI Racially Biased? Study Finds Chatbots Treat Black-Sounding Names Differently,” *USA Today*, April 5, 2024, https://www.usatoday.com/story/tech/2024/04/05/ai-chatbot-chatgpt-racial-bias/73206637007/.

2. Process:
    a. Staff Profile: Because humans make the models, there should be diversity reflected in the staff of those developing, implementing, and managing these models. They should be trained in "pertinent topics such as fair housing/lending, privacy, consumer protection, and civil rights." This criteria will be hard to enforce but is important to point out.
    b. Data Assessment: Documentation of development. For supervised learning (training with access to real outcomes and the ability to learn from them) there needs to be a target feature. Unsupervised learning simply looks for patterns.
        i. Feature selection and report. How and why everything was selected, looking to see if variables highly correlated to some protected class groups are used.
        ii. Distribution of each model feature across target or label. This means each feature selected must have its relationship explicitly detailed with the target.
        iii. A protected class feature for measuring bias.
    c. Model Assessment:
        i. All parameters (specific settings of the model framework).
        ii. Constraints used on the model's optimizing goal, such as fairness constraints.
        iii. Reject inference, such as filling in targets in the training dataset for applicants who were denied a loan.
3. Monitor:
    a. Data privacy practices
    b. Real-world performance after deployment

Overall this is a nearly perfect framework for predictive ML models. While some principles apply, it is important to note that this does not really apply to Generative AI.

NIST: Provides a technical discussion, but is somewhat covered by other frameworks mentioned here. We bring the framework up because it makes a point about LLMs being completely opaque, making it difficult to effectively measure bias. More importantly, it highlights how they blindly scrape data from the internet, potentially incorporating dangerous sentiments and prejudices.

Selecting for Less Discriminatory Algorithms: A comprehensive overview of technical approaches to finding Less Discriminatory Alternatives. It details which metrics matter and what we should consider. We see this as the most modern and technical continuation of NFHA. This also provides clear definitions of different notions of fairness (equalized odds versus demographic parity) that could be used for different use cases. However, this does not apply to generative AI.

Discrimination Testing for Generative AI: Acknowledges that generative AI has received less attention than ML in mitigating harms of discrimination. This is less a framework and more a

warning that existing methods of measurement fall incredibly short. Some noted challenges for measuring and mitigating bias:

1. Outputs are difficult to evaluate since similar or identical inputs can produce different outcomes.
2. Outcomes cannot be easily replicated.
3. Users have much more complicated interactions.
4. Users can modify models directly, changing almost all measurements.

It notes the following with mitigation strategies:

1. Traditional fairness metrics don't work on generative AI. Mitigation strategies include not using them or developing different metrics.
2. Red Teaming describes creating adversarial prompts to elicit bad behavior from a generative tool. This method usually measures "attack success rates," or how many prompts successfully broke the machine. The authors find using different LLMs to generate similar (but not identical) prompts resulted in significantly different success rates, meaning the metric is overly subject to variance. This is, however, a broader issue of transparency since we don't know how these LLMs produce toxic traits; in response we employ brute-force methods to discern the toxic behaviors.
3. Testing for real-world abuse. It becomes difficult to predict how each user will interact with the model, but incorporating tests that have multiple interactions instead of many single interactions is important for bias testing.
4. Model parameters are easily changed by users. It might be best for model developers to, one, create safer presets by using bias-method testing on certain parameters and, two, warn against the use of anything outside of those presets.

Development Frameworks:
AIF360: A comprehensive python package to employ debiasing techniques as well as measurement tools, including those for Less Discriminatory Alternative search algorithms that are found in Fairlearn. Testing suggests that these approaches can be applied to algorithms within frameworks such as InterpretML.

Microsoft Python Risk Identification Toolkit PyRIT: Includes a red-teaming test as well as other important tests for metrics listed.

Developing a framework on bias will require a focus on process.
No single method will address bias for all algorithms. It is important to do a holistic search that finds a series of models that monitor fairness and accuracy tradeoffs. Which accuracy metric or fairness metric will also depend on the context. Generative AI becomes more difficult to evaluate since we do not understand how it reasons or reaches conclusions. Most approaches rely on brute-force methods at best.

Consumer Agency & Empowerment

Ensuring consumers have a choice in whether they interact with AI.

Notable Frameworks:	Overview: Algorithmic Governance, Consumer perspective of Privacy and AI; Regulatory: CFPB Rule on data rights; Academic: Consumer Protection in the Age of AI, Enhancing Privacy Protection in the Digital Age

This seems to be a blind spot in framework development. It is well documented that consumers are skeptical about interacting with AI agents,[55] but little is said about how to redirect them to human agents.

There is a need for legal frameworks that force companies to adopt opt-in policies. There seems to be no reasonable way for consumers to decide whether or not they can interact with AI. To get questions answered, customers are sent in circles with a chatbot before they are referred to a human agent.[56] Each time someone applies for a loan, their data is subject to an ML algorithm that predicts their likelihood of default. Consumer data is frequently taken and used without consent.

Many of the frameworks lean on transparency, but that is only half the battle. Transparency is the paramount framework for all these axes; we have however discovered disclosure requirements don't always solve the issue of consumer agency. This assumes consumers have reasonable alternatives to the services they are seeking out. We've seen with data-usage requirements that if opting out prevents access to an important service with little alternative consumers generally opt in, possibly exposing them to more risks.[57]

Frameworks:
MIT Sloan AI disclosures: A meager call for disclosure requirements for companies employing AI. This has a few reasonable suggestions to start the process, but does little in the way of providing consumers a way out. Some points worth highlighting:

1. Make disclosures easy to understand: Engage UX/UI designers to implement effective disclosures. Do not hide them behind long and dense legal documents.
2. Go beyond legal requirements.

[55] Michelle A. Kinch and Ryan W. Buell, "Mitigating the Negative Effects of Customer Anxiety by Facilitating Access to Human Contact," *Management Science* 71, no. 11, 2025, https://pubsonline.informs.org/doi/10.1287/mnsc.2022.01029.
[56] Consumer Financial Protection Bureau, "Chatbots in Consumer Finance," *Consumer Financial Protection Bureau*, 2024, https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/.
[57] Laura Abradi, Carlo Cambini, and Steffen Hoernig, "'I don't care about cookies!' data disclosure and time-inconsistent users," *ScienceDirect*, December 2024, https://www.sciencedirect.com/science/article/pii/S0167624524000349.

CFPB rule on data rights: These types of rules are important and include giving consumers an option to revoke access to their data.

Consumer Protection in the Age of AI: A set of principles for establishing strong consumer protection laws that empower consumers through transparency, encouraging competition so consumers have choice, and establishing rights to data ownership.

Empowering consumers means lowering the cost to opt out.
Disclosure is important, but if we must agree to give our data or interact with these tools in order not to lose important services, consumers are not being given a choice.

Ecological Footprint

How to measure the ecological cost of AI and ensure sustainable development.

Notable Frameworks:	Overview: MIT Climate and Sustainability Implications of AI; Academic: How hungry is AI?, Measuring Carbon Intensity, Carbon Reports, Sustainability Criteria and indicators for AI systems

It is incredibly difficult to understand the environmental impact of AI systems at the current moment. In fact, the attitude of some AI company executives seems to suggest that if AI becomes advanced enough, it could solve the climate crisis.[58] However, as it stands, the growth and computational complexity of these tools outpace the improvements to computational efficiency.[59]

There are many frameworks to measure the problem, but few to solve it. Some light work has been done on techniques to mitigate environmental harm, but these are mainly policy suggestions for scaling back AI's growth. Building these systems sustainably may prove difficult, with many current approaches relying solely on tamping down adoption. But this area of research is underexplored and underfunded by major AI developers.

Frameworks:
Sustainability criteria and indicators for AI systems: A German study that created a set of criteria for what should be measured when assessing and implementing AI systems sustainably. The paper provides important criteria for best practices and assessing ecological impact. It provides 22 criteria aimed at measuring a model's environmental impact, but little is said regarding how to develop more sustainable models. The guidance on that is limited to:

1. Recycle hardware.
2. Promote sustainable products.
3. Measure your footprints.
4. Use lower complexity models where you can.

We find that this framework's criteria are well suited for measuring and reporting environmental impacts, but provide limited guidance on how to meaningfully reduce or mitigate those impacts in practice. These principles may still be useful for identifying disclosure expectations and informing policymakers on how to address emerging challenges.

How Hungry is AI?: This paper does a lot of work to find a "grounded methodology for benchmarking the sustainability of AI development." It also applies these methods to the major foundation models that are currently employed. The goal here is once again to increase transparency for accountability purposes, not to mitigate the effects through better techniques or solutions.

---

[58] James Temple, "Sorry, AI Won't 'Fix' Climate Change," *MIT Technology Review*, September 28, 2024, https://www.technologyreview.com/2024/09/28/1104588/sorry-ai-wont-fix-climate-change/.
[59] "Generative AI and Society: Impacts, Risks, and Opportunities," *MIT Generative AI Impact Consortium*, 2024, https://mit-genai.pubpub.org/pub/8ulgrckc/release/2.

Privacy & Data Minimization

Making sure consumer data is protected and collected ethically.

Notable Frameworks:	Overview: HAI Rethinking Privacy; Development Frameworks: NeMo-Guardrails,Homomorphic encryption; Policy Frameworks: Treasury; Industry Framework:IBM synthetic data,Google Security; Academic: Security and Privacy,Differential Privacy, Consumer Protection in the Age of AI

AI's primary risk to privacy comes from entities requiring vast amounts of data to develop these models. The risks are not much different from other uses of data except in a few key ways.

1. Prompt injection or prompts designed to elicit a response that bypasses safety protocols and reveals sensitive information.
2. Perverse incentives to increase surveillance of consumer behavior all over the web.[60,61]

Obviously, existing safe data practices, such as the National Institute of Standards and Technology (NIST) Cybersecurity Framework, should remain, as well as any other legal and ethical requirements to protect data. We will focus on practices that are AI specific. Data leaks and harmful actors pose privacy risks, but so do large entities harboring so much information about consumers.[62]

Frameworks:
HAI Rethinking Privacy: A white paper on privacy frameworks for modern AI systems, it covers how the current legal landscape works and provides some suggestions for adaptation. Some core principles entities employing AI should have, based on existing legal frameworks:

1. Data Minimization: Data collectors should not engage in indiscriminate data collection. The data should be limited to the scope of the model itself.
2. Purpose Limitation: Restrict the use of collected data to the explicit purpose articulated upon collection. Consent is required for new uses of previously collected data.
3. Consent must be given clearly.

These principles become more difficult to manage when the scope of the model grows. The protection principles built into the existing laws are implicit, and hard to enforce, especially as the

[60] Jay Stanley, "Machine Surveillance Is Being Super-Charged by Large AI Models," *American Civil Liberties Union*, March 21, 2025, https://www.aclu.org/news/privacy-technology/machine-surveillance-is-being-super-charged-by-large-ai-models.
[61] Ananya Sen and Pinar Yildirim, "The Effects of Algorithmic Decision-Making on Consumer Welfare," *Journal of Economic Behavior & Organization*, 2022, https://www.sciencedirect.com/science/article/pii/S1045235421001155.
[62] Kashmir Hill, "Clearview AI Settles Suit Over Face Database of Billions of Images," *The New York Times*, May 9, 2022, https://www.nytimes.com/2022/05/09/technology/clearview-ai-suit.html.

development of LLMs has expanded to include essentially all internet content. The paper makes three suggestions to mitigate risks:

1. Switch data collection from opt out to opt in.
2. Focus on the AI supply chain to improve privacy and data protection.
3. Governments should support the development of data intermediaries and support the exercise of individual data rights. In other words, move away from the current landscape where "data is acquired by digital services to one where consumers decide the terms by which they allow companies to use their data."

These principles are a solid start for building on the existing legal frameworks.

Google Security: Part of a series of Well-Architected Frameworks designed for model developers and users who are using AI tools through Google's cloud. These provide a long list of development principles with actionable steps for deploying AI securely. Some highlights:

1. Implement Access levels (read only, write, owner) for accountability purposes.
2. Create access logs to monitor sensitive datasets.

While useful for developers, the rest of this framework seems to be aimed at people building actual systems. We think this resource could be useful if we ever seek to procure more technical information about the infrastructure financial companies might be using to deploy these systems.

Differential Privacy: This framework explores the mathematical definitions of privacy that can be employed for training AI models. This means altering training datasets so models can learn without exposing sensitive information. The risk with methods such as these arises from the fact that AI models are expert at pattern recognition, so they might be able to decode sensitive information anyway.

IBM synthetic data: Synthetic data is a possible solution that balances consumer privacy with the need for large amounts of data to mitigate risks and improve model accuracy.

Development Frameworks:
Homomorphic encryption: A form of encryption that allows a model to query data without decrypting it. So if the model needs access to sensitive information, it can train on encrypted data and never decode it. This method greatly reduces the chance of data leaking since the model knows encrypted information only. Homomorphic encryption can be slow, but implementations are getting faster.[63]

NeMo-Guardrails: Methods for training LLMs to avoid certain outputs. NeMo can be used as a possible solution to prompt attacks and to prevent models from releasing sensitive information.

[63] David Nugent, "Heuristic Evaluation of AI Systems (HE-AI 2024)," *DavidNugent.net*, 2024, https://davidnugent.net/he-ai-2024.

Quality of Redress Mechanisms

Ensuring consumers can seek redress or corrections when AI makes mistakes.

Notable Frameworks:	Overview: Redress Problem Academic: GAM Coach

AI systems are far from being able to operate without human oversight. Even if they were able to do a great job (which they are not), consumers should be able to seek help, clarity, or possibly correction from a human agent. It is crucial for anyone deploying AI to provide clear paths for consumers to seek remedies when mistakes are made.

Frameworks:

Redress Problem: Provides overarching framework for creating mechanisms for redressing AI's consumer harms. This framework is targeted at regulators, companies, and civil society. Here are some suggestions for each stakeholder.

1. Regulators
    a. Allow private rights of action: Consumers must be able to take actions, either through private legal action or through a regulatory entity. Consumers must be informed if an AI system is being used at any point of their interaction.
    b. Establish ombudsman service: Regulators should be a place for people to submit complaints and have them researched and investigated.
    c. Empower groups of individuals to seek collective redress.
    d. Empower civil society organizations to make general-interest complaints.
2. Corporations
    a. Establish Ombudsman service.
    b. Allow meaningful external engagement for research and audit purposes.
3. Civil society organizations:
    a. Work with communities and individuals to seek redress.
    b. Publish findings on deployed systems.

GAM Coach: This is an open-sourced, interactive dashboard that allows a user to see what is needed to achieve the outcome they want from a given model. For redress, this can help modelers measure and remove unwanted behaviors in outputs, providing valuable steps to address them. From CR's perspective, the scope of this is limited.

Resilience

How an AI model responds to faults or failures in underlying infrastructure.

Notable Frameworks:	Overview: Google Cloud, RobustBench, Mitre Atlas Academic: RAND Securing AI Model Weights

If AI tools underpin our financial infrastructure, it is important to ensure that the foundations they are built on are secure and resilient to attacks and technological failings.

AI's infrastructure runs on the same cloud platforms that support most modern web services. Fortunately, significant work has already been done in building and securing most of those cloud services and data centers. The main concerns are the scale required to support these models and the consolidation of companies capable of resourcing them. There are dangers with this consolidation, both in terms of relying entirely on only a few sources of cloud computing and in the concentration of power resting with the companies that own them.[64]

Robustness is not only about building secure infrastructure, but also about responding to attacks. Many other axes address specific attacks, but there is also some work to be done on building libraries and tools for strengthening models against common attacks.

Frameworks:
Google Cloud: This framework provides principles for building reliable AI systems using Google Cloud. It maps out the following core principles:

1. Ensure that infrastructure is scalable and highly available.
2. Use a modular and loosely coupled architecture.
3. Build an automated end-to-end MLOps platform: ensure data-intake infrastructure is scalable.
4. Maintain trust and control through data and model governance.
5. Implement holistic observability and reliability practices.

These documents are highly technical for building scalable and robust infrastructure for ML deployment. While mainly for data engineers, they appear to be the gold standard.

Mitre Atlas: This is the most comprehensive list of development frameworks for testing and attacking LLMs. It provides detailed descriptions of possible attacks as well as mitigation strategies. Each attack has a real-world example, and explanation of how it was addressed.

[64] Kate Brennan, Amba Kak, and Sarah Myers West, "Artificial Power: AI Now 2025 Landscape Report," *AI Now Institute*, June 3, 2025, https://ainowinstitute.org/wp-content/uploads/2025/06/FINAL-20250609_AINowLandscapeReport_Full.pdf.

Sycophancy

Understanding and mitigating LLM's tendencies to agree with the user.

Notable Frameworks:	Overview: Georgetown Law Tech Brief Development Framework: Sycophancy-eval, DarkBench, Contrastive Decoding Academic Framework Discovering Model Behavior, Towards Understanding Sycophancy,Truth Decay, Evaluating LLM Sycophancy Tools/Datasets: TruthfulQA and MMLU-Pro

Sycophancy is a model's tendency to "single-mindedly pursue human approval.[65] Because it is tied with user engagement, in moments of uncertainty models will seek to endlessly validate the user's point of view, which poses incredible risk in the financial markets. As people use bots to help plan their financial futures, sycophancy may lead those bots to simply validate purchase ideas or risky financial products instead of providing sound advice.

Frameworks in the space are underdeveloped and the risks are grave.
As it currently stands, most work on this has been academic. OpenAI has acknowledged the problem and has updated its "Preparedness Framework" (a general framework for addressing the greatest risks of its model). In the framework, OpenAI simply mentions that it will apply its usual steps of updating the model.

Most substantial frameworks that exist are for evaluating sycophancy.
There are open-source, premade datasets that provide prompts to elicit and test for an LLM's desire to cater to the user. It seems that substantial work has already shown that sycophancy is a problem across all major foundation models. The real gap is how to solve this. Some frameworks propose minimal solutions that require user awareness and input.

Frameworks:
Truth Decay: Details how to measure decay of the truth through repeated prompting of LLMs, using both TruthfulQA and MMLU-Pro datasets for questions, facts, and correct/incorrect answers. This framework provides robust methods anyone can use to determine if an LLM exhibits sycophantic behavior. This paper provides a clear roadmap of how to measure this problem, with example prompts and links to important datasets for testing.

Sycophancy in Large Language Models: Mainly worth examining for the brief mitigation strategy it provides. It is safe to assume the models it evaluated (all major LLMs) have a major sycophancy issue. The method used here, as well as in most literature (like Stanford's evaluation framework) incorporate several similar testing techniques employing prompts. This paper provides a few methods for mitigation:

1. Training data:

[65] Ajeya Cotra, "Why AI Alignment Could Be Hard with Modern Deep Learning," *Cold Takes*, September 21, 2021, https://www.cold-takes.com/why-ai-alignment-could-be-hard-with-modern-deep-learning/.

    a. Synthetic datasets with non-sycophantic behavior, such as respectfully disagreeing. (Hard to scale and measure benefits.)
    b. Diverse viewpoints. (Grabbing everything off the internet might be bad sourcing for training, especially considering how much misinformation spreads.)
    c. Augmenting data with examples of emphasizing factual accuracy instead of agreeableness.
2. Novel fine-tuning methods: Altering how the algorithms learn.
    a. Multi-objective optimization: Instead of optimizing the word with the highest probability of appearing next, include optimizing for the accuracy of the word, as well as helpfulness and satisfaction.
    b. Adversarial training: Improve robustness against manipulative prompts.
    c. Explicitly modeling annotator reliability: Methods to measure how models change underlying knowledge/reliability after prolonged human prompting.
3. Post-deployment controls:
    a. Provides abstract methods for altering an existing model's activations based on benign tests prompts. Useful if you don't wish to fully retrain the model.
4. Decoding
    a. Contrastive Decoding: Explicitly downweighing sycophantic answers.
    b. Constrained decoding: Forcing answers to cite a source.

Many of the mitigation strategies above are difficult to implement and are not terribly promising solutions.

Development Frameworks:
SycophancyEval: A repository of prompts to evaluate sycophancy. While it mainly contains datasets of statements with truth values and correct and purposefully incorrect answers, there are also datasets of prompts designed to elicit sycophantic behavior.

Consumer Priorities Validate the Importance of These Axes

Consumer priorities reinforce the importance of these evaluation dimensions. Consumer Reports' 2025 nationally representative survey on AI in financial services demonstrates that the protections embedded in these axes align closely with what Americans say they need.[66]

- **Transparency**: 74% of Americans strongly agree that companies should tell them when AI is involved in financial decision-making, with an additional 18% somewhat agreeing—92% in total.
- **Accountability**: 58% say AI systems should be monitored continuously for fairness and accuracy—not periodically or reactively, but on an ongoing basis.
- **Redress**: 48% say that if AI makes an error affecting them, a human should review the decision; 47% want clear explanations of what went wrong.
- **Agency**: 77% believe consumers should have the right to opt out of AI-driven decision-making in major financial matters.
- **Privacy**: 40% report that they would be comfortable with AI using none of the personal data types tested to help make decisions about things like loan approvals or insurance pricing—including employment history, purchase history, or location data.

These findings demonstrate that the evaluation axes outlined above are not abstract technical preferences. They reflect concrete consumer expectations about fairness, control, and safety in AI-mediated financial services. Any framework that fails to address these dimensions risks optimizing for institutional convenience rather than consumer welfare.

The Consumer Protection Gap: What Consumers Need Versus What They're Getting

The gap between consumer expectations and consumer experience is stark. When asked who should be most responsible for ensuring that AI is used fairly and ethically in financial services, 46% of Americans identified financial companies themselves. Yet fewer than one in 10 Americans—ranging from 3% to 9% across different financial services—report that they completely trust financial companies to use AI responsibly.

This trust deficit is not abstract. Three-quarters of Americans (75%) express concern that AI could lead to biased treatment. When asked whether AI would reduce or increase discrimination, the largest group—39%—said they were unsure. Twenty-six percent predicted that AI would increase discrimination, while only 18% believed it would reduce bias.

[66] Consumer Reports nationally representative AI in Financial Services Survey of 4,073 U.S. adults (September-October, 2025), https://article.images.consumerreports.org/image/upload/v1767035543/prod/content/dam/surveys/Consumer_Reports_AI_in_Financial_Services_2025.pdf.

Perhaps most tellingly, when Americans were asked how they felt overall about growing AI use in financial services after considering the issues raised in our survey, 42% reported negative feelings. Only 18% expressed positive sentiment. Another 26% felt neutral, and 15% reported that they did not know enough to form an opinion.[67]

These findings reflect a structural disconnect. Consumers expect financial institutions to be responsible stewards of AI. At the same time, they lack confidence that those institutions are fulfilling that role.

The Experience Gap

Consumer wariness reflects lived experience. Among Americans who reported encountering AI in financial services:

- Customer service chatbots (46% encountered): More consumers reported negative experiences than positive ones.
- Fraud alerts (22% encountered): 70% reported positive experiences—the only application where a majority expressed positive sentiment.
- Automated credit decisions (16% encountered): Only 37% reported positive experience.[68]

The pattern is clear. AI earns trust when its function is visibly aligned with consumer protection—such as fraud detection. Trust erodes when AI's purpose is ambiguous or appears primarily to serve institutional efficiency—such as replacing human service or automating consequential decisions without meaningful explanation or recourse.

This experience gap highlights why both frameworks and regulation must move beyond abstract principles and directly address how AI systems function in real consumer contexts.

[67] Consumer Reports nationally representative AI in Financial Services Survey of 4,073 U.S. adults (September-October, 2025), https://article.images.consumerreports.org/image/upload/v1767035543/prod/content/dam/surveys/Consumer_Reports_AI_in_Financial_Services_2025.pdf.
[68] ibid.

Regulatory Overview

This section examines the most relevant laws, regulations, and proposals affecting the use of AI in consumer financial products and services (CFPS). Each subsection highlights a regulatory concept and compares how it appears in different jurisdictions. While a large number of legal examples are considered (see Appendix #1 for a comprehensive list), the purpose of this analysis is not to provide an exhaustive catalog of laws, but rather to identify ideas that could inform future frameworks while noting where approaches diverge or leave gaps. There are therefore instances where redundant laws or bills are not fully described. Similarly, there are instances where laws are tangentially relevant but insufficiently salient to describe in detail.

Colorado remains one of the most prominent state-level actors, having enacted a comprehensive AI framework in 2024[69] governing "high-risk" systems used in consequential decisions, including in financial services.[70] Although its effective date has been delayed to June 30, 2026,[71] the core statutory provisions—including impact assessments, discrimination safeguards, and consumer-disclosure obligations—remain intact. California has shifted from prescriptive frontier-model legislation (vetoed in 2024)[72] to a transparency-focused regime enacted in 2025 that requires large AI developers to publish governance frameworks and disclose catastrophic risk assessments.[73] New York similarly enacted its RAISE Act in late 2025, establishing governance and transparency requirements for certain large-scale model developers,[74] with implementation scheduled for 2027.[75] Other states, such as Texas and Utah, have taken narrower approaches emphasizing notice, disclosure, and other governance principles rather than comprehensive risk-based regimes.[76,77] Across the nation, many state legislatures are considering following these examples.[78]

---

[69] Colorado Senate Bill 24-205, APPROVED by Governor May 17, 2024, https://leg.colorado.gov/bill_files/47770/download: 26.
[70] SB 24-205 "(1) On and after February 1, 2026, a developer of a high-risk artificial intelligence system shall use reasonable care to protect consumers from any known or reasonably foreseeable risks of algorithmic discrimination arising from the intended and contracted uses of the high-risk artificial intelligence system," https://leg.colorado.gov/bill_files/47770/download: 6.
[71] "Overview. The bill delays implementation of Senate Bill 24-205 from February 1, 2026, to June 30, 2026," https://content.leg.colorado.gov/sites/default/files/documents/2025B/bills/fn/2025b_sb25b-004_f1.pdf: 1.
[72] https://www.gov.ca.gov/wp-content/uploads/2024/09/SB-1047-Veto-Message.pdf:1.
[73] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB53.
[74] https://www.nysenate.gov/legislation/bills/2025/A6453/amendment/A.
[75] https://assembly.state.ny.us/leg/?default_fld=&leg_video=&bn=A09449&term=&Text=Y#:~:text=%C2%A7%20%203.%20%20This%20act%20shall%20take%20effect%20%5Bon%20the%20ninetieth%20day%20after%20it%20shall%0A%20%20%20%2032%20%20have%20become%20a%20law%5D%20January%201%2C%202027.
[76] Texas, https://capitol.texas.gov/tlodocs/89R/analysis/html/HB00149S.htm.
[77] Utah, https://le.utah.gov/~2024/bills/static/SB0149.html.
[78] Gregory S. Dawson et al., "How different states are approaching AI," *Brookings Institution*, August 18, 2025, https://www.brookings.edu/articles/how-different-states-are-approaching-ai/.

At the federal level, comprehensive AI legislation remains stalled.[79] While proposals such as the AI Civil Rights Act[80] (introduced in late 2025) and Algorithmic Accountability Act (referred to committee in June 2025) would add bias testing, audits, and transparency requirements, none has been enacted.[81] A December 2025 Executive Order signaled federal opposition to certain state AI laws and called for a more uniform national framework,[82] but absent congressional action state-level statutes continue to operate. Internationally, regulatory standards remain more developed: the European Union's AI Act[83] and General Data Protection Regulation (GDPR) impose structured obligations on high-risk systems, and China's algorithmic recommendations laws[84] already impose structured, binding obligations on high-risk AI systems, including requirements related to risk classification, documentation, testing, monitoring, and accountability.

Because the U.S. CFPS framework is fragmented across statutes like the Equal Credit Opportunity Act (ECOA), Fair Credit Reporting Act (FCRA), Consumer Financial Protection Act (CFPA), the Gramm-Leach-Billey Act (GLBA), and a number of other laws and regulations, and relies heavily on sectoral enforcement by the Consumer Financial Protection Bureau, the Securities and Exchange Commission, and the Federal Trade Commission, significant regulatory gaps remain. At the time of this writing, it is difficult to know precisely which relevant regulators will enforce the law or how they would do so. This report therefore interprets enacted and proposed laws and regulations, guidance, active or revoked executive orders, and legal examples that are under judicial review or have been revoked or rescinded as potential sources of advocacy or inspiration. Together, these comparisons illustrate a complex, transitional regulatory landscape in which lawmakers, regulators, advocates, companies, and consumers must navigate a piecemeal and evolving AI CPFS marketplace.

---

[79] "Artificial Intelligence Legislation Tracker," *Brennan Center for Justice*, September 2025, https://www.brennancenter.org/our-work/research-reports/artificial-intelligence-legislation-tracker.
[80] AI Civil Rights Act of 2025 https://www.congress.gov/bill/119th-congress/house-bill/6356/text see "Actions" tab for introduction date (December 2, 2025).
[81] https://www.congress.gov/crs-product/R48555 "No federal legislation establishing broad regulatory authorities for the development or use of AI or prohibitions on AI has been enacted."
[82] https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/.
[83] https://artificialintelligenceact.eu/ai-act-explorer/.
[84] *Cyberspace Administration of China,* 2022, Internet Information Service Algorithmic Recommendation Management Provisions

Assessments & audit requirements

Entities may be required to conduct evaluations of AI systems and outcomes.

Legal examples:	Colorado Law - Colorado Privacy Act Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems (Proposed) Federal Law - Artificial Intelligence Civil Rights Act (Proposed) Federal Law - Algorithmic Accountability Act of 2025 (Proposed) Hawaii Law - Relating to Algorithmic Discrimination (Proposed) Illinois Law - Artificial Intelligence Safety and Security Protocol Act New York Department of Financial Services (DFS) Guidance - Use of Artificial Intelligence Systems [...] (Proposed) New York Law - NY Artificial Intelligence Act (Proposed) New York Law - An act to amend the banking law, in relation to the use of automated lending decision-making tools to [...]
Relevant axes:	Bias & Fairness · Accuracy

This regulatory concept requires organizations to perform structured risk assessments and, in some cases, independent audits of AI systems that affect consumers. In consumer financial products and services (CFPS), these assessments would be relevant for any decision engine; applications could therefore include fraud detection systems, product recommendation engines, targeting algorithms, credit decisioning systems, debt collection evaluation systems, P2P lending tools, and a range of other consumer-facing products and services. By surfacing and sharing potential risks and ongoing issues, this regulatory concept may in theory increase transparency and support identification of bias-related risks.

State laws provide distinct assessment frameworks.
Colorado's privacy laws require “data protection assessments” for certain profiling activities that present heightened risks of harm, with an emphasis on bias, fairness, and privacy.[85] These assessments must be retained for three years and made available to the Colorado attorney general upon request.[86] Colorado’s AI Act (SB 24-205), enacted in 2024 and effective June 30,

[85] Colorado Privacy Act (CPA) Implementing Rules, Colorado Attorney General (4 CCR 904-3), Rule 9: Profiling and Automated Decision Making, https://www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=11819&fileName=4%20CCR%2090 4-3: 38, part 8.
[86] Colorado Privacy Act (CPA) Implementing Rules, Colorado Attorney General (4 CCR 904-3), Rule 9: Profiling and Automated Decision Making, https://www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=11819&fileName=4%20CCR%2090 4-3: 42, item E.

2026, takes a broader approach, requiring risk assessments and impact assessment documentation of developers and deployers of "high-risk AI systems" that influence consequential decisions.[87] These reports must document the intended use, data sources, known or foreseeable risks of algorithmic discrimination, and mitigation measures. They must also include public plain-language summaries.[88]

Illinois' proposed law[89] creates requirements for developers to regularly publish and follow a "safety and security protocol," create and "conspicuously publish a risk assessment report," and annually hire a "reputable third-party auditor" to assess compliance with the protocol.[90] A proposed Hawaii law[91] would require annual audits of algorithmic decision engines that document discrimination and disparate impact risks, illegal actions, data used, data sources, methodology used to establish the algorithm, model performance, and decisions made by the model.

Federal proposals would expand assessment requirements across more systems.
The proposed federal Algorithmic Accountability Act[92] directs the Federal Trade Commission to create rules that require impact assessments for a wide range of "automated decision systems" (definitionally interchangeable with algorithmic decision engines and AI-like systems) and "augmented critical decision processes"; it remains under consideration in Congress.[93] These assessments must document the system's purpose, data governance, risk assessment, and results of testing. The proposed federal Artificial Intelligence Civil Rights Act[94] focuses more narrowly on "consequential decisions" that materially impact consumers, requiring assessments as part of a broader civil rights compliance program.

New York bills and guidance incorporate more specific obligations.
Certain proposed New York legislation has contemplated impact assessment requirements for

---

87 https://www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=11819&fileName=4%20CCR%2090 4-3: 39.

88 Colorado, Concerning Consumer Protections in Interactions With Artificial Intelligence Systems (SB 24-205, signed May 17, 2024).

89 Illinois HB 3506, Artificial Intelligence Safety and Security Protocol Act (proposed), https://www.ilga.gov/legislation/BillStatus?DocNum=3444&GAID=18&DocTypeID=SB&LegId=166392&se ssionid=114.

90 Illinois HB 3506, Artificial Intelligence Safety and Security Protocol Act (proposed), https://www.ilga.gov/legislation/BillStatus?DocNum=3444&GAID=18&DocTypeID=SB&LegId=166392&se ssionid=114.

91 Hawaii SB 59, Relating to Algorithmic Discrimination (proposed), https://www.capitol.hawaii.gov/session/archives/measure_indiv_Archives.aspx?billtype=SB&billnumber=5 9&year=2025.

92 U.S. Senate Bill 2164, Algorithmic Accountability Act of 2025 (proposed, introduced June 25, 2025). https://www.congress.gov/bill/119th-congress/senate-bill/2164/text.

93 U.S. Senate Bill 2164, Algorithmic Accountability Act of 2025 (proposed, introduced June 25, 2025). https://www.congress.gov/bill/119th-congress/senate-bill/2164/text.

94 U.S. Senate Bill 5152, 118th Congress, Artificial Intelligence Civil Rights Act (proposed). https://www.congress.gov/bill/119th-congress/house-bill/6356/text.

high-risk AI systems used in consumer transactions.[95] Proposed amendments to New York banking law would impose pre-deployment and annual audits for automated lending decision-making tools, with a focus on discriminatory outcomes and model accuracy.[96] Separately, the New York Department of Financial Services (DFS) insurance guidance states that insurers must conduct regular assessments of AI systems and external consumer data sources.[97] These assessments must monitor for discrimination and document vendor reliance.

Assessments can be a helpful first step, but self-assessment biases and consumer accessibility must be managed.
This regulatory concept generates reports that most consumers cannot understand. While impact assessment requirements may increase documentation of AI-related risks, accessibility and comparability remain limited. Many assessment reports are not standardized, not easily accessible to the public, or rely heavily on self-evaluation. The effectiveness of such regimes depends on regulatory capacity, technical expertise, and independent review mechanisms, all of which vary across jurisdictions. Centralized, searchable repositories of regulator-facing attestations (for example, NYDFS submissions), banking impact assessments (where required), and public high-risk AI statements could aid in comparison. Unfortunately, these practices may not alleviate the fundamental limitation of this idea: Self-assessments, including those already enacted or proposed by some jurisdictions, may have glaring gaps that are largely undetectable by regulators and consumers. The effectiveness of self-assessment regimes depends heavily on regulatory capacity and independent technical review, both of which vary across jurisdictions.

[95] New York Senate Bill S1169, New York Artificial Intelligence Act (proposed, 2025). The RAISE Act is enacted (developer governance). Broader "AI Act" proposals remain pending, https://www.nysenate.gov/legislation/bills/2025/S1962.
[96] New York Assembly Bill A00773, an act to amend the banking law in relation to automated lending decision-making tools (proposed), https://www.nysenate.gov/legislation/bills/2025/A773/amendment/C.
[97] New York Department of Financial Services, Circular Letter No. 2024-07: Use of Artificial Intelligence Systems and External Consumer Data and Information Sources in Insurance Underwriting and Pricing (July 11, 2024) https://www.dfs.ny.gov/industry-guidance/circular-letters/cl2024-07.

Bans on algorithmic discrimination

Entities may need to prevent discrimination across diverse applications.

Legal examples:	Consumer Financial Protection Bureau (CFPB) Circular - Adverse action notification requirements [...] Consumer Financial Protection Bureau Comment - Comment on Request for Information on Uses, Opportunities, and Risks of Artificial [...] Consumer Financial Protection Bureau Rule - Quality Control Standards for Automated Valuation Models Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems (Revoked) Executive Order - Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Proposed) Federal Law - Artificial Intelligence Civil Rights Act; Federal Law / Regulation - Equal Credit Opportunity Act (ECOA) / Reg B (Proposed) Georgia Law - Fair and Future Ready Housing Act (Proposed) Hawaii Law - Relating to Algorithmic Discrimination New York Department of Financial Services (DFS) Guidance - Use of Artificial Intelligence Systems [...] (Proposed) New York Law - New York Artificial Intelligence Consumer Protection Act (Proposed) Rhode Island Law - Artificial Intelligence Act
Relevant axes:	Bias & Fairness

A number of states, officials, and regulators are aware that AI may exacerbate or trigger discrimination in CFPS. These measures seek to prevent unlawful discrimination by prohibiting differential treatment or disparate impact based on protected characteristics in lending, housing finance, insurance, or other financial services. These laws, orders, and regulations could apply to credit decisioning, fraud detection, pricing algorithms, algorithmic targeting and advertising, and a very broad range of other algorithmic decision systems.

Existing protections are extended, and new ones are created.
The Equal Credit Opportunity Act (ECOA) and Regulation B apply regardless of whether lending decisions are made using automated or AI-driven systems, and CFPB guidance has reaffirmed that adverse action and nondiscrimination obligations extend to algorithmic decision-making contexts.[98] Seeking to extend these protections, several jurisdictions have created AI-specific

[98] CFPB, Circular 2022-03: Adverse Action Notification Requirements in Connection with Credit Decisions Based on Complex Algorithms, (May 26, 2022), 15 U.S.C. §§ 1691–1691f; 12 C.F.R. Part 1002; CFPB Circular 2022-03.

anti-discrimination legislation. Colorado's AI Act (SB 24-205), enacted in 2024 and effective June 30, 2026, requires deployers of "high-risk AI systems" to exercise "reasonable care" to prevent algorithmic discrimination and to provide notice and information regarding consequential decisions.[99] Certain proposed New York and Rhode Island legislation includes similar AI-specific anti-discrimination protections.[100]

More specific protections also seek to target AI discrimination risk.
The CFPB's final rule on Automated Valuation Models (AVMs) requires mortgage originators and secondary-market issuers to maintain quality-control standards designed to ensure compliance with nondiscrimination laws, including ECOA and the Fair Housing Act.[101] The rule broadly requires entities to set up "policies, practices, procedures, and control systems" that will prevent the AVMs from violating nondiscrimination laws (including ECOA and the Fair Housing Act). New York Department of Financial Services' more specific guidance requires insurers to assess and document that the use of external data and AI does not result in unfair discrimination.[102] Proposed legislation in Georgia seeks restrictions on discriminatory AI use in housing-related decisions.[103] Proposed Hawaii legislation (e.g., SB 59) has sought to prohibit algorithmic discrimination across specified sectors, including credit, insurance, education, employment, housing, or place of public accommodation. The October 2023 Executive Order on AI governance included directives encouraging federal agencies to address unlawful discrimination in AI systems; that order was rescinded in January 2025 and did not create binding statutory obligations.[104]

Modern risk vectors are increasingly incorporated into protected class definitions.
Certain newer laws have expanded their list of protected characteristics. (Older definitions have often included race, color, religion, national origin, sex, marital status, age, disability or receipt of public assistance, veteran status, and certain other more specific characteristics.) For example, Colorado's enacted law highlights "genetic information," "reproductive status," and limited English proficiency.[105] New York's proposed bill also considers "sexual orientation, gender identity, gender expression, pregnancy, pregnancy outcomes, and reproductive healthcare

[99] Colo. Rev. Stat. §§ 6-1-1703, 6-1-1704.
[100] Rhode Island Senate Bill S0627, Artificial Intelligence Act (proposed, 2025).
[101] CFPB, Quality Control Standards for Automated Valuation Models (Final Rule, June 2024), 12 C.F.R. Part 1002; 88 Fed. Reg. 84150 (December 2023).
[102] New York Department of Financial Services, Circular Letter No. 2024-07: Use of Artificial Intelligence Systems and External Consumer Data and Information Sources in Insurance Underwriting and Pricing (July 11, 2024), https://www.dfs.ny.gov/industry_guidance/circular_letters/cl2019_01.
[103] Georgia House Bill (2025-2026 session), Fair and Future Ready Housing Act (proposed), https://www.legis.ga.gov/legislation/71101.
[104] Executive Order 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (November 1, 2023, revoked by Executive Order 14148, January 20, 2025), https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.
[105] Colo. Rev. Stat. § 6-1-1701 et seq.

choices" to be relevant protected classes. The proposed federal AI Civil Rights Act[106] includes "income level (not including the ability to pay for a specific good or service being offered)," "biometric information," and "medical conditions."

Broad bans mean opportunities exist to support regulators and empower consumers. Given the breadth of these bans, enforcement capacity and monitoring mechanisms will determine whether statutory protections translate into meaningful consumer outcomes. In particular, collecting, sharing, and potentially publishing data and provable examples of protected class discrimination could help regulators efficiently improve the market and support consumers in comparing CFPS providers.

[106] Proposed AI Civil Rights Act https://www.congress.gov/bill/119th-congress/house-bill/6356 pending as of this writing.

Consumer protection laws still apply to AI

Statements from around the country seek to reaffirm enforcement powers.

Legal examples:	Joint statement from CFPB, FTC, EEOC, and DOJ - Joint Statement On Enforcement Efforts Against Discrimination And Bias In Automated Systems Statements, Guidance, & Advisories from Attorneys General - Including California, Massachusetts, and New Jersey
Relevant axes:	Accountability & Transparency

Statements from federal regulators and executives continue to emphasize that the use or stated use of AI and algorithmic decision-making does not diminish or displace existing consumer protection obligations, and that these laws will continue to be enforced. In 2023, the CFPB, FTC, DOJ, and EEOC issued a joint statement affirming that longstanding anti-discrimination and consumer protection statutes—including ECOA, FCRA, and the FTC Act—apply equally to automated and algorithmic systems.[107] That position remains legally operative: the introduction of AI into financial decision-making does not create exemptions from civil rights, fair lending, or unfair and deceptive practices laws. This is a broad regulatory concept, supporting most axes by solidifying accountability.

While a 2023 federal Executive Order directed agencies to prioritize AI-related civil rights enforcement, it was rescinded in January 2025.[108] A subsequent Executive Order in December 2025 articulated a federal preference for a more uniform national AI policy framework. However, neither executive action alters the applicability of existing statutory obligations governing lending, credit reporting, privacy, or unfair practices. Enforcement authority under ECOA, FCRA, CFPA,[109] and related statutes remains intact.

Attorneys general and federal regulators have issued individual and joint statements. Similarly, multiple state attorneys general—including those in California, Massachusetts, and New Jersey—have issued public statements reinforcing that algorithmic systems must comply with existing consumer protection and anti-discrimination statutes.[110]

---

[107] https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

[108] Exec. Order No. 14110, 88 Fed. Reg. 75191 (Oct. 30, 2023) - (rescinded January 2025).

[109] 12 U.S.C. §§ 5531, 5536, 5565(c).

[110] See, e.g., Office of the New Jersey Attorney General, Statement on Algorithmic Bias and Consumer Protection New Jersey Attorney General, DCR Guidance on Algorithmic Discrimination (January 8, 2025); Office of the California Attorney General, Guidance on AI and Consumer Protection Enforcement (California Attorney General, Legal Advisory: Application of Existing California Laws to Artificial Intelligence; 2023–2024); Massachusetts Attorney General Advisory on the Application of the

These statements vary in emphasis. New Jersey's guidance highlights discrimination risks specifically,[111] while others reference broader concerns including civil rights compliance, data privacy obligations, and the potential for misleading or false information generated by automated systems. Across the board, they underscore a consistent position: The use of AI in consumer financial products and services does not displace existing statutory obligations.

The practical implication is that AI systems are governed by a layered and fragmented regulatory structure rather than one that is AI-specific and unified. Existing consumer protection laws provide baseline guardrails, but questions remain regarding interpretability, monitoring capacity, and consistent enforcement across jurisdictions.

Commonwealth's Consumer Protection, Civil Rights, and Data Privacy Laws to Artificial Intelligence (Massachusetts Attorney General, AI Advisory April 16, 2024).

[111] https://www.nj.gov/oag/newsreleases25/2025-0108_DCR-Guidance-on-Algorithmic-Discrimination.pdf.

Cybersecurity and data protection

Entities may need to prevent discrimination across diverse applications.

Legal examples:	Connecticut Law - Connecticut Data Privacy Act Delaware Law - Personal Data Privacy Act Federal Law - Federal Deposit Insurance Act; Federal Law - Gramm-Leach-Bliley Act; Federal Trade Commission Regulation - Standards For Safeguarding Customer Information ("Safeguards Rule") Multi-Agency Regulation (Office of the Comptroller of the Currency [OCC], Board of Governors of the Federal Reserve, Federal Deposit Insurance Corporation [FDIC], Office of Thrift Supervision [OTS]) - Interagency Guidelines Establishing Information Security Standards; Multi-Agency Regulation (Office of the Comptroller of the Currency [OCC], Board of Governors of the Federal Reserve System, and the Federal Deposit Insurance Corporation [FDIC]): Computer-Security Incident Notification; National Institute of Standards and Technology (NIST) Guidance - Cybersecurity Framework; New York Department of Financial Services (DFS) State-Level Regulation - Cybersecurity Requirements For Financial Services Companies
Relevant axes:	Data Security · Privacy ·

This set of regulatory concepts requires financial service providers—and, in some cases, the core AI service providers supporting them—to identify, report, and remediate security risks to consumer data and related systems. In an AI CFPS context, this could relate specifically to safeguarding or minimizing training data for fraud or credit decisioning, protecting data from interactions with chat or other interfaces, and ensuring that the infrastructure that serves models is secure. Strong security controls reduce the risk of prompt injection, data theft, or data leakage, thereby strengthening data protection and related privacy safeguards.

Federal regulations typically require reporting and security controls.
At the federal level, the FTC's Safeguards Rule requires certain non-bank financial institutions to implement a written information security program and, as of its recent amendments, to report within 30 days certain security events that affect more than 500 consumers.[112] Covered entities must designate a qualified individual to oversee the information security program and implement

[112] FTC Standards for Safeguarding Customer Information ("Safeguards Rule"), 16 C.F.R. Part 314 (amended 2021, effective June 2023).

administrative, technical, and physical safeguards, including access controls, encryption, continuous monitoring, and vendor oversight.[113]

Banks, by contrast, are subject to tighter timelines and broader reporting requirements. The joint federal banking agency incident-notification rule requires banking organizations to notify their primary federal regulator as soon as possible and no later than 36 hours after determining that a "computer-security incident" meeting defined thresholds has occurred.[114] This measure includes breaches, major service disruptions, or any other security concerns about sensitive customer data. Setting somewhat stricter standards for banks, the Interagency Guidelines Establishing Information Security Standards require banking organizations to implement administrative, technical, and physical safeguards to protect customer information, including oversight of service providers. These measures must be documented and regularly reviewed by management.[115]

NIST guidance provides a lens into regulator expectations.
The NIST Cybersecurity Framework organizes cybersecurity practices around functions including: Identify, Protect, Detect, Respond, and Recover, and the added "Govern" function.[116] Regulators often reference the NIST Framework, meaning that adoption can help CFPS providers demonstrate adherence to federal and state data protection expectations.[117] The specifics in the framework are extensive and could guide entities to best practices on how to secure training data, model parameters, and supporting infrastructure, as well as how to prepare for AI-specific threats such as data poisoning or adversarial manipulation.

States extend and sharpen these obligations.
NYDFS's cybersecurity regulation[118] (which applies to banks, insurers, and other CFPS) requires covered entities to conduct periodic risk assessments, designate a chief information security officer, oversee third-party service providers, and report certain cybersecurity incidents to the department within specified timeframes. Somewhat less comprehensive, Connecticut's privacy law requires controllers to implement reasonable administrative, technical, and physical data security practices appropriate to the volume and nature of the personal data processed.[119] Delaware law similarly requires "reasonable" security practices, including staff training, secure

[113] 16 C.F.R. § 314.4.
[114] Interagency Notification Rule: Computer-Security Incident Notification Requirements for Banking Organizations and Their Bank Service Providers, 12 C.F.R. Parts 53, 225, 304, and 748 (effective April 1, 2022); 12 C.F.R. § 53.3 (OCC); 12 C.F.R. § 225.301 (Fed); 12 C.F.R. § 304.22 (FDIC).
[115] Appendix B to 12 C.F.R. Part 30 (OCC); 12 C.F.R. Part 208, App. D-2 (Fed); 12 C.F.R. Part 364, App. B (FDIC) Interagency Guidelines Establishing Information Security Standards (12 C.F.R. Part 30, App. B for OCC-supervised banks; 12 C.F.R. Part 208, App. D-2 for state member banks).
[116] National Institute of Standards and Technology, Cybersecurity Framework (CSF) 2.0 (released February 2024), https://www.nist.gov/cyberframework.
[117] National Institute of Standards and Technology, Cybersecurity Framework (CSF) 2.0 (released February 2024), https://www.nist.gov/cyberframework.
[118] 23 N.Y.C.R.R. Part 500 NYDFS Cybersecurity Requirements for Financial Services Companies, 23 N.Y.C.R.R. Part 500 (amended October 2023).
[119] Connecticut Data Privacy Act (Public Act 22-15, eff. July 1, 2023). Conn. Gen. Stat. § 42-521.

network configurations, and controlled data storage.[120] More specific to insurance, Ohio's data security law mandates risk evaluations, board reporting, vendor due diligence, and notification sent to regulators within three business days of qualifying cybersecurity incidents.[121]

While cybersecurity laws cover many elements of CFPS, private sector-related AI-specific cybersecurity regulation is nascent, as jurisdictions have only begun to examine the issue. Reported data breaches and incident notifications are not systematically synthesized in ways that allow consumers to compare providers' security performance, creating informational asymmetries that may persist despite reporting requirements. While public reporting enhances transparency, variability in disclosure formats and aggregation practices limit comparability across institutions. As AI systems increasingly rely on complex data pipelines and model infrastructure, questions remain regarding how consistently training data, production systems, and model governance practices are evaluated under existing cybersecurity regimes.

[120] Delaware Personal Data Privacy Act (signed September 11, 2023), 6 Del. C. § 1205C.
[121] Ohio Rev. Code § 3965.04 Ohio Data Protection Act (R.C. §§ 1354.01–1354.05) and Ohio Data Security Law for Insurance Companies (R.C. Chapter 3965).

Deceptive Design prohibitions

Manipulative choice architecture may be restricted in certain AI contexts.

Legal examples:	California Law - California Consumer Privacy Act, Update; Colorado Law - Colorado Privacy Act; Connecticut Law - Connecticut Data Privacy Act; Consumer Financial Protection Bureau (CFPB) Statement - Policy Statement on Abusive Acts or Practices; Consumer Financial Protection Bureau (CFPB) Statement - Request for Information on Uses, Opportunities, and Risks of Artificial [...]; Federal Law - Consumer Financial Protection Act; (Proposed) Federal Law - Deceptive Experiences To Online Users Reduction (DETOUR) Act Texas Law - Responsible Artificial Intelligence Governance Act
Relevant axes:	AI Safety & Human Centeredness Consumer Agency & Empowerment Accountability & Transparency · Bias & Fairness · Privacy

This regulatory concept focuses on prohibiting the use of deceptive, harmful, or manipulative design patterns and choice architecture (also known as “dark patterns”) in consumer interactions, particularly when they impact consent, comprehension, or the ability to exercise rights. Given the breadth of potentially deceptive design in present consumer technology products, these regulations could impact AI-driven customer acquisition flows, marketing content, contracts and cancellations, natural language processing-based chatbots explaining or establishing agreements, checkout processes, product add-on offers, targeted enrollment nudges, and a range of other touchpoints.[122] These measures aim to protect consumer autonomy and informed choice. As such, they intersect with broader concerns regarding consumer agency, transparency, and human oversight in AI-mediated interactions. This can impact fairness and privacy downstream as well.

Jurisdictions vary in the scope and framing of the dark patterns prohibitions. Texas law, including the Texas Data Privacy and Security Act, prohibits the use of deceptive design to obtain consumer consent, requiring plain language, clear, and conspicuous notices.[123] California’s and Colorado’s privacy rules both specify that consent obtained through dark patterns is invalid.[124,125] Connecticut’s privacy laws incorporate a definition of “dark pattern”

[122] Federal Trade Commission. Bringing Dark Patterns to Light. FTC Staff Report, 2022.
[123] Texas Data Privacy and Security Act (HB 4), 88th Legislature, 2023, § 541.055 Texas HB 1709, Texas Artificial Intelligence in Critical Infrastructure Act (effective. September 1, 2023).
[124] California Code of Regulations. Title 11, § 7004, California Privacy Protection Agency, Enforcement Advisory No. 2024-02, 2024, California Consumer Privacy Act (as amended by CPRA, Civil Code §1798.100 et seq.), Colorado Privacy Act (Colo. Rev. Stat. §6-1-1301 et seq.).
[125] Colorado Department of Law, Colorado Privacy Act Rules, 4 CCR 904-3, Rule 7.

aligned with the previous two states mentioned, but extend the prohibition to all covered consent interactions.[126] At the federal level, the CFPB has identified manipulative "negative option" choice architecture (where inaction or continuing is treated as consent) as potentially unlawful under existing consumer protection statutes in financial services contexts.[127] The CFPB's statement on abusiveness also frames manipulative choice architecture as abusive.

Proposed federal legislation on manipulative design.
The proposed federal DETOUR Act would have restricted certain manipulative design practices by prohibiting large online operators from "obscuring, subverting, or impairing user autonomy, decision making, or choice" using an interface. The bill additionally contemplated stricter standards for obtaining consent, particularly for minors. However, the DETOUR Act has not advanced, and no comprehensive federal statute specifically targeting deceptive design patterns has been enacted.[128] As a result, enforcement continues to rely primarily on existing unfair and deceptive practices authorities and state privacy regimes.

Deceptive practices bans are a valuable tool, but enforcement and comparison require support.
Despite these prohibitions, enforcement challenges and definitional ambiguity may limit the deterrent effect of deceptive practices bans in AI-mediated financial services. Variability in interpretation and application across jurisdictions further complicates comparative assessment. As AI-driven personalization becomes more prevalent, the boundary between persuasive design and unlawful manipulation may become increasingly difficult to delineate under existing statutory frameworks.

[126] Connecticut General Assembly. Public Act No. 22-15: An Act Concerning Personal Data Privacy and Online Monitoring, 2022.
[127] Comment Letter on Request for Information on Uses, Opportunities, and Risks of Artificial Intelligence in the Financial Services Sector, Consumer Financial Protection Bureau, August 12, 2024. CFPB, Statement of Policy Regarding Abusive Acts or Practices (April 3, 2023), and CFPB comments on Request for Information on Uses, Opportunities, and Risks of Artificial Intelligence in the Financial Services Sector (CFPB-2023-0006).
[128] Deceptive Experiences To Online Users Reduction Act (DETOUR Act, S. 2708, 118th Congress, proposed, not enacted).

Data deletion rights

Consumers can request deletion of certain kinds of data in varying contexts.

Legal examples:	California Law - California Consumer Privacy Act; California Law - Data broker registration: accessible deletion mechanism; California Privacy Protection Agency Regulation - California Consumer Privacy Act Regulations. Colorado Attorney General Regulation - Colorado Privacy Act Rules; Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems; Federal law – Fair Credit Reporting Act; (Proposed) Federal Law - National DELETE Act; Virginia law – Virginia Consumer Data Protection Act
Relevant axes:	Accuracy · Data Security · Consumer Agency & Empowerment

This regulatory concept provides consumers the right to request deletion of their personal data held by covered entities, either to address privacy concerns or to correct or remove data that is either inaccurate or no longer necessary. In the context of AI-mediated consumer financial products and services, deletion rights may encompass data collected directly, inferred or algorithmically derived, extrapolated algorithmically, or purchased from third-party data brokers. These rights can apply to information used in machine learning-based underwriting, algorithmic pricing, targeted advertising, and other AI-influenced financial services. The right to data deletion serves as a mechanism of consumer control over personal information and may support data minimization practices that reduce the amount of retained data.

Specific laws and regulations vary in data category and deletion mechanism.
While many jurisdictions have passed or proposed deletion rights, they diverge significantly in terms of covered entities, data categories, and mechanisms for execution. California's privacy framework provides the strongest deletion rights, as it mandates that data brokers register with the state and must honor deletion requests.[129] California's framework contains certain carve-outs for entities already subject to sector-specific privacy regimes such as the Gramm-Leach-Bliley Act (GLBA). Even so, deletion rights and data broker registration obligations can affect third-party data holders and intermediaries that contribute to financial profiling, targeted advertising, or alternative credit modeling. Virginia's Consumer Data Protection Act similarly grants consumers the right to delete personal data that a controller has

[129] California CCPA/CPRA, Civil Code § 1798.105 (right to deletion) and California Delete Act (SB 362, eff. January 1, 2024), requiring data brokers to register and honor deletion requests.

collected about them, including inferred data or data derived from other information, subject to certain exceptions.[130]

At the federal level, the Fair Credit Reporting Act (FCRA) gives consumers the right to correct and dispute inaccurate information in consumer reports.[131] While this is not a general "deletion right," it functions as a deletion or amendment mechanism for covered credit reporting data. The Consumer Financial Protection Bureau (CFPB) continues to affirm that FCRA obligations apply regardless of whether AI or automated processes are used in credit reporting or credit decisioning. The proposed National DELETE Act[132] would extend this significantly by empowering the FTC to create a centralized system for consumers to delete personal data held by data brokers. As of early 2026, this bill remains under consideration.

Deletion triggers vary by regime / jurisdiction.
California's obligations apply broadly upon request from consumers, Virginia's law requires an affirmative request, and FCRA applies only to documented inaccuracies. Furthermore, while some laws are operationalized through centralized tools (e.g., California's data broker portal), others place the burden on consumers to individually contact companies.

Data deletion is a meaningful tool, but inefficiency and consumer burden create gaps.
Although deletion rights offer a mechanism for consumer control over personal information, practical barriers remain. Many jurisdictions require affirmative consumer action and differing documentation, often far more investigation and labor than the average consumer can reasonably provide. Additionally, the application of these requirements to only certain kinds of personal data means consumers remain under threat from hacks, leaks, and privacy invasion. Because most deletion regimes require affirmative consumer action and vary in scope, practical accessibility remains uneven across jurisdictions.

[130] Virginia CDPA (Va. Code Ann. §59.1-577). Virginia Consumer Data Protection Act (Va. Code § 59.1-571 et seq., effective January 1, 2023).
[131] Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681i (procedure for disputed accuracy).
[132] National DELETE Act (bill text / introduced April 2, 2025).

Direct disclosure of AI use

Entities may be required to inform consumers of AI interactions and decisions.

Legal examples:	(Proposed) Federal Law - Artificial Intelligence Civil Rights Act; (Revoked) Executive Order - Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence; California Law - SB 1001: Bots: disclosure. California Law - CA AI Transparency Act Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems; (Proposed) New York Law - NY Artificial Intelligence Act Utah Law - Artificial Intelligence Policy Act; Utah Law - Artificial Intelligence Consumer Protection Amendments
Relevant axes:	Accountability & Transparency Consumer Agency & Empowerment

This regulatory concept centers around requiring entities to disclose to consumers when consumers are interacting with or subject to decisions made by AI. This idea would generally be most relevant to developers and deployers of AI-driven financial advisors or brokers, algorithmic content on comparison-shopping sites, credit and lending decisions, customer service chatbots, and other AI touchpoints or decision engines. These laws, orders, and regulations seek to increase transparency into the use of AI in consumer interactions and may support consumers' ability to make informed comparisons among services.

Legal examples diverge with regard to the specific AI function to which they apply. One high-level articulation of AI disclosure principles appeared in President Biden's 2023 Executive Order directing federal agencies to advance transparency to consumers when they are interacting with or using AI.[133] That Executive Order was rescinded in January 2025 and did not itself create binding disclosure obligations.[134]

Other legal approaches are more narrowly tailored and vary significantly in scope. Utah's Artificial Intelligence Policy Act (2024) requires disclosure when a consumer interacts with generative AI in certain contexts, unless it would be obvious to a reasonable person that the content is AI-generated.[135] Colorado's AI Act, on the other hand, links disclosure obligations

[133] Executive Order 14110 (Oct. 30, 2023), rescinded January 2025. Executive Order 14110 (revoked Jan. 20, 2025), Section 8 (consumer protection).

[134] https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

[135] Utah Artificial Intelligence Policy Act (2024), Utah Code §13-71-201 et seq; Utah Artificial Intelligence Consumer Protection Amendments (SB 226, 2025 session).

more broadly to "high-risk systems," "high-risk artificial intelligence systems," or "high-risk AI systems [used] for a consequential decision."[136] In New York, certain proposed measures have contemplated advance notice requirements when high-risk AI systems are used in consequential decision-making, though such provisions remain under legislative consideration.[137] California's laws apply to the use of "bots to communicate" to "incentivize a purchase or sale" (2019) and to AI-generated content (2024), respectively.[138]

Requirements also differ with regard to logistical specifics of the disclosure. Utah's laws require disclosure only if an individual asks whether they are interacting with artificial intelligence. While other disclosures are either meant to be timely or are ambiguous in timeline, New York's proposed law requires a disclosure at least five business days prior to the use of "high-risk" AI for a "consequential" decision.

This idea boosts transparency and agency, but is only a first step towards impact. While disclosure requirements may enhance transparency, disclosure alone may not make a substantial difference in the power balance of consumer financial markets. Furthermore, the behavioral impact of such disclosures remains uncertain, particularly in markets with limited competition or meaningful opt-out options.

[136] Colorado SB 24-205 (2024). §6-1-1703 et seq.
[137] Responsible AI Safety and Education Act, N.Y. Gen. Bus. L. § 1421, 2025.
[138] California Bot Disclosure Law (Cal. Bus. & Prof. Code §17941).

Developers-to-deployers disclosure requirements

AI creators may be required to provide reports for those who launch their models.

Legal examples:	Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems; (Proposed) Federal Law - Artificial Intelligence Civil Rights Act; (Proposed) Federal Law - Algorithmic Accountability Act of 2025; (Proposed) New York Law - An act to amend the banking law, in relation to the use of automated lending decision-making tools to [...] (Proposed) New York Law - NY Artificial Intelligence Act (Proposed) New York Law - Artificial Intelligence Consumer Protection Act (AICPA); Texas Law - Responsible Artificial Intelligence Governance Act; Utah Law - Artificial Intelligence Policy Act
Relevant axes:	Accountability & Transparency · Data Security · Consumer Agency & Empowerment · Resilience

This set of regulatory concepts requires developers to disclose descriptions of AI systems, risks, limitations, and/or intended use to deployers. Such disclosures can establish clearer accountability structures across the AI supply chain, potentially improving auditability, monitoring, and allocation of responsibility when harms occur.

Legal examples provide differing expectations for disclosure content.
Colorado's AI Act (effective June 30, 2026) obligates developers of high-risk AI systems to provide deployers with documentation regarding intended use, known limitations, and how they manage algorithmic discrimination risk, and to notify both the Colorado attorney general and effective deployers if such risks are identified.[139] New York's Artificial Intelligence Act, enacted in December 2025, would similarly require certain large-scale developers of high-risk systems to publish governance and risk-assessment frameworks. While not structured explicitly as a developer-to-deployer documentation package, it increases transparency regarding developer safety practices. In contrast, Texas[140] and Utah laws[141] provide certain governance or disclosure requirements that would be relevant for developers, but largely stop short of requiring a developer-to-deployer disclosure package in the CFPS context.

Some disclosure requirements are not directly developer-to-deployers, but are still relevant.
At the federal level, the proposed Algorithmic Accountability Act (pending as of February 2026) would direct the FTC to require impact assessments that summarize the system's purpose, data

[139] Colorado AI Act (SB 24-205) – Colo. Rev. Stat. §6-1-1701 et seq.
[140] https://capitol.texas.gov/tlodocs/89R/analysis/html/HB00149S.htm.
[141] Utah Code §13-71-101 et seq.

use, risks, and mitigation measures.[142] The proposed Artificial Intelligence Civil Rights Act would also require FTC rulemaking to ensure notice, disclosure, opt-out rights, and access to human review for consumers affected by consequential algorithmic decisions.[143] The New York proposal to amend state banking law targets automated lending tools;[144] while this proposal primarily imposes requirements on lenders themselves, lender compliance may require documentation from developers that the state may mandate or lenders may request.

Jurisdictions vary in definitions (and titles) for deployers and developers. Broadly, developers are considered the entities that create AI models or algorithms, and deployers are considered those that build systems to apply these technologies to consumers or outside parties[145] While these distinctions can seem subtle, they may substantially impact the range of AI CFPS to whom these measures apply. Colorado law establishes developers as entities that create or substantially modify high-risk AI systems, while deployers are those who use such systems in consequential decisions.[146] The proposed federal laws take a similar approach, framing developers as algorithm creators and deployers as the entities operationalizing the technology.[147]

Texas law defines a developer as a person who creates an AI system that is sold, leased, or otherwise provided, while a deployer "deploys artificial intelligence systems for use;"[148] the law provides limited specificity regarding operational thresholds. New York's proposed AI Act is more specific, defining a developer as an entity who "designs, codes, [...] produces," or substantially changes an AI system and a deployer as an entity who "offers or uses an AI system for commerce" or "for use by the general public."[149]

The proposed federal AI Civil Rights[150] and New York AICPA bills[151] take a nearly identical approach. Relatedly, proposed federal legislation such as the Algorithmic Accountability Act[152] would establish obligations for "covered entities" based on size and use of automated decision systems, but does not explicitly define developer and deployer roles in a manner consistent with the aforementioned state-level AI laws. Instead, the bill focuses on requiring impact assessments and documentation of automated decision systems used in consequential

---

[142] https://www.congress.gov/bill/119th-congress/house-bill/6266/text.
[143] https://www.congress.gov/bill/119th-congress/house-bill/6356/text.
[144] https://www.nysenate.gov/legislation/bills/2025/A773/amendment/C.
[145] "AI Developers and Deployers: An Important Distinction," *Business Software Alliance*, https://www.bsa.org/files/policy-filings/03162023aidevdep.pdf.
[146] https://leg.colorado.gov/bill_files/47770/download: 3.
[147] https://www.congress.gov/bill/119th-congress/house-bill/6356/text#toc-H4025353877DF48FFB54C537B3F21B469.
[148] https://statutes.capitol.texas.gov/?tab=1&code=BC&chapter=BC.552&artSec=552.102.
[149] https://legislation.nysenate.gov/pdf/bills/2025/S1169A: 4.
[150] https://www.congress.gov/bill/118th-congress/senate-bill/5152/text.
[151] https://legislation.nysenate.gov/pdf/bills/2025/A768: 2.
[152] United States Congress, Algorithmic Accountability Act of 2023, H.R. 6580 / S. 3572, 118th Cong., 2023–2024.

contexts. Finally, Utah law provides relatively limited formal distinction between developer and deployer roles. Earlier statutory language referred broadly to persons who "use, prompt, or otherwise cause" generative AI systems to interact with individuals, but this provision was repealed in 2025.[153] Current Utah law focuses more narrowly on disclosure obligations in generative AI interactions rather than supply chain-role definitions.

Comprehensive disclosures boost transparency, but coverage and analysis gaps remain. Structured disclosure measures may clarify accountability across the AI supply chain and assist regulators in identifying the sources of system risks. However, implementations of this regulatory concept that are not simultaneously expansive and specific, and are not accompanied by other investigatory or regulatory measures, may leave substantial gaps while creating the assumption that AI development is being sufficiently monitored.

While these disclosure requirements may improve accountability along the AI supply chain, differences in definitions, scope, and enforcement create uneven transparency. Where disclosures are incomplete or not standardized, deployers and consumers may struggle to assess risk exposure. This fragmentation underscores variability across jurisdictions in AI-mediated financial services.

[153] Utah Legislature, Utah Code § 13-2-12 (2024) (repealed May 2025).

Disclosure of discrimination risks and instances

Entities may be required to publish observed discrimination and bias risks.

Legal examples:	Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems; Federal Law - Equal Credit Opportunity Act (ECOA); Federal Law - Home Mortgage Disclosure Act (HMDA); New York Department of Financial Services (DFS) Guidance - Use of Artificial Intelligence Systems [...]
Relevant axes:	Accountability & Transparency · Bias & Fairness · Quality of Redress Mechanisms

This regulatory concept centers around requiring entities to share or publish instances of discrimination in their AI systems, as well as their analysis of potential discrimination risk. In consumer financial products and services (CFPS), this could include credit underwriting, pricing, fraud models, advertising, and a range of other CFPS that rely on AI decision engines. Expanded disclosure requirements may increase transparency, subsequently aid in the fight against discrimination, and therefore boost downstream redress attempts.

Colorado leads this issue with regards to AI-specific protections.
Colorado's AI Act (SB 24-205), enacted in 2024 and effective on June 30, 2026, requires developers and deployers of "high-risk AI systems" to document and disclose how they manage the foreseeable risks of algorithmic discrimination.[154] The statute also requires notification to the Colorado attorney general when developers or deployers identify known or reasonably foreseeable risks of algorithmic discrimination arising from system use.[155] These disclosures could provide regulators and consumers with additional information relevant to identifying, detecting and contesting discriminatory outcomes.

Other laws and guidance apply in a secondary or more specific capacity.
The Equal Credit Opportunity Act (ECOA) requires creditors to provide applicants with specific reasons for an adverse action, including credit denial.[156] ECOA obligations apply regardless of whether decisions are made using automated or AI-driven systems. Prior CFPB guidance, including Circular 2022-03, clarified that adverse action notice requirements extend to algorithmic decision-making contexts. Relatedly, the Home Mortgage Disclosure Act (HMDA) requires the disclosure of lending data to facilitate monitoring for potential discriminatory lending

[154] Colo. Rev. Stat. § 6-1-1704 (developer documentation requirements); Colo. Rev. Stat. § 6-1-1706 (deployer duty).
[155] Colo. Rev. Stat. § 6-1-1704(4): 10 (developer notification requirement).
[156] Equal Credit Opportunity Act (ECOA), 15 U.S.C. §1691 et seq. (adverse action notice requirement); 12 C.F.R. §1002.9 (Regulation B - adverse action notice content).

patterns.[157] Finally, the NYDFS Insurance circular states that insurers must provide evidence and analysis to assess and document that the use of external data and AI does not result in unfair discrimination.[158]

Gaps exist in discrimination disclosure adoption, analysis, and consumer accessibility. While Colorado's documentation and notification requirements create a formal record of foreseeable algorithmic discrimination risks, no centralized public repository aggregates or standardizes these disclosures for comparative analysis. Variability in reporting formats, jurisdictional scope, and enforcement mechanisms may limit cross-state or cross-institution comparison.

[157] 12 U.S.C. § 2801 et seq. 12 C.F.R. Part 1003.
[158] NYDFS Insurance Circular Letter No. 1 (2019) (Use of External Consumer Data and Information Sources in Underwriting).

Environmental impact

The environmental impacts of AI may lead to reporting and further requirements.

Legal examples:	(Proposed) Federal Law - Artificial Intelligence Environmental Impacts Act; (Proposed) California Law - Water Resources: Data Centers; (Proposed) Connecticut Law - An Act Requiring A Study Concerning Energy Efficiency Requirements For Artificial Intelligence Data Centers; (Proposed) Minnesota Law - An Act Relating To Environment; Requiring Study Of Environmental Impacts Of Artificial Intelligence; Requiring [...]
Relevant axes:	Ecological Footprint

Concerns regarding the water and energy consumption associated with AI infrastructure have prompted regulatory and legislative proposals focused primarily on measurement, study, and reporting of the environmental impacts. Although these measures are not specifically related to consumer finance, increased infrastructure costs may indirectly affect consumer financial markets if energy and operational expenses are passed through to service providers and, ultimately, consumers. Because AI-driven financial products rely on large-scale computing infrastructure, regulatory approaches targeting data center efficiency or resource use may have downstream implications for the cost structure of AI-enabled consumer financial services.

Federal and state proposals differ in scope and enforcement.
Proposed federal legislation[159] has contemplated directing the National Institute of Standards and Technology (NIST) to develop standard methodologies for measuring and reporting AI-related energy and resource consumption. As of February 2026, such proposals remain under consideration and have not been enacted.[160] Draft legislation in Connecticut has proposed quarterly reporting of AI data centers' water and electricity consumption and contemplated granting state authorities the power to set efficiency standards. As of this writing, this proposal had not advanced, reflecting the early and still-developing nature of environmental regulation in this area.[161] Proposed legislation in Minnesota has similarly considered reporting requirements for AI-related energy consumption and also includes the analysis of potential disproportionate environmental impacts to certain communities.[162] More tangentially, California has considered legislation requiring data centers to disclose water use in certain situations[163] and that imposes efficiency requirements at the local level that could impact AI service providers [164] and subsequently CFPS. While these measures are not specific to financial services, they

[159] https://www.congress.gov/bill/118th-congress/senate-bill/3732/text.
[160] https://www.congress.gov/bill/118th-congress/senate-bill/3732/text.
[161] https://legiscan.com/CT/bill/SB01292/2025.
[162] https://www.revisor.mn.gov/bills/94/2025/0/SF/1117/versions/latest./.
[163] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260AB2619.
[164] https://alcl.assembly.ca.gov/system/files/2025-04/ab-93-papan.pdf.

may indirectly affect AI service providers that supply models or infrastructure used by consumer financial institutions.

Gaps remain in the transparency and comparability of AI-related environmental impacts. While these measures may begin to create downstream transparency around AI's environmental footprint, the issues of energy consumption and water scarcity are already concerns.[165] Publicly available disclosures regarding the energy and water consumption required by AI infrastructure are not standardized or easily comparable across providers.[166] As energy-intensive AI infrastructure expands, questions arise regarding how environmental costs are allocated[167] and whether such costs ultimately affect the affordability of consumer financial services.

[165] Marie Grimm, Nell Green Nylen, and Michael Kiparsky, "Regulating Data Center Water Use in California," *Center for Law, Energy & the Environment, UC Berkeley School of Law*, February 2026, https://www.law.berkeley.edu/data-center-water-use.
[166] Mitul Jhaveri and Vijaykumar Palat, "Measuring and Standardizing AI's Energy and Environmental Footprint to Accurately Access Impacts," *Federation of American Scientists*, June 27, 2025, https://fas.org/publication/measuring-and-standardizing-ais-energy-footprint/.
[167] Mahmut Kandemir, "Why AI uses so much energy — and what we can do about it," *Penn State Institutes of Energy and the Environment*, April 8, 2025, https://iee.psu.edu/news/blog/why-ai-uses-so-much-energy-and-what-we-can-do-about-it.

Explainability requirements

Entities may be required to justify their algorithmic decisions.

Legal examples:	(Proposed) Federal Law - Artificial Intelligence Civil Rights Act; (Revoked) Executive Order - Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence; Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems; (Proposed) New York Law - New York Artificial Intelligence Consumer Protection Act
Relevant axes:	Accountability & Transparency · AI Safety & Human Centeredness · Bias & Fairness · Consumer Agency & Empowerment

These requirements typically focus on ensuring that AI developers or deployers provide understandable explanations of how an AI system contributed to or made a decision. In the context of AI-driven CFPS, this idea may be applicable to credit decisioning, fraud detection and freezing of accounts, financial product recommendations, and AI customer support systems that automate or influence financial decisions. Explainability requirements are intended to provide consumers with greater transparency into decisions that affect them. Such transparency may assist consumers in assessing, contesting, or seeking correction of adverse or potentially biased AI-driven financial outcomes.

Enacted and proposed laws emphasize explainability and differ on execution.
Colorado's AI Act, SB 24-205, establishes explainability-related obligations for "high-risk" AI systems used in "consequential decisions."[168] The statute requires deployers to provide affected consumers with notice of the use of AI and information sufficient to understand the reasoning for the decision, how AI contributed, data usage, and data sources.[169]
Certain proposed New York legislation has contemplated disclosure requirements for the use of "high-risk" AI systems in consequential decisions. It mandates that deployers provide a statement disclosing the purpose and description of the AI system, the "nature" of the decision, and their contact information before using the system to make a "consequential decision."[170] If the decision that is subsequently made is "adverse to the consumer," the deployer must provide reasons for the decision, how the AI contributed to the decision, the data type, and the data usage.[171] The October 2023 Executive Order on AI governance included non-binding directives to ensure individuals are informed when they are subject to automated decision-making;[172] this

[168] Colo. Rev. Stat. §§ 6-1-1703, 6-1-1704 Colorado SB 24-205 ("Concerning Consumer Protections in Interactions With Artificial Intelligence Systems").
[169] Colo. Rev. Stat. §§ 6-1-1703, 6-1-1704.
[170] https://www.nysenate.gov/legislation/bills/2025/A768.
[171] https://www.nysenate.gov/legislation/bills/2025/A768.
[172] Executive Order 14110 (October 30, 2023), rescinded January 2025.

order did not create enforceable explainability requirements and was rescinded in January 2025.[173] Proposed federal legislation has contemplated studies or rulemaking related to explainability and automated decision transparency,[174] but no comprehensive federal explainability mandate has been enacted as of February 2026.

Explainability is valuable, but it is primarily a first step.
Even where explainability obligations exist, the clarity and accessibility of explanations may vary significantly across providers.

Similarly, entities may attempt to use difficult language to otherwise obfuscate understanding. Information about an AI decision alone may not be sufficient to support action or advocacy. The utility of explanations depends not only on disclosure, but also on whether the information is presented in a manner that consumers can reasonably understand and act upon. Variability in formatting, timing, and detail of explanations may limit their effectiveness in supporting complaints, appeals, or review processes.

173 https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.
174 https://www.congress.gov/crs-product/R48555.

Financial penalties

Entities may encounter fines and monetary penalties for breaking the law.

Legal examples:	CFPB Comment - Comment on Request for Information on Uses, Opportunities, and Risks of Artificial [...]; CFPB Statement - CFPB Acts to Protect the Public from Black-Box Credit Models Using Complex Algorithms CFPB Regulation - Quality Control Standards for Automated Valuation Models Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems; Colorado Legislative Council Staff Fiscal Note - Consumer Protections For Artificial Intelligence Federal Law - Consumer Financial Protection Act; Federal Law / Regulation - Equal Credit Opportunity Act (ECOA) / Reg B; New York Department of Financial Services (DFS) Guidance - Use of Artificial Intelligence Systems [...]; (Proposed) New York Law - New York Artificial Intelligence Consumer Protection Act
Relevant axes:	AI Safety & Human Centeredness · Data Security · Privacy · Accountability & Transparency

Financial penalties for illegal activity or noncompliance provide a key backstop for AI safety in CFPS. Potentially relevant to credit decisioning, payments platforms, insurance and risk underwriting, and fintech interfaces, these measures establish civil penalty frameworks that may influence compliance with security, privacy, and anti-discrimination requirements in AI-mediated financial services.

Broader AI Penalties vary in enforcement pathway and value.
Violations of Colorado's AI Act are enforceable under the Colorado Consumer Protection Act, which authorizes civil penalties of up to $20,000 per violation, and as much as $50,000 in cases involving certain protected classes, in addition to injunctive relief.[175] Less specifically, certain proposed New York legislation would authorize attorney general enforcement and, in some versions, a private right of action for harms arising from unlawful AI use;[176] however, such provisions remained under legislative consideration as of February 2026.[177] The FTC may seek

[175] Colo. Rev. Stat. § 6-1-112; § 6-1-1707,
[176] https://www.nysenate.gov/legislation/bills/2025/S1169/amendment/A.
[177] https://www.nysenate.gov/legislation/bills/2025/S1169/amendment/A.

civil penalties for unfair or deceptive AI-related practices, with maximum per-violation penalties adjusted annually for inflation (currently exceeding $50,000 per violation).[178]

More specific pathways also create potential for unlawful AI CFPS penalties.
Under the Equal Credit Opportunity Act (ECOA), creditors that engage in unlawful discrimination may be liable for actual damages, punitive damages up to $10,000 in individual actions, limited class-action damages, equitable relief, and attorney fees.[179] Discriminating creditors may be liable for actual damages, punitive damages up to $10,000 in individual actions, class-action caps, equitable relief, and attorney fees.[180]

More broadly, the CFPB has stated its ability to leverage its financial enforcement authority to penalize AI-driven financial wrongdoing, including lending discrimination and unfair, deceptive, or abusive practices.[181] The Consumer Financial Protection Act authorizes the CFPB to seek tiered penalties, with maximum daily amounts that increase based on the severity of the violations, including higher penalties for reckless or knowing misconduct.[182]

Penalties may change behavior and guide decisions, but efficacy weakens with larger actors.
While the standard financial penalties can be effective tools, the volume of the current enacted and proposed fees may not sufficiently deter bad behavior. Questions have been raised regarding whether existing civil penalty structures are sufficient to deter large-scale technology providers operating in AI-enabled financial markets.[183] Variability in enforcement actions and penalty amounts may influence market behavior unevenly across entities and sectors. This analysis could also help consumers decide between vendors or products and encourage entities to minimize violations. Differences in enforcement approaches across jurisdictions may result in inconsistent application of penalty standards to AI-mediated financial services.

[178] Federal Trade Commission Act, 15 U.S.C. § 45(m); FTC Civil Penalty Adjustments (adjusted annually for inflation).
[179] 15 U.S.C. § 1691e.
[180] https://www.law.cornell.edu/uscode/text/15/1691e.
[181] https://www.banking.senate.gov/imo/media/doc/gorfine_testimony_9-20-23.pdf: 6-7.
[182] 12 U.S.C. § 5565(c) - around $7,000 daily penalties for rule and order violations, $36,000 penalties for law violations, $1.4M penalties for knowing violations, and substantially larger redress payments (sometimes totalling billions of dollars).
[183]Richie Koch, "Big Tech Earns Enough in Less Than 3 Weeks to Pay All 2024 Fines," *Proton*, January 17, 2025, https://proton.me/blog/big-tech-pays-fines-under-3-weeks.

Human review and algorithmic decision appeals

Consumers may be able to challenge AI-driven decisions and opt for those made by humans.

Legal examples:	(Proposed) California Law - Automated Decisions Safety Act Colorado Law - Concerning Consumer Protections in Interactions With Artificial Intelligence Systems; (Proposed) Federal Law - Artificial Intelligence Civil Rights Act
Relevant axes:	Quality of Redress Mechanisms Consumer Agency & Empowerment · Bias & Fairness · Accuracy

This set of regulatory concepts addresses mechanisms through which consumers may challenge AI-driven outcomes. This either centers on an appeal, a switch to a human decision, or a combination of both. In CFPS, this may be relevant to credit decisioning, price or credit-limit determinations, AI-driven financial planners, mortgage or rental applications, fraud suspicion leading to account holds or closures, and other consequential decisions where model-driven outputs may be opaque or prone to error. Appeal and human review mechanisms are intended to provide an additional layer of oversight in situations where automated decisions may be opaque or contested.

Proposed and enacted laws vary in mechanism but recognize the value of human input. Colorado's AI Act (SB 24-205), enacted in 2024 and effective June 30, 2026,[184] requires deployers of "high-risk AI systems" used in "consequential decisions" to provide notice and an "opportunity to appeal" adverse decisions.[185] Additionally, it allows for human review if "technically feasible," unless providing such review would be impracticable or would harm the consumer.[186]

Proposed federal legislation potentially expands the range of appealable situations, requiring a human review or appeal mechanisms for certain consequential algorithmic decisions.[187] However, such provisions remain under consideration and have not been enacted as of March 2026. Certain proposed California legislation has included structured human review and appeal requirements for AI-facilitated consequential decisions;[188] deployers would need to disclose, after a decision, whether a human reviewed either the AI outputs or the decision itself. After finalization, consumers would have 30 business days to appeal.[189]

[184] https://leg.colorado.gov/bills/sb25b-004.
[185] https://leg.colorado.gov/bills/sb24-205.
[186] https://leg.colorado.gov/bills/sb24-205.
[187] https://www.congress.gov/bill/119th-congress/house-bill/5511/text.
[188] https://legiscan.com/CA/text/AB1018/id/3267938.
[189] https://legiscan.com/CA/text/AB1018/id/3267938.

Gaps remain in breadth, enforceability, and accessibility of appeal and human review rights.
The effectiveness of appeal and human review rights depends on how clearly they are disclosed, how quickly they are processed, and whether reviews meaningfully reassess automated determinations. In practice, variability in implementation may affect whether human review functions as a substantive safeguard or as a procedural formality. Differences in statutory scope and enforcement mechanisms result in uneven availability of appeal and human review rights across jurisdictions.

Juvenile protections

Data and interaction requirements may vary for CFPS that interact with minors.

Legal examples:	Federal Law - Children's Online Privacy Protection Act (COPPA); Federal Trade Commission Regulation - Children's Online Privacy Protection Rule (COPPA Rule); California Law - California Online Privacy Protection Act (CalOPPA); (Enjoined as of writing) California Law - California Age-Appropriate Design Code Act; Colorado Law - Colorado Privacy Act Delaware Law - Delaware Personal Data Privacy Act
Relevant axes:	AI Safety & Human Centeredness · Data Security · Privacy · Consumer Agency & Empowerment

Juvenile protections focus on special protections for children's personal data, requiring overt parental consent, the restriction of targeted advertising, and sometimes allowing minors (or their agents) to opt out of algorithmic decisions. In the context of AI-driven CFPS, these rules and laws could apply to AI-driven financial literacy tools, minor- and teen-oriented banking products, algorithmic risk-scoring for student accounts, gamified savings platforms, and advertisements for any of the aforementioned products. The data-oriented elements of these laws and regulations aim to limit targeted advertising, profiling, and unnecessary data collection involving minors, reducing risks associated with misuse, unauthorized disclosure, or exploitation of children's data. Other elements of these measures may reduce attempts to manipulate children.

Data-protection laws and regulations vary in age coverage and scope.
The Children's Online Privacy Protection Act (COPPA) and the FTC's COPPA Rule apply to online services and children 13 and under, requiring "verifiable parental consent" before collecting personal information from such children.[190] Colorado's Privacy Act requires opt-in consent before processing personal data of known children under 13, for purposes such as targeted advertising or profiling.[191] Delaware law goes further, extending certain enhanced protections to individuals aged 13-17, including restrictions on targeted advertising and limits on the sale of personal data without consent.[192] California's Online Privacy Protection Act (CalOPPA) provides fewer protections, primarily requiring operators of commercial websites and online services to simply post a privacy policy and disclose certain data practices.[193]

[190] 15 U.S.C. §§ 6501–6506; 16 C.F.R. Part 312.
[191] Colo. Rev. Stat. § 6-1-1308: 153.
[192] https://legis.delaware.gov/json/BillDetail/GenerateHtmlDocument?legislationId=140388&legislationTypeId=1&docTypeId=2&legislationName=H#:~:text=(7)%20Not%20process,years%20of%20age.
[193] Cal. Bus. & Prof. Code §§ 22575–22579.

Comprehensive, behaviorally relevant design protections are early and challenged. California's Age-Appropriate Design Code Act (CAADCA) imposes design and data protection obligations on services "likely to be accessed by children,"[194] although elements of the law are currently enjoined.[195] The law requires greater privacy and data minimization standards, risk assessments, language that can be understood by children, notifications for collection of location data, and a ban on dark patterns manipulating children into providing personal information or payment.[196]

Child protections make progress towards data equity but leave meaningful gaps in AI-specific safeguards.
While these measures limit data collection and certain data uses, most are not designed specifically for AI and do not address the distinct harms of algorithmic decisioning or profiling. Age limits also expose older children to greater potential harm with no additional support, and data limits still leave substantial gaps. Although these measures address certain data protection and design concerns, most are not tailored specifically to AI-driven profiling or automated decision-making. Age thresholds and consent frameworks vary across jurisdictions, resulting in uneven protections. As AI-mediated financial tools expand to younger users, questions remain regarding how existing juvenile privacy regimes intersect with algorithmic transparency, appeal rights, and anti-discrimination safeguards.

[194] Cal. Civ. Code §§ 1798.99.28–1798.99.40.
[195] Kevin Angle, Rachel Marmor, "Ninth Circuit Issues Mixed Ruling on California Age-Appropriate Design Code Act," *Holland & Knight,* March 19, 2026, https://www.hklaw.com/en/insights/publications/2026/03/ninth-circuit-issues-mixed-ruling-on-california-age-appropriate-design.
[196] https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.47.&part=4.&chapter=&article=.

Opt-out rights

Consumers may have the right to opt out of the use of AI, directly or indirectly.

Legal examples:	(Proposed) California Law - Automated Decisions Safety Act (Proposed) Federal Law - Artificial Intelligence Civil Rights Act; (Proposed) New York Law - NY Artificial Intelligence Act State Law(s) - Universal Data Collection & Use Opt-Out Rights: CA 1, CA 2, CO, CT, DE, MT, NE, NJ, OR, and TX
Relevant axes:	Consumer Agency & Empowerment · Data Security · Privacy

Opt-out rights serve to provide consumers with the ability to avoid AI-driven decisions or to mitigate the collection and use of data to power those decisions. In CFPS, this could apply to essentially any product or system driven by AI, but is likely salient only for those that still function without algorithmic decision engines. Such measures are intended to enhance consumer control over automated decision-making and may indirectly affect data minimization and privacy practices.

True AI opt-outs are relatively rare, but may be emerging at the federal and state levels. Proposed federal law would direct the FTC to establish rules permitting individuals to opt out of certain automated consequential decisions and request a human decision;[197] however, no such federal opt-out requirement has been enacted as of this writing.[198] Certain proposed New York legislation includes provisions to let consumers opt out of an automated "consequential" decision and receive a human decision within a defined timeframe;[199] these measures remain under consideration and have not been enacted.[200] Initial drafts of the automated decision-making regulations in California's Consumer Privacy Act allowed individuals to opt out of the use of "automated decision systems" that affect them. Those opt-out provisions were reduced in July 2025 and took effect in January 2026. As of this writing, no U.S. jurisdiction guarantees a broad right to opt out of the use of AI.

Broader data-use opt-outs are common and can impact AI training.

---

197 https://www.congress.gov/bill/118th-congress/senate-bill/5152/text#toc-idc1bd46aa7f824437b9f552770b1a317e:~:text=III%E2%80%94TRANSPARENCY-,SEC.%20301.%20NOTICE%20AND%20DISCLOSURE.,-%28a%29%20In: 203.

198 https://www.congress.gov/bill/118th-congress/senate-bill/5152/all-actions.

199 https://legislation.nysenate.gov/pdf/bills/2025/s1169a: 5.

200 https://www.nysenate.gov/legislation/bills/2025/S1169/amendment/A.

Several states, including Colorado,[201] Connecticut,[202] Delaware,[203] Montana,[204] New Jersey,[205] Oregon,[206] and Texas,[207] provide consumers with the right to opt out of certain profiling activities used in "significant" or "consequential" decisions under their respective privacy statutes. In consumer financial contexts, these profiling opt-out rights may limit certain use of personal data in automated decision-making systems, depending on statutory scope and exemptions. In contrast, some states, including California[208] and Nebraska,[209] focus on opt-out primarily around data sharing and the sale of personal data. None of these provisions establish a general right to opt out of AI-driven financial services in their entirety.

Opt-out provisions lack accessibility, need advocacy, and put the burden on consumers. The practical impact of opt-out rights depends on how clearly they are disclosed, how easily they can be exercised, and the extent to which compliance is monitored. Variability in statutory scope and implementation may result in uneven consumer protections across jurisdictions. Opt-out regimes typically place the burden on consumers to take affirmative action, and their effectiveness depends on usability, awareness, and enforcement. Broader structural questions remain regarding whether opt-out frameworks sufficiently address power imbalances in AI-mediated financial markets.

[201] https://leg.colorado.gov/bills/sb21-190.
[202] https://www.cga.ct.gov/asp/cgabillstatus/cgabillstatus.asp?selBillType=Bill&bill_num=SB00006&which_year=2022.
[203] https://legis.delaware.gov/BillDetail?LegislationId=140388.
[204] https://bills.legmt.gov/#/bill/20231/LC1086.
[205] https://www.njleg.state.nj.us/bill-search/2022/S332.
[206] https://olis.oregonlegislature.gov/liz/2023R1/Measures/Overview/SB619.
[207] https://capitol.texas.gov/BillLookup/Text.aspx?LegSess=88R&Bill=HB4.
[208] https://vig.cdn.sos.ca.gov/2020/general/pdf/topl-prop24.pdf.
[209] https://nebraskalegislature.gov/FloorDocs/108/PDF/Slip/LB1074.pdf.

Whistleblowing paths, rewards, and protections

Paths may be required to help insiders share critical information.

Legal examples:	(Proposed) Federal Law - AI Whistleblower Protection Act; Federal Law - Anti-Money Laundering Act Federal Law - Bank Secrecy Act Federal Law - Consumer Financial Protection Act (CFPA); (Proposed) Illinois Law - Artificial Intelligence Safety and Security Protocol Act; (Proposed) New York Law - NY Artificial Intelligence Act; Federal Regulator Tiplines - Consumer Financial Protection Bureau (CFPB), Federal Reserve, Department of Justice Criminal Division, Financial Crimes Enforcement Network, National Credit Union Administration, Office of the Comptroller of the Currency (OCC), Securities and Exchange Commission
Relevant axes:	Accountability & Transparency · AI Safety & Human Centeredness

This regulatory concept creates options and legal protections for individuals with inside knowledge of AI systems to report violations of law or ethics, potentially including discrimination, deceptive interfaces, data-protection failures, or false statements. In CFPS, this could apply to virtually any entity, but may be most applicable for those who directly engage with consumers or consumer data. These measures are intended to enhance transparency and support accountability, safety, and a range of other priorities in situations where internal actors identify potential risk.

AI-specific proposals create paths and protections without awards.
Certain proposed New York legislation has contemplated requiring developers and deployers of "high-risk" AI systems to establish internal processes for employees to report potential legal violations or undisclosed risks and to protect employees from retaliation.[210] As of this writing, such whistleblower-specific provisions remain under legislative consideration.[211] Separately, New York's enacted RAISE Act (2025) emphasizes governance and risk-documentation obligations but does not establish a standalone AI whistleblower compensation framework.[212] The proposed federal AI Whistleblower Protection Act would establish a range of retaliation and relief protections from current or former employees involved in the development or deployment of AI.[213] Under this law, an employee could choose to report to a range of government and law

[210] https://legislation.nysenate.gov/pdf/bills/2025/s1169a: 6.
[211] https://www.nysenate.gov/legislation/bills/2025/S1169/amendment/A.
[212] https://legislation.nysenate.gov/pdf/bills/2025/a6453a.
[213] https://www.congress.gov/bill/119th-congress/senate-bill/1792/text.

enforcement officials[214] that the AI creation includes federal law violations or a failure to respond to serious risk to public safety, public health, or national security. New Jersey's proposed resolution also urges entities to voluntarily commit to whistleblower protocols and employee protections.[215]

Proposed Illinois legislation has included whistleblower protections in situations involving significant risks arising from the development or deployment of foundational AI models, including risks associated with severe physical harm, large-scale financial damage, or violations of criminal law.[216] These provisions remain proposed and have not been enacted.[217] Employees in this scenario could share information with internal leadership and/or government officials and would be protected against retaliation. It is worth highlighting that the standard of risk in this bill is substantially more severe than other instances.

Some CFPS whistleblower channels similarly provide protection but no compensation. The CFPB maintains channels for reporting potential violations of federal consumer financial law,[218] and banking regulators—including the Office of the Comptroller of the Currency,[219] the Federal Reserve,[220] and National Credit Union Administration[221]—provide mechanisms for reporting unsafe or unlawful practices. These channels may be used to report AI-related compliance concerns where applicable statutory violations are implicated.

Other CFPS and financial crime-oriented agencies offer protection as well as compensation.
The Department of Justice's Criminal Division has implemented whistleblower initiatives to encourage reporting of certain categories of corporate misconduct, including financial and sanctions-related violations.[222] These programs are not AI-specific and predate recent shifts in federal AI policy, but they may be implicated where AI systems contribute to underlying unlawful conduct. The Securities and Exchange Commission administers a whistleblower program, established under the Dodd-Frank Act, that provides monetary awards for information leading to

---

214 https://www.congress.gov/bill/119th-congress/senate-bill/1792/text.
215 https://www.njleg.gov/bill-search/2024/AR158/bill-text?f=AR&n=158_I1.
216 https://www.ilga.gov/Legislation/BillStatus/FullText?GAID=18&DocNum=3506&DocTypeID=HB&LegId=0&SessionID=114.
217 https://www.ilga.gov/Legislation/BillStatus?GAID=18&DocNum=3506&DocTypeID=HB&LegId=0&SessionID=114.
218 https://www.consumerfinance.gov/complaint/.
219 https://www.occ.treas.gov/topics/supervision-and-examination/dispute-resolution/consumer-complaints/index-consumer-complaints.html.
220 https://www.federalregister.gov/documents/2025/10/30/2025-19711/unsafe-or-unsound-practices-matters-requiring-attention.
221 https://ncua.gov/about/inspector-general/hotline/portal.
222 https://www.justice.gov/criminal/criminal-division-corporate-whistleblower-awards-pilot-program.

successful enforcement actions involving securities law violations.[223] Although not AI-specific, the program may apply where AI-enabled brokerage or advisory platforms are implicated in securities law violations. The Financial Crimes Enforcement Network administers a whistleblower program under the Bank Secrecy Act that provides monetary awards for information leading to enforcement actions involving violations of anti-money laundering requirements.[224] While not AI-specific, the program may be implicated where AI-driven transaction monitoring systems contribute to underlying compliance failures. All of these programs offer meaningful financial incentives depending on accuracy and outcomes. These incentives could be particularly relevant when AI failures lead to false security claims, market manipulation, fraud, or other financial violations.

Protections are a strong starting point, but whistleblowers and lawmakers need support. The effectiveness of whistleblower regimes depends on awareness, accessibility, and protection against retaliation. Variability in statutory scope and incentive structures may influence whether individuals report AI-related risks. Differences in compensation structures and reporting mechanisms across agencies may affect participation rates and enforcement outcomes. As AI deployment expands in consumer financial services, policymakers continue to debate whether existing whistleblower frameworks sufficiently address risks specific to AI development and deployment.

[223] https://www.sec.gov/enforcement-litigation/whistleblower-program.
[224] https://www.fincen.gov/whistleblower-program.

(EU) European Union Artificial Intelligence Act

This far-reaching, extensive regulation specifically calls out certain AI CFPS.

Link:	https://artificialintelligenceact.eu/ai-act-explorer/
U.S. Legal Examples:	(Proposed) California Law - Automated Decisions Safety Act
Relevant axes:	Bias & Fairness · Resilience · Accountability & Transparency · Consumer Agency & Empowerment · Accuracy · Data Security · Resilience · Quality of Redress Mechanisms

The European Union Artificial Intelligence Act (EUAIA) is the most comprehensive AI regulation enacted globally at the time of this writing. It features a range of analyses and statutes that outpace the U.S. regulatory system in breadth and specificity. It establishes a tiered regulatory framework based on risk level, with the most stringent obligations applied to AI systems that can significantly impact individuals' rights, safety, or livelihoods—including AI CFPS, which are considered "high-risk" systems. This is particularly relevant to any AI CFPS involved in calculating risk, credit scoring, or insurance pricing, but it could also apply to systems that impact professional outcomes, welfare, and medical finance.

AI risk management and oversight expectations are greater.
High-risk AI systems under the EUAIA (which explicitly includes certain AI CFPS products and may extend to a broader range of systems) must follow a detailed risk-management process.[225] This requirement applies both to deployment and throughout the AI's use.[226] Developers must identify and mitigate foreseeable risks to "health, safety, and fundamental rights," while also designing systems to allow effective human oversight and intervention.[227] Comparable U.S. laws and regulations, such as the Equal Credit Opportunity Act (ECOA) and Regulation B, impose nondiscrimination rules and mandate consumer notices about negative actions regarding their credit,[228,229] but these rules do not require the proactive, design-stage risk evaluation or embedded human-in-the-loop controls that the EUAIA mandates. While the proposed federal AI

[225] https://artificialintelligenceact.eu/article/6/.
[226] https://artificialintelligenceact.eu/article/9/.
[227] https://artificialintelligenceact.eu/article/9/.
[228] https://www.consumerfinance.gov/rules-policy/regulations/1002/.
[229] https://www.justice.gov/crt/equal-credit-opportunity-act-3.

Civil Rights Act,[230] Colorado's AI Act,[231] and California's proposed Automated Decisions Safety Act[232] introduce pre-deployment assessments and certain oversight obligations for high-risk systems, they are narrower in scope and lack the continuous, legally binding risk-management lifecycle the EUAIA imposes. AI laws from Utah[233] and Texas[234] lack the EU's substantive risk-management duties as well, but do include disclosure and consumer notice requirements.

Entities must follow standards of data quality, documentation, and transparency
The EUAIA requires AI training, validation, and testing datasets to be as error-free as possible, demographically representative, and free from bias.[235] Developers must prepare comprehensive technical documentation describing the AI's design, intended use, and performance and supply deployers with substantial instruction.[236] This contrasts with U.S. laws like the Fair Credit Reporting Act[237] and Gramm–Leach–Bliley Act,[238] which address data accuracy and security but do not mandate representativeness testing, bias-mitigation protocols, or pre-deployment documentation to this level of specificity.

Certain features from this portion of the EUAIA are found distributed across proposed and enacted U.S. AI laws, though none are without gaps. The proposed federal AI Civil Rights Act includes representativeness checks for training data and performance testing across demographic groups,[239] and the proposed federal Algorithmic Accountability Act mandates data provenance documentation and system summaries.[240] Colorado law requires disclosure of data

---

230 https://www.congress.gov/bill/119th-congress/house-bill/6356/text#toc-H33854E7B53FC49D7A0F81E2C4E70543A.
231 https://leg.colorado.gov/bill_files/47770/download: 12.
232 https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB1018.
233 https://le.utah.gov/~2024/bills/static/SB0149.html.
234 https://capitol.texas.gov/tlodocs/89R/analysis/html/HB00149S.htm.
235 https://artificialintelligenceact.eu/article/10/.
236 https://artificialintelligenceact.eu/article/11/.
237 https://www.ftc.gov/system/files/ftc_gov/pdf/fcra-march-2026.pdf.
238 https://www.ftc.gov/business-guidance/privacy-security/gramm-leach-bliley-act.
239 https://www.congress.gov/bill/118th-congress/senate-bill/5152/text/is#:~:text=bb)%20defined%20benchmarks.such%20demographic%20groups%3B.
240 https://www.congress.gov/bill/119th-congress/house-bill/5511/text#:~:text=(7)%20Maintain%20and.of%20data%20fields%3B.

types and sources in impact assessments,[241] while Texas[242] and Utah[243] have no dataset-quality mandates, focusing instead on transparency and disclosure.

Continuing testing and resilience are fundamental to EUAIA-compliant model development.
Developers must log data to conduct live testing after a model has been released into use,[244] meet measurable accuracy thresholds,[245] and maintain resilience against errors or malicious attacks.[246] These requirements go beyond the U.S. Interagency Guidelines on Information Security—which are extensive and focus on safeguarding customer information through a range of administrative, technical, and physical controls—by adding performance and robustness standards for AI models as well.[247] California's proposed Automated Decisions Safety Act[248] and Colorado's AI law add recurring testing and transparency requirements,[249] and the Federal Algorithmic Accountability Act calls for ongoing evaluation and reporting to the FTC.[250] However, none require the kind of centralized post-market monitoring and binding accuracy standards found in the EUAIA. Texas and Utah laws provide no traceability or model-performance obligations. The EUAIA also requires monitoring and incident reporting to national authorities,[251] whereas the U.S. lacks a unified AI incident-reporting framework.

241 https://leg.colorado.gov/bills/sb24-205#:~:text=a%20high%2Drisk%20system%3B-,Making%20a%20publicly%20available%20statement%20summarizing%20the%20types%20of%20high%2Drisk,days%20after%20the%20discovery%2C%20that%20the%20high%2Drisk%20system%20has%20caused,-A%20person%20doing%20business.

242 https://capitol.texas.gov/tlodocs/89R/analysis/html/HB00149S.htm#:~:text=in%20this%20state.-,Sec.%20551.003.,-CONSTRUCTION%20AND%20APPLICATION.

243 https://le.utah.gov/~2024/bills/static/SB0149.html#:~:text=%E2%96%B8requires%20disclosure%20when%20an%20individual%20interacts%20with%20AI%20in%20a%20regulated%20occupation%3B.

244 https://artificialintelligenceact.eu/article/12/.

245 https://artificialintelligenceact.eu/article/15/.

246 https://artificialintelligenceact.eu/recital/115/.

247 https://www.ecfr.gov/current/title-12/chapter-III/subchapter-B/part-364/appendix-Appendix%20B%20to%20Part%20364.

248 https://leginfo.legislature.ca.gov/faces/billStatusClient.xhtml?bill_id=202520260AB1018.

249 https://leg.colorado.gov/bill_files/47770/download: 9.

250 https://www.congress.gov/bill/119th-congress/house-bill/5511/text.

251 https://artificialintelligenceact.eu/article/70/.

(EU) General Data Protection Regulation (GDPR)

This far-reaching, extensive regulation specifically calls out certain AI CFPS.

Link:	https://gdpr-info.eu/
U.S. Legal Examples:	Federal Law - Gramm-Leach-Bliley Act; FCRA California's CPRA Equal Credit Opportunity Act TC's "Safeguards Rule" and the Interagency Security Guidelines
Relevant axes:	Bias & Fairness · Resilience · Accountability & Transparency · Consumer Agency & Empowerment · Accuracy · Data Security · Resilience · Quality of Redress Mechanisms

The EU's General Data Protection Regulation (GDPR) is a far-reaching privacy and data protection law. The regulation establishes strict requirements around data processing, data-purpose limitations, transparency, accuracy, and accountability. It grants individuals access, correction, deletion, and appeals rights over their own data. Regarding AI CFPS, GDPR's definitions of "high-risk processing" effectively include algorithmic profiling used in credit, lending, and insurance.

GDPR data minimization and purpose limitation requirements are stricter than in the U.S. GDPR requires that personal data be collected only when necessary, used for specified purposes, and processed only on the basis of consent, a contract, or legal duty.[252] For AI CFPS, this means a credit risk model or fraud detector cannot hoard irrelevant personal data or repurpose it for marketing without new legal justification. In the U.S., the Gramm-Leach-Bliley Act (GLBA) limits some financial data sharing,[253] and the Fair Credit Reporting Act (FCRA) restricts acceptable uses for a credit report,[254] but there is no nationwide rule requiring categorical data minimization or lawful bases for processing. California privacy laws also include partial limits but lack the GDPR's binding framework.[255]

[252] https://gdpr-info.eu/art-6-gdpr/.
[253] https://www.ftc.gov/business-guidance/privacy-security/gramm-leach-bliley-act.
[254] "The Fair Credit Reporting Act (FCRA)", *Electronic Privacy Information Center,* https://epic.org/fcra/#:~:text=The%20FCRA%20limits%20the%20use,%2C%20promotion%2C%20reassignment%20or%20retention.
[255] Alice Marini et al., "Comparing privacy laws: GDPR v. CCPA," *DataGuidance* and *Future of Privacy Forum,* https://fpf.org/wp-content/uploads/2018/11/GDPR_CCPA_Comparison-Guide.pdf: 5.

Accuracy, fairness, and the right to challenge automated decisions extend to AI. GDPR mandates that personal data be accurate and kept current. It also gives individuals the right to data correction[256] and deletion.[257] Furthermore, the regulation prohibits situations in which consumers are subject solely to automated decisions with legal effects (unless safeguards such as human review exist).[258] For AI CFPS, this means models must update and correct credit or fraud data and customers must be able to contest AI outcomes. U.S. laws including FCRA[259] and the Equal Credit Opportunity Act (ECOA)[260] impose accuracy and notice duties, but they apply unevenly across CFPS types. The proposed federal AI Civil Rights Act approximates GDPR's fairness obligations but remains narrower in reach.[261] Colorado's Privacy Act also grants some access and deletion rights,[262] but lacks GDPR's explicit obligations to explain decision logic or provide portability.

Security, breach notification, and accountability duties are broader under GDPR. The regulation requires technical and organizational safeguards for confidentiality, integrity, and resilience, compels notification of breaches within 72 hours, and forces organizations to prove compliance.[263] U.S. legal examples like the FTC's "Safeguards Rule[264]" and the Interagency Security Guidelines[265] impose similar controls on banks, and state breach laws require disclosures, but timelines and coverage vary. The U.S. lacks GDPR's uniform 72-hour breach rule and universal accountability duty.

Sensitive data and profiling face stricter limits under GDPR than in the U.S. GDPR also restricts the use of special category data (e.g., health, race, political beliefs) and automated profiling with significant effects, unless explicit consent or narrow legal bases exist.[266] AI CFPS tools using such data—for example, in insurance underwriting—must include human

[256] https://gdpr-info.eu/art-16-gdpr/.
[257] https://gdpr-info.eu/art-17-gdpr/.
[258] https://gdpr-info.eu/art-22-gdpr/.
[259] https://www.ftc.gov/system/files/ftc_gov/pdf/fcra-march-2026.pdf: 47.
[260] https://www.govinfo.gov/content/pkg/USCODE-2011-title15/html/USCODE-2011-title15-chap41-subchapIV.htm.
[261] Sofia Gracias, "Comparing the EU AI Act to Proposed AI-Related Legislation in the US," *The University of Chicago Business Law Review*, March 14, 2024, https://businesslawreview.uchicago.edu/online-archive/comparing-eu-ai-act-proposed-ai-related-legislation-us.
[262] https://coag.gov/resources/colorado-privacy-act/#:~:text=The%20CPA%20grants%20Colorado%20Consumers.or%20certain%20kinds%20of%20profiling.
[263] https://gdpr-info.eu/art-33-gdpr/.
[264] https://www.ecfr.gov/current/title-16/chapter-I/subchapter-C/part-314 Notification is to the FTC, and they have 30 days.
[265] https://www.federalreserve.gov/supervisionreg/interagencyguidelines.htm must notify consumer "as soon as possible" but no firm timeline.
[266] https://gdpr-info.eu/art-9-gdpr/.

review and strict legal justification. In the U.S., ECOA prohibits credit discrimination[267] and CPRA protects some "sensitive personal information,"[268] but no broad law bans categories of data from CFPS use outright. Recent U.S. AI proposals focus on testing and discrimination impacts, not categorical restrictions.

267 https://uscode.house.gov/view.xhtml?req=granuleid%3AUSC-prelim-title15-chapter41-subchapter4&edition=prelim

268 https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5: 1798.121.

(China) Provisions on the Management of Algorithmic Recommendations in Internet Information Services

This far-reaching, extensive regulation specifically calls out certain AI CFPS.

Link:	English translation: https://www.chinalawtranslate.com/en/algorithms/ Original: https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm
U.S. Legal Examples:	Privacy laws in California, Colorado, and Connecticut FTC click to cancel ECOA / Reg B
Relevant axes:	Bias & Fairness · Resilience · Accountability & Transparency · Consumer Agency & Empowerment · Accuracy · Data Security · Resilience · Quality of Redress Mechanisms

These binding national laws govern how "algorithmic recommendation" services operate across China's digital ecosystem. With regards to AI CFPS, these statutes create disclosure requirements, allow consumers to opt out of targeted content, place guardrails on algorithmic pricing, require audits, and set oversight obligations that would affect large platforms offering financial products or steering consumers toward them.

Transparency, user choice, and control are mandatory (and very specific).
Relevant entities must clearly tell consumers when they use algorithmic recommendations and publicly explain the "basic principles, purpose, and main operating mechanisms" of the algorithm.[269] They must also provide an easy way either to turn personalization off entirely or to choose a non-personalized option.[270] These providers must also let users view or delete the tags used for personalization.[271] If an algorithm significantly affects user rights, the provider must "give an explanation" and assume corresponding responsibility.[272]
In the U.S., state privacy laws in California,[273] Colorado,[274] and Connecticut[275] offer opt-outs from targeted ads or certain kinds of profiling, but they generally do not require a prominent, built-in "non-personalized feed" alternative or tag-deletion tools. Colorado's AI law creates disclosure

[269] https://www.chinalawtranslate.com/en/algorithms/, Article 16.
[270] https://www.chinalawtranslate.com/en/algorithms/, Article 17.
[271] https://www.chinalawtranslate.com/en/algorithms/, Article 17.
[272] https://www.chinalawtranslate.com/en/algorithms/, Article 17.
[273] https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5: 1798.185, 18 A.
[274] https://coag.gov/app/uploads/2022/01/SB-21-190-CPA_Final.pdf: 18.
[275] https://www.cga.ct.gov/2022/ACT/PA/PDF/2022PA-00015-R00SB-00006-PA.PDF: 9-10.

and post-decision explanation duties for "high-risk" consequential decisions,[276] but it lacks a mandate for a universal non-personalized mode. The Equal Credit Opportunity Act (ECOA) / Regulation B houses reasoning for adverse actions in credit files[277] (often delivered weeks later by mail) rather than in proactive, system-level interactions.

Pricing fairness and anti-manipulation are front and center.
For any consumer sale or service, relevant AI CFPS providers may not use algorithms to impose "unreasonable differential treatment in price or other trading conditions" based on consumer preferences or habits.[278] The rules also forbid manipulating rankings, search results, or reviews and social media cues to distort presentation.[279] In the U.S., there is no broad ban on algorithmic price discrimination that affects CFPS.[280] Credit-specific discrimination remains illegal (ECOA/Reg B),[281] and the FTC has targeted "dark patterns," especially negative-option traps, but neither creates a general bar on data-driven price differentials in financial services marketing nor fully prevents manipulative presentation to achieve a sale.[282] Colorado's AI Act prohibits "algorithmic discrimination," but its focus is on protected-class harm in consequential decisions rather than blanket prohibition on dynamic pricing.[283]

Governance, audits, and safety reviews are required.
Under China's law, relevant AI CFPS providers must periodically audit and verify algorithm mechanisms, models, data, and outputs, and they may not design models that induce addiction or over-consumption.[284] By contrast, U.S. financial institutions follow sector-specific security and model-risk practices (e.g., bank safety-and-soundness, state privacy rules on consent and dark patterns), and Colorado's AI law additionally includes impact assessments and discrimination-risk management for "high-risk" decisions.[285] There is, however, no national or cross-sector mandate to periodically audit recommendation algorithms or to prevent over-consumption.

Special-population and harmful-behavior guardrails are explicit.

---

[276] https://leg.colorado.gov/bill_files/47770/download: 14.
[277] https://www.consumerfinance.gov/rules-policy/regulations/1002/9/.
[278] https://www.chinalawtranslate.com/en/algorithms/, Article 21.
[279] https://www.chinalawtranslate.com/en/algorithms/, Article 14.
[280] Max Morgan, "How States Are Taking on Algorithmic Pricing," *Tech Policy Press*, March 20, 2026, https://www.techpolicy.press/how-states-are-taking-on-algorithmic-pricing/.
[281] https://ncua.gov/regulation-supervision/manuals-guides/federal-consumer-financial-protection-guide/compliance-management/lending-regulations/equal-credit-opportunity-act-regulation-b.
[282] https://www.ftc.gov/system/files/ftc_gov/pdf/P214800%20Dark%20Patterns%20Report%209.14.2022%20-%20FINAL.pdf.
[283] https://leg.colorado.gov/bills/sb24-205.
[284] https://www.chinalawtranslate.com/en/algorithms/, Article 8.
[285] https://leg.colorado.gov/bill_files/47770/download: 6.

Providers must meet minor protection requirements by minimizing unsafe or addictive content.[286] They must also follow access and scam prevention obligations meant to protect elderly consumers.[287] In the U.S., the Children's Online Privacy Protection Act focuses on data collection for those under 13. California's Age-Appropriate Design Code Act (CAADCA), currently enjoined and under judicial review, imposes design and data protection obligations for services "likely to be accessed by children." Financial consumer rules protect against unfair, deceptive, and abusive practices (and regulators take particular notice of young people and the elderly), but there are no CFPS-specific algorithmic duties for minors or seniors akin to these provisions.

[286] https://www.chinalawtranslate.com/en/algorithms/, Article 18.
[287] https://www.chinalawtranslate.com/en/algorithms/, Article 19.

(Report summaries) Notable gaps across axes

A high-level summary of notable absences in frameworks and regulations.

Framework Overview

Across the axes identified in this report, many existing frameworks identify risks associated with generative AI but provide limited operational guidance for mitigating these risks in practice. Two structural factors appear to contribute to this gap. First, generative AI systems involve the highly variable and unstructured nature of a user's interactions, which complicates both the testing and the design of comprehensive safeguards. Second, many generative models lack sufficient explainability, making it difficult to trace how popular outputs are produced and to predict how risks may manifest across different use contexts.

Mitigation efforts often focus on addressing certain known failure modes through retraining or targeted adjustments. While such interventions may reduce identified harms, they do not necessarily prevent similar issues from emerging in new forms. Because generative systems optimize probabilistically based on training objectives, rather than on rule-based reasoning,[288] ensuring consistent alignment across evolving use cases remains technically and institutionally challenging. These dynamics raise questions about whether existing regulatory and evaluative tools are sufficient to oversee systems whose behavior is neither fully interpretable nor procedurally determined.

Industry-developed frameworks similarly vary in depth and specificity. Many articulate high-level commitments to safety, fairness, or transparency but provide limited detail regarding implementation mechanisms, measurement methodologies, or independent verification. While some firms have published more structured governance documentation,[289] coverage across risk categories remains uneven. In addition, certain compliance-oriented frameworks are proprietary or not publicly accessible,[290] limiting comparative analysis.

Lastly, some of the most technically advanced proposals for risk mitigation remain concentrated in academic literature and may require significant expertise and resources to implement. The feasibility of deploying such measures at scale—absent regulatory requirements—remains uncertain. Even where regulations exist, they are only beginning to address the complexity of this technology, and monitoring compliance across rapidly evolving AI systems will present significant challenges.

Regulatory Overview

---

[288] Parshin Shojaee et al., “The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity,” *Apple Machine Learning Research*, June 2025, https://machinelearning.apple.com/research/illusion-of-thinking.

[289] See, for example, “InterpretML: Interpretable Machine Learning,” *InterpretML*, 2024, https://interpret.ml/docs/#.

[290] Nicholas Berente et al., “Why invest in AI ethics and governance?,” *IBM Institute for Business Value*, December 16, 2024, https://www.ibm.com/thought-leadership/institute-business-value/report/roi-ai-ethics.

Across the 12 axes of evaluation on which we relied for this report, two axes failed to be represented in drafted or enacted regulations, laws, or guidance. Here, we examine this gap and explore needs and opportunities.

Resilience has been discussed by some regulators, but we found no legal examples. Consumers depend on CFPS for nearly every major aspect of their lives, nearly every day. As AI controls a larger share of CFPS and the services that drive them, AI systems' ability to respond to challenges is increasingly important. However, these systems are dependent on an interconnected web of secondary service providers including cloud hyperscalers like AWS, Microsoft Azure, and Google Cloud Platform. As these models become increasingly complex, the web of infrastructure and attendant dependencies expands to include cross-cloud agentic AI, various serverless compute microservices, and a range of other secondary tools.

When these providers suffer outages,[291] AI CFPS' resilience will be tested. In that situation, the most effective systems will leverage backup procedures to manage workload and continue to function at a consistent accuracy. Less effective, but relatively safe (depending on context) CFPS may fail transparently and safely. However, the least resilient—and most high-risk—services may ostensibly continue to perform their tasks but silently degrade in quality. This outcome could cause tremendous damage as credit, financial advice, investment fraud, compliance, customer service, transaction tracking, and a host of other AI-dependent CFPS processes would proceed incorrectly and unmonitored.

While financial regulators impose operational resilience, incident reporting, and third-party risk management obligations, we did not identify regulatory frameworks that specifically address AI model resilience—particularly risks such as silent degradation, cross-cloud dependency failures, or performance deterioration under stress conditions in AI-driven CFPS. The Consumer Financial Protection Bureau and other regulators have expressed concerns regarding concentration risk and reliance on a small number of critical service providers,[292] but these discussions have not yet resulted in AI-specific resilience requirements for model performance or degradation risk. Similarly, the Federal Trade Commission has raised concerns around "single points of failure",[293] warning that the small number of dominant cloud service providers could lead to cascading technology outages (including essential CFPS), but has not taken substantial steps related to this issue.

---

[291] *Georgetown University*, "Is Global Tech Infrastructure Too Vulnerable? Professor Responds to CrowdStrike, Microsoft Outage," July 25, 2024, https://www.georgetown.edu/news/ask-a-professor-crowdstrike-outage/.

[292] Rohit Chopra, "Director Chopra's Opening Remarks to the Community Bank and Credit Union Advisory Councils," *Consumer Financial Protection Bureau*, April 7, 2022, https://www.consumerfinance.gov/about-us/newsroom/director-chopras-opening-remarks-to-the-community-bank-and-credit-union-advisory-councils/.

[293] Nick Jones, "Cloud Computing RFI: What We Heard and Learned," *Federal Trade Commission*, November 16, 2023, https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/11/cloud-computing-rfi-what-we-heard-learned.

As AI systems become more deeply embedded in consumer financial infrastructure, questions remain regarding how existing operational resilience frameworks apply to AI-specific risks, including cascading infrastructure failures and non-transparent performance degradation.

Advocates and companies warn about sycophancy, but no salient legal examples emerge.

Sycophancy is the tendency of AI to be overly agreeable. This tendency manifests not simply as positivity but rather as a deeply inset deceptive pattern that can purportedly encourage dangerous health decisions,[294,295] support drug use,[296] isolate consumers from the outside world,[297] or even support violence.[298] Industry and academic discussions have identified over-agreeability and overconfidence in generative AI systems as emerging risk factors,[299] particularly where users rely on AI outputs for consequential decisions. In May 2025, OpenAI rolled back a complete update to its conversational AI because the update's tendency to be "overly flattering or agreeable" triggered "safety concerns".[300]

Within the context of AI CFPS, this poses substantial concerns to both AI products used within financial institutions and those that interact directly with consumers. Internally, depending on the structure of the AI, there are concerns that over-agreeability could amplify existing biases in human-AI interactions if model outputs reinforce user assumptions without sufficient constraint or corrective signals. This could additionally lead to poorer, potentially biased customer service scripts and strategic decisions. For external-facing AI, this could lead to accuracy issues (for example a scenario in which the AI describes credit card features or company policies that do not exist[301]), poorer customer service chatbots, or problematic financial advice that hurts a consumer's financial future (and exposes a company to additional liability).

---

[294] Ziv Ben-Zion et al., "Assessing and alleviating state anxiety in large language models," *Nature Digital Medicine*, March 3, 2025, https://www.nature.com/articles/s41746-025-01512-6.
[295] Tom Gerken, "Update that made ChatGPT 'dangerously' sycophantic pulled," *BBC News*, April 30, 2025, https://www.bbc.com/news/articles/cn4jnwdvg9qo.
[296] Cade Metz, "As A.I. Chatbots Get Smarter, They May Also Become More Convincing Conspiracy Theorists," *The New York Times*, June 13, 2025, https://www.nytimes.com/2025/06/13/technology/chatgpt-ai-chatbots-conspiracies.html.
[297] Kerrin Artemis Jacobs, "Digital loneliness—changes of social recognition through AI companions," *Frontiers in Digital Health*, March 5, 2024, https://pmc.ncbi.nlm.nih.gov/articles/PMC10949182/.
[298] Kevin Collier, "OpenAI Rolls Back ChatGPT Update After Users Complain the Bot Is Too Sycophantic," *NBC News*, May 2025, https://www.nbcnews.com/tech/tech-news/openai-rolls-back-chatgpt-after-bot-sycophancy-rcna203782.
[299] Alex Engler, "Breaking the AI Mirror," *Brookings Institution*, 2024, https://www.brookings.edu/articles/breaking-the-ai-mirror/.
[300]Kevin Collier, "OpenAI Rolls Back ChatGPT Update After Users Complain the Bot Is Too Sycophantic," *NBC News*, May 2025, https://www.nbcnews.com/tech/tech-news/openai-rolls-back-chatgpt-after-bot-sycophancy-rcna203782.
[301] Maria Yagoda, "Airline held liable for its chatbot giving passenger bad advice - what this means for travellers," *BBC Travel*, February 23, 2024, https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know.

We did not identify regulatory frameworks that explicitly address AI "sycophancy" as a model-design risk. While existing unfair, deceptive, or abusive practices doctrines may reach certain harms resulting from overconfident or misleading AI outputs, current statutes do not specifically evaluate or measure over-agreeability as a safety dimension in AI-driven financial services. Former regulatory leaders from the Consumer Financial Protection Bureau and Federal Trade Commission have raised concerns about this issue,[302] but there does not appear to be any prominent regulatory activity addressing it.

The absence of explicit standards addressing overconfidence or validation bias in AI systems highlights a potential gap between existing consumer protection doctrines and emerging model-level behavioral risks.

[302] Erie Meyer et al., "Tech Brief: AI Sycophancy & OpenAI," *Georgetown Law Institute for Technology Law & Policy*, July 30, 2025, https://www.law.georgetown.edu/tech-institute/research-insights/insights/tech-brief-ai-sycophancy-openai-2/.

What This Landscape Demands: A Consumer Mandate for AI in Financial Services

This landscape analysis reveals not only technical gaps in AI evaluation frameworks and regulatory approaches, but also a profound disconnect between AI deployment and consumer welfare.

The evidence is unambiguous:

- **Consumers know they're unprotected:** 57% say current laws do not adequately protect them from risks related to AI in financial services; only 8% disagree.[303]
- **Consumers know what they need:** 92% want to be informed when AI is involved in financial decision-making; 77% want the right to opt out of AI-driven decisions on major financial matters; and 58% believe AI systems should be monitored continuously for fairness and accuracy.[304]
- **Consumers aren't getting it**: Fewer than 10% of Americans completely trust financial companies to use AI responsibly in any financial service surveyed. After considering the issues raised, 42% report feeling negatively about AI's growing role in finance, compared with only 18% who feel positively.[305]

These findings are not abstract expressions of technological anxiety. They reflect concrete experience. Among consumers who report encountering AI in financial services, positive experiences are concentrated in contexts where AI clearly protects consumer interests—such as fraud detection. In other applications, including customer service chatbots and automated credit decisions, positive sentiment declines sharply. Trust erodes when AI appears to serve institutional efficiency rather than consumer protection.

This trust deficit exists because—as this landscape demonstrates—no coherent, consumer-centered standard defines what AI systems owe to the people whose financial lives they influence.

Frameworks address pieces of the problem: bias metrics in one domain, transparency guidance in another, privacy safeguards elsewhere. Regulations remain fragmented across jurisdictions and subject areas. Meanwhile, deployment frequently outpaces understanding. Generative AI systems are integrated into financial interactions before their behavior can be reliably explained, transforming technical opacity into a substantial governance challenge.

---

[303] Consumer Reports nationally representative AI in Financial Services Survey of 4,073 U.S. adults, (September-October, 2025), https://article.images.consumerreports.org/image/upload/v1767035543/prod/content/dam/surveys/Consumer_Reports_AI_in_Financial_Services_2025.pdf.
[304] Ibid.
[305] Ibid.

At the same time, market incentives work against consumer welfare. Infrastructure providers profit from increased AI adoption regardless of downstream consumer outcomes. Corporate leadership is rewarded for adoption speed, automation gains, and efficiency metrics—not for demonstrable improvements in fairness, accountability, or redress. In such an environment, voluntary principles predictably fail to function as effective guardrails.

Taken together, these dynamics reveal three structural gaps:

1. **No Coherent Consumer-Centered Standard:** There is no integrated definition of what AI systems owe consumers across transparency, fairness, agency, redress, resilience, and reliability.
2. **Deployment Outpaces Understanding:** Systems are embedded in consequential financial decisions despite persistent limitations in explainability, confidence calibration, and oversight.
3. **Market Forces Work Against Consumers**: Incentive structures reward adoption and scale over validation and protection.

The axes of evaluation presented throughout this report are therefore more than analytical categories. Collectively, they define the potential contours of a consumer-centered evaluation standard that the current market and regulatory landscape lacks.

Addressing these gaps will require institutions capable of translating fragmented principles into coherent, testable standards, making systemic harms visible, and ensuring that evaluation tools and metrics are legible to consumers as well as to technical experts.

Addressing these gaps will require:

- **Integrated Standards:** The translation of fragmented principles into coherent, testable consumer welfare criteria that define what AI systems owe consumers.
- **Visible Accountability:** Independent evaluation and investigation capable of identifying and illustrating framework failures in everyday financial interactions.
- **Public Legibility:** The development of standards and evaluation tools that are interpretable and usable by consumers, policymakers, and companies—not solely by technical specialists.

Absent such a coherent, legible standard, the trajectory of AI in financial services will continue to be shaped primarily by market incentives rather than consumer priorities.

Major open questions

This landscape analysis reveals not only fragmentation in frameworks and regulation, but a set of unresolved governance questions at the core of AI deployment in consumer financial services. These questions are structural, not technical; they persist because existing tools address components of the problem without integrating them into a coherent consumer-centered standard.

Generative AI Explainability and Accountability
Large language models can be evaluated for bias, accuracy, and sycophancy, yet their internal reasoning processes remain opaque—even to their developers. In financial contexts, this opacity complicates accountability.

Open questions include:

- How can consumers challenge decisions that companies themselves struggle to explain?
- What constitutes “reasonable care” when deploying systems that cannot be fully interpreted?
- Should explainability thresholds apply to consequential financial decisions?

When technical opacity becomes embedded in credit, pricing, or advisory systems, it evolves into a governance problem.

Dynamic and Algorithmic Pricing
AI-mediated pricing challenges long-standing assumptions about visibility and contestability in financial markets. Survey findings indicate that many consumers believe personalized pricing is used, yet few report knowingly encountering it—suggesting either opaque deployment or detection barriers.

Open questions include:

- How should regulators distinguish between legitimate risk-based pricing and discriminatory algorithmic pricing?
- What transparency obligations apply when pricing is dynamically determined?
- Can existing fair-lending laws adequately address algorithmic pricing harms?

Sycophancy and Overconfident Advice
Generative systems tend toward validation rather than principled disagreement. In financial contexts, this may encourage risky decisions while projecting unwarranted confidence.

Open questions include:

- Should financial AI systems be tested for overconfidence and validation bias prior to deployment?
- How do fiduciary or suitability standards apply to engagement-optimized AI advisors?

- Can regulators meaningfully differentiate between personalization and harmful validation?

The Adequacy Gap
A majority of Americans believe current laws do not adequately protect them from AI-related financial risks. Positive experiences are concentrated in clearly protective uses, such as fraud detection; skepticism increases when AI replaces human judgment in consequential decisions.

This gap prompts broader questions:
- Do existing statutes (ECOA, FCRA, CFPA) sufficiently address AI-specific harms?
- Should regulation focus on technological processes or measurable outcomes?
- How should federal and state roles interact in shaping baseline protections?

Agentic AI and Responsibility
Emerging systems capable of autonomous action further complicate accountability.

Open questions include:
- What level of opt-in should be required for AI-initiated financial actions?
- How should liability be allocated when autonomous systems cause harm?
- What monitoring standards apply to evolving, goal-directed systems?

Market Structure and Incentives
AI deployment incentives prioritize speed, scale, and cost efficiency. Consumers, however, prioritize security, affordability, and service quality.

Key questions include:
- How can market incentives be recalibrated toward consumer-welfare outcomes?
- Can independent evaluation create competitive pressure for safer systems?
- What mechanisms ensure efficiency gains translate into consumer benefit?

Cross-Border Governance
AI systems deployed domestically may be developed and trained globally, under divergent regulatory regimes.

Open questions include:
- How should domestic standards apply to internationally developed models?
- How can regulators address jurisdictional gaps in enforcement?

What These Questions Signal
Taken together, these unresolved issues reflect a deeper structural problem: fragmentation across frameworks, regulatory approaches, and accountability chains. This landscape clarifies the terrain but does not resolve these issues. The next phase of this work builds on this mapping exercise to develop an integrated, consumer-centered evaluation standard capable of addressing these gaps in practice.

Appendix: Key Consumer Findings on AI in Financial Services

This appendix summarizes core findings from Consumer Reports' 2025 nationally representative survey on AI in financial services. These data provide the empirical foundation for the analysis presented throughout this report and illustrate the magnitude of the consumer protection gap.

Potential groupings:

- Trust & Regulatory Adequacy
- Transparency & Opt-Out Expectations
- Redress & Human Review Preferences
- Privacy Comfort Levels
- Experience by Use Case
- Demographic Differences in Trust and Concern

This appendix enables readers to reference underlying data directly and reinforces the report's evidentiary grounding.

Acknowledgements

Consumer Reports' Digital Marketplace strategy is supported by a host of staff members and consultants across the organization. We would like to thank the following people for their contributions to this report: Delicia Hand, Stephanie Landry, Kevin Doyle, Noemi Altman, Wendy Greenfield, James Brock, Chris Griggs, Tim La Palme, Jonea Gurwitt, Tracy Anderman, Adam Pickersgill, Varun Gadh, Matthew McBride, Pragathi Balasubramanian, and Maria Bazan.

Contributions to this report included strategy, research, analysis, drafting, and editorial development.

Consumer Reports' exploration of artificial intelligence in financial services is part of a broader initiative to monitor, evaluate, and strengthen consumer protections in the marketplace, work that is made possible, in part, by a grant from the Sloan Foundation.

We thank everyone for their support of this work.