August 12, 2024

The Honorable Janet Yellen
Secretary, Department of the Treasury
1500 Pennsylvania Ave NW
Washington, DC 20220

RE: Uses, Opportunities, and Risks of Artificial Intelligence in the Financial Services Sector

Dear Secretary Yellen,

Consumer Reports[1] appreciates the opportunity to comment on the Department of Treasury's Request for Information on Uses, Opportunities, and Risks of Artificial Intelligence in the Financial Services Sector. Financial institutions have long been at the forefront of leveraging technology for core business operations, and artificial intelligence (AI) is no different. For example, algorithms have been used for decades in underwriting and trading. However, recent advances in AI, particularly in machine learning (ML), represent game-changing advancements.

The ability of these new models to learn by themselves, combined with the availability of enhanced computing powers, has opened the door to advanced analytics leveraging alternative and unstructured data. AI/ML provides the ability to analyze and automate with greater sophistication and efficiency and is increasingly being employed by financial institutions for both back-end and front-end operations. There are a wide range of AI use cases in finance, ranging from powering digital chatbots and virtual assistants, to augmenting or even automating credit underwriting, to digital marketing and fraud monitoring.

AI/ML can have many potential benefits for financial consumers, including increasing access to credit for traditionally underserved consumers with limited credit histories, expanding the availability of new and innovative products potentially at lower cost, and providing faster customer service.

But in nearly all of these instances, AI/ML is a double-edged sword. The same AI use cases that can benefit consumers also pose risks depending on how AI/ML is deployed. AI/ML models that can increase access to finance can also perpetuate and exacerbate bias against certain segments of the population. Digital targeted marketing can be used for aggressive marketing of predatory products to vulnerable consumers that exploits behavioral biases. GenAI may allow for quicker responses to customer service

[1] Founded in 1936, Consumer Reports (CR) is an independent, nonprofit and nonpartisan organization that works with consumers to create a fair and just marketplace. Known for its rigorous testing and ratings of products, CR advocates for laws and company practices that put consumers first. CR is dedicated to amplifying the voices of consumers to promote safety, digital rights, financial fairness, and sustainability. The organization surveys millions of Americans every year, reports extensively on the challenges and opportunities for today's consumers, and provides ad-free content and tools to 6 million members across the U.S.

queries, but may also result in inaccurate responses or prevent consumers from reaching live agents to resolve urgent matters.

It is therefore critical to have clear and strong safeguards in place to mitigate risks to consumers, so that AI/ML is deployed in a safe and responsible manner that ensures that consumers can reap the benefits, rather than facing the harms that AI/ML can cause. CR is working to support responsible innovation in AI that can expand access to financial services and improve consumer outcomes, particularly for marginalized communities, while simultaneously identifying and promoting safeguards and requirements that need to be in place to mitigate new and enhanced risks from AI/ML. Earlier this year, we published a general set of our AI policy recommendations.[2]

For any framework to effectively safeguard consumers from the risks that arise from AI/ML, multiple dimensions must be covered. Strong risk management frameworks and governance systems are key components, as are internal and external audits and impact assessments. However, these are only one dimension of the puzzle. Affirmative obligations must also be placed on developers and deployers, for example to take proactive measures throughout the AI/ML model development pipeline to mitigate the potential for discrimination in resulting models and to meet high standards of accuracy when employing consumer-facing AI applications such as GenAI chatbots and virtual assistants. In addition, these obligations should be complemented by individual rights to consumers, including regarding disclosure of AI usage and the right to contest AI-driven decisions.

In the sections below, we provide responses to Questions 7, 9, 10, 14, and 18 from the RFI.

<u>Oversight of AI – Explainability and Bias</u>

<u>Question 7: What challenges exist for addressing risks related to AI explainability? What methodologies are being deployed to enhance explainability and protect against potential bias risk?</u>

The lack of transparency into the inner workings of complex AI/ML models is a well-known issue. The opaque "black box" nature of more complex ML models (such as neural networks) can make it difficult for financial institutions themselves to explain a model's decision process. For example, a 2021 study found that 65% of respondent companies were not able to explain how specific AI decisions or predictions were made.[3]

Without transparency into an AI model, even in instances where a financial institution recognizes that their model produces unintended consequences or leads to poorer performance for certain segments of the population, it can be challenging to identify the sources of bias and address them. In addition, without transparency into what factors impact pricing or approval outcomes, financial institutions cannot comply with adverse action notice requirements that require disclosure of the key factors that led to an adverse action with respect to credit decisions.

As a result, the risk of algorithmic discrimination may be enhanced. Consumers also face barriers in understanding the factors behind a negative decision, whether due to errors, unfair practices, or

[2] Gedye, Grace and Justin Brookman. Consumer Protection Policies for the AI Era. Consumer Reports, March 2024. https://docs.google.com/document/d/1SobfUSSvaFNWMGxH2uU4c6tWLReiI0A10UrpRdxUYYY/edit
[3] The State of Responsible AI: 2021. FICO, 2021. https://www.fico.com/de/latest-thinking/market-research/state-responsible-ai-2021

legitimate factors, and will lack the knowledge on how they can change their behavior to improve their outcome or eligibility in the future. This leaves consumers with limited ability to hold financial institutions accountable for unjustified results or discriminatory and illegal practices and limits their tools to appeal decisions from an AI/ML model.

In order to be able to effectively assess for discrimination, particularly for variables that are functioning as proxies for protected characteristics, and provide clear, specific, and accurate reasons in adverse action notices, it will be critical for financial institutions to ensure they have sufficient transparency into AI/ML models. While the realm of explainable AI (XAI) is a rapidly evolving space, two main approaches that been put forth to enhance transparency are (1) *post hoc* explainability and (2) inherent interpretability.

Post hoc explainability involve applying techniques on top of "black box" models to provide visibility into how an AI/ML model works and reaches its decisions. There are a range of *post hoc* explainability techniques, but two common techniques are SHAP values and LIME. Explainability techniques such as SHAP (Shapley Additive Explanations)[4] values are used to explain which features or input variables have the biggest impact on a model's conclusions. SHAP can be used for both local explanations (i.e. particular inputs and outputs) as well as global explanations (i.e. how the model works overall). Techniques like LIME (Local Interpretable Model-Agnostic Explanations) essentially leverage simpler models designed to mimic the original, more complex model. They can provide a local explanation around a particular input and output. They are trained on the more complex model's predictions, such as a linear surrogate model around a particular data point to approximate a complex model's output, and can identify what features led to that particular prediction.[5]

While *post hoc* explainability techniques are popular and useful, they have certain limitations. In particular, there are concerns that these techniques may not be sufficiently robust. For example, they have difficulty capturing latent variables (i.e. variables that are not directly observed in the data but are inferred) or interaction effects between variables within complex ML models, both of which can play a significant role in a model's decision-making process. It is therefore unclear whether employing current *post hoc* explainability techniques provides sufficient specificity or accuracy for the purposes of adverse action notices, as indirectly alluded to in a 2022 CFPB circular.[6] In addition, these techniques do not enable financial institutions to fine-tune models to address any issues identified, nor do they provide understandable explanations for internal or external stakeholders with less technical expertise.[7]

---

[4] SHAP omits individual input features over multiple iterations and analyzes the resulting changes in model performance to generate a cumulative measure of a feature's relative importance to the model's prediction and how much a specific feature contributed to changing an outcome.

[5] For further details on transparency tools and techniques as well as their limitations, see Explainability & Fairness in Machine Lending for Credit Underwriting: Policy Analysis. FinRegLab, November 2023. https://finreglab.org/wp-content/uploads/2023/12/FinRegLab_2023-12-07_Research-Report_Explainability-and-Fairness-in-Machine-Learning-for-Credit-Undewriting_Policy-Analysis.pdf

[6] Footnote 1 of the CFPB circular states that, *"While some creditors may rely upon various post-hoc explanation methods, such explanations approximate models and creditors must still be able to validate the accuracy of those approximations, which may not be possible with less interpretable models."* See CFPB Circular 2022-03. Adverse action notification requirements in connection with credit decisions based on complex algorithms. CFPB, May 2022. https://www.consumerfinance.gov/compliance/circulars/circular-2022-03-adverse-action-notification-requirements-in-connection-with-credit-decisions-based-on-complex-algorithms/

[7] For further discussion on the pros and cons of *post hoc* explainability versus inherent interpretability, see Transparency, Explainability, and Interpretability in AI/ML Credit Underwriting Models. Consumer Reports'

By contrast, interpretability is a very different approach to transparency that involves building a model that is inherently interpretable and transparent from the start. Essentially, as ML processes are used to analyze different variables and their interactions and predictive power, these learnings are made visible to users. As a result, the resulting model and how it reaches its predictions can be easily understood by a range of stakeholders, models can be fine-tuned with greater ease and specificity, and clear, accurate, and more actionable adverse action notices can be provided to consumers. However, the main drawback that is raised with respect to inherently interpretable approaches is that such approaches necessarily limit how complex a ML model can be, which may in turn limit a model's performance capabilities.

Greater clarity as well as consistency from regulators on transparency requirements and expectations would be beneficial for all stakeholders, including both industry and consumers. There is still much ongoing debate regarding which transparency techniques work best, as well as limited regulatory guidance on these issues. There are currently no uniform definitions or benchmarks for determining what level of information or transparency is sufficient for what particular regulatory purpose or audience.

One potential approach to the issue of transparency would be to consider the different reasons and contexts for which transparency may be needed and follow a proportionate approach that is use case specific. For example, for providing general explanations regarding how a model works (such as with respect to AI/ML models for marketing or fraud monitoring purposes), *post hoc* explainability techniques may be sufficient, particularly as such techniques continue to advance and improve. Reliance on such techniques could be complemented (and compensated) with stronger governance measures such as enhanced human oversight, stronger data governance measures, and more robust searches for less discriminatory alternative (LDA) models (further discussed under Question 10). However, if transparency is needed in order to be able to pinpoint the specific factors that led to a consequential decision impacting the rights of a consumer, then *post hoc* explainability techniques may be insufficient. In which case, there is a strong argument that inherently interpretable approaches should be used.

Clear policy guidance would be particularly useful on complex questions such as the level of accuracy that *post hoc* explainability techniques should be required to meet, in what contexts and use cases minimum interpretability requirements may be required, and appropriate justifications when using less interpretable approaches.

CR's position is that if a tool is so complex that the company using it cannot provide specific, accurate, clear, and actionable explanations for the outputs it generates, it should not be used in consequential decisions, including financial decisions.[8] For example, the European Insurance and Occupational Pensions Authority (EIOPA) has stated that insurance firms should endeavor to use as much as possible interpretable AI models, particularly where the AI use case has a significant impact on consumers, such as the use of AI in pricing and underwriting of insurance. EIOPA specifically notes that a high level of transparency is necessary in order to achieve fairness for consumers and existing state-of-the-art *post*

Innovation Blog, March 2024. https://innovation.consumerreports.org/transparency-explainability-and-interpretability-in-ai-ml-credit-underwriting-models/

[8] Gedye, Grace and Justin Brookman. Consumer Protection Policies for the AI Era. Consumer Reports, March 2024. https://docs.google.com/document/d/1SobfUSSvaFNWMGxH2uU4c6tWLReiI0A10UrpRdxUYYY/edit

hoc explainability techniques can only provide rough explanations when it comes to "black box" systems analyzing non-traditional factors, such as use of telematics in auto insurance pricing and underwriting.[9]

<u>Fair Lending, Data Privacy, Fraud, Illicit Finance, and Insurance</u>

<u>Question 9: How are financial institutions evaluating and addressing any increase in risks and harms to impacted entities in using emerging AI technologies? What are the specific risks to consumers and other stakeholder groups, including low- to moderate-income consumers and/or underserved individuals and communities (e.g., communities of color, women, rural, tribal, or disadvantaged communities)? How are financial institutions protecting against issues such as dark patterns – user interface designs that can potentially manipulate impacted entities in decision-making – and predatory targeting emerging in the design of AI? Please describe specific risks and provide examples with supporting data.</u>

Use of emerging AI technologies introduces a range of new and enhanced risks to consumers. In this section, we focus on three key risks: (1) algorithmic discrimination, (2) lack of external transparency and due process rights, and (3) risks from use of GenAI in customer service.

Algorithmic discrimination across the customer lifecycle

In recent years, the adoption of algorithmic decision-making tools by financial institutions, particularly more complex machine learning (ML) models, has surged. While these advancements have the potential to enhance efficiency and advance financial inclusion, there is growing evidence that they can also perpetuate and exacerbate existing and historical biases, leading to discriminatory outcomes that adversely affect marginalized and underserved communities. The risk of algorithmic systems resulting in biased outcomes that perpetuate and even exacerbate existing societal biases has been well-established in a wide range of research across multiple sectors.[10] Algorithmic discrimination occurs when an algorithmic decision system repeatedly creates unfair or inaccurate outcomes for a protected class. **These biases in AI/ML models have been shown to result in incorrect, inaccurate, or biased decisions for certain groups, leading to real harm to consumers due to financial exclusion or unfair pricing, among other harms.**

Discriminatory results can arise from multiple sources. Discrimination can arise from incorrect, incomplete, or unrepresentative training data, as well as from biased data collection and processing methods. For example, a 2021 study from Stanford University and the University of Chicago found that one of the underlying reasons for differences in mortgage approval rates between minority and majority groups was the limited credit history data on mortgages for minorities and low-income groups, leading to less precise predictions for such groups.[11]

[9] Note that this report uses the term "explainability" similar to how the term interpretability is used in this letter. See Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. EIOPA, June 2021. https://www.eiopa.europa.eu/system/files/2021-06/eiopa-ai-governance-principles-june-2021.pdf

[10] For example, see Barocas, Solon and Andrew D. Selbst. "Big Data's Disparate Impact." 104 California Law Review 671, 2016, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899; O'Neil, Cathy. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Penguin Books, 2016; and Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine Bias." ProPublica, 2016, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[11] Blattner, Laura and Scott Nelson. "How Costly is Noise? Data and Disparities in Consumer Credit." Papers 2105.07554, arXiv.org, 2021. https://arxiv.org/pdf/2105.07554.pdf

Data may also reflect historical biases, particularly the types of data sources used for underwriting, which are tainted by past discriminatory practices. Due to past policies that institutionalized systemic racism, studies have found that Black and Latinx Americans are more likely to have damaged credit or lower credit scores compared to their white counterparts[12] and are more likely to be sold high-cost, unmanageable loans.[13] AI/ML credit underwriting models that rely on this historical data can perpetuate these disparities.

Biases can also be embedded into models through the design process, such as via use of proxies for protected characteristics. Complex ML models utilizing hundreds or thousands of input features may inadvertently use proxies for race or ethnicity or other protected characteristics, leading to discriminatory outcomes even without direct use of protected class data. For example, a 2022 NYU study found that popular mortgage pricing ML algorithms consider factors that correlate with race, such as nearby college education levels, leading to inflated interest rates for minorities.[14]

It is worth noting that traditional systems (algorithmic or not) can and do lead to discriminatory results as well. AI/ML models do not necessarily cause more or less discrimination as compared to traditional models. However, they are not as objective and impartial as they may appear to be, given that these systems are designed by humans and heavily impacted by the data and processes used to trained them. And unlike traditional algorithmic systems, AI/ML models can be deployed at scale in a more cost-effective manner, potentially impacting more consumers; be used for automated decisions, resulting in discriminatory results going unchecked; and lack transparency, making it harder to identify and correct discrimination. The potential for automated systems to produce biased outcomes and "automate discrimination" has already been recognized by the CFPB and other civil rights agencies in an April 2023 joint statement, which also highlighted a commitment to enforce respective laws and regulations to address these very issues arising from the use of new technologies.[15] **Importantly, there is the potential for AI/ML models to actually be less discriminatory than current approaches, but only if clear and robust standards are established to ensure the deployment of less discriminatory AI/ML models.**

AI/ML tools are being utilized for multiple purposes in retail banking, raising the risk of algorithmic discrimination across the customer lifecycle. The application that has received the most attention to date is credit underwriting and pricing. Financial institutions, particularly fintech lenders, are increasingly leveraging AI/ML to augment or even replace traditional underwriting with automated methods that often rely on alternative data.

---

[12] "Past Imperfect: How Credit Scores and Other Analytics 'Bake In' and Perpetuate Past Discrimination." National Consumer Law Center, May 2016. https://www.nclc.org/resources/past-imperfect-how-credit-scores-and-other-analyticsbake-in-and-perpetuate-past-discrimination/.

[13] Marte, Jonnelle. "Wells Fargo steered blacks and Latinos toward costlier mortgages, Philadelphia lawsuit alleges." Los Angeles Times, May 16, 2017. https://www.latimes.com/business/la-fi-wells-fargo-philadelphia-20170516-story.html

[14] Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A. (2022). Predictably Unequal? The Effects of Machine Learning on Credit Markets. The Journal of Finance, 77: 5-47. https://doi.org/10.1111/jofi.13090

[15] Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems. CFPB, Department of Justice, Equal Employment Opportunity Commission, and Federal Trade Commission, April 2023. https://files.consumerfinance.gov/f/documents/cfpb_joint-statement-enforcement-against-discrimination-bias-automated-systems_2023-04.pdf

But AI/ML tools are also being increasingly utilized for other stages of the customer lifecycle as well, including for targeted marketing, fraud monitoring, automated valuation models for mortgages, etc. Instances of harm to consumers arising from algorithmic discrimination has already been identified for these applications. For example, biased algorithms used for targeted advertising of financial products can result in certain groups being unfairly excluded from better products and services ("digital redlining"), or conversely being targeted for offers for inferior products ("steering").[16] With respect to fraud monitoring, there is an ongoing lawsuit against State Farm claiming that its fraud detection software has a disparate impact on Black customers, requiring additional documentation and delaying claims processing for Black customers compared to White customers.[17] There have also been increasing reports of consumers locked out of their bank accounts due to false positives in fraud monitoring and facing a substantial difficulties in regaining access to their accounts.

Policymakers should ensure that measures and safeguards to address algorithmic discrimination (such as those discussed under Question 10) apply for all other relevant AI/ML use cases that could negatively impact financial consumers. The risks of consumer harm from algorithmic discrimination during other stages of the customer lifecycle have received less attention to date from industry or from policymakers.

Lack of external transparency and due process rights

The risks to consumers arising from the use of AI are exacerbated when consumers are not even made aware that they have been subject to a consequential decision wholly or partly based on an AI model or given any rights to contest such decisions. Consumers deserve due process rights when AI is being used to make consequential decisions that affect them, including notice regarding the use of AI, adverse action notices, and the right of appeal. Several of these rights are reflected in the White House's Blueprint for an AI Bill of Rights as well as the subsequent memorandum from the Office of Management and Budget (OMB) on agency use of AI.[18] This combination of disclosure and due process rights provide a key safeguard (when in combination with other safeguards) for consumers to be in a position to identify where AI-driven decisions have negatively impacted them and to try to address the situation. It also incentivizes greater accountability on the part of financial institutions deploying such models.

At a minimum, CR advocates for consumers to be provided with clear disclosure when an AI tool is being used to help make a consequential decision about them, such as whether they qualify for a

---

[16] See "Is Your Marketing Biased? Financial Regulators Are About To Start Checking." Forbes, 2022. https://www.forbes.com/sites/kareemsaleh/2022/12/07/is-your-marketing-biased-financial-regulators-are-about-to-start-checking/. See also Lambrecht, A. and Tucker, C. (2019). "Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads." Management Science, 65(7), 2976-2991. This study investigates how advertising algorithms can result in gender discrimination. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2852260

[17] https://www.nytimes.com/2022/12/14/business/state-farm-racial-bias-lawsuit.html

[18] See Blueprint for an AI Bill of Rights. The White House, October 2022. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf. See also Memorandum on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. Office of Management and Budget, March 2024. https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

loan. It is critical that consumers be made aware of the use of AI in order to be able to take further action if needed. For example, the aforementioned OMB memo states that, *"Where people interact with a service relying on the AI and are likely to be impacted by the AI, agencies must also provide reasonable and timely notice about the use of the AI and a means to directly access any public documentation about it in the use case inventory."*[19] The 2023 EU Consumer Credit Directive requires creditors to inform consumers in a clear and comprehensible manner when they are presented with a personalized offer that is based on automated processing of personal data.[20] The recently passed Colorado AI Act includes provisions that require deployers of high-risk AI systems to notify consumers that they have deployed such systems before a consequential decision is made and provide information to the consumer regarding the purpose of the system and the nature of the consequential decision.[21]

Consumers should also be provided with clear explanations when they receive a consequential adverse decision. These explanations should provide specific and accurate information on the key factors and the underlying data that resulted in adverse decisions. This type of explanation will be critical to enable consumers to identify and take steps to correct any inaccurate information that may have contributed to an adverse decision. It will also enable consumers to better understand how they can change their behavior to achieve a better outcome in the future, rather than feeling powerless in the face of such decisions, while also placing greater onus on financial institutions to ensure decisions are not discriminatory or inaccurate. For example, the Colorado AI Act requires deployers of high-risk AI systems to provide to consumers, if such systems are used to make a consequential decision adverse to the consumer, a statement disclosing the principal reason or reasons for the consequential decision, including the type and source of data processed in making the decision.[22]

While adverse action notices are already required for credit products, such notices could be improved to be made more clear and actionable for consumers. The CFPB has taken some actions to date on adverse action notices, including hosting a "tech sprint" in 2020 on innovative ways to notify consumers of adverse credit decisions[23] and issuing circulars clarifying that adverse action notices must include specific and accurate reasons for adverse actions regardless of the underlying technology used to reach such decisions[24] and should not rely on a checklist of reasons in sample forms or overly broad or vague reasons.[25] However, further clarifications would be beneficial to ensure that, particularly in the context of AI/ML-driven decisions, information is provided to consumers in a manner that balances sufficient granularity with usefulness and relevance to consumers, how factors contribute to adverse decisions is clearly explained in plain language, and information is actionable for consumers.

[19] Ibid.

[20] Art. 13, Directive (EU) 2023/2225 on credit agreements for consumers. European Parliament and the Council of the European Union, October 2023.

[21] Art. 6-1-1703(4), Colorado AI Act (SB-205)

[22] Art. 6-1-1703(4), Colorado AI Act (SB-205)

[23] https://www.consumerfinance.gov/rules-policy/competition-innovation/cfpb-tech-sprints/electronic-disclosures-tech-sprint/

[24] Circular 2022-03 on Adverse action notification requirements in connection with credit decisions based on complex algorithms. CFPB, May 2022. https://www.consumerfinance.gov/compliance/circulars/circular-2022-03-adverse-action-notification-requirements-in-connection-with-credit-decisions-based-on-complex-algorithms/

[25] Circular 2023-03 on Adverse action notification requirements and the proper use of the CFPB's sample forms provided in Regulation B. CFPB, September 2023.
https://www.consumerfinance.gov/compliance/circulars/circular-2023-03-adverse-action-notification-requirements-and-the-proper-use-of-the-cfpbs-sample-forms-provided-in-regulation-b/

As it is, adverse action notices can be difficult for consumers to understand or to act upon; these challenges are only heightened in the context of AI/ML-driven decisions. It will be critical to consider how best to present such information to consumers to be practically useful. Ideally, adverse action notices should include action-oriented language on what specific steps a consumer can take to achieve a better result. For example, the Monetary Authority of Singapore (MAS)'s Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector[26] highlight transparency as a core principle and include accompanying materials on how to assess key principles. The assessment methodology on transparency notes that consumers/data subjects ideally require a combination of general explanations for the decision along with information on actions they can take to change a model's behavior, particularly highlighting the benefit of providing counterfactuals to demonstrate how a decision could be improved by a change in the consumer's behavior.[27] Financial institutions should ensure they have the answers to the following questions in their explanations:

- How the decision was made;
- What were the top reasons behind the decision (both positive and negative factors);
- What actions could have enabled a more favorable outcome for the consumer;
- How did the AIS decision impact the consumer; and
- What redress options are available to the consumer.

In addition, adverse action notice requirements should be extended beyond just credit products and applied more broadly and consistently to consequential decisions impacting financial consumers that are driven by AI. For example, the New York Department of Financial Services' (NYDFS) recent circular to insurers on the use of AI systems and external consumer data and information sources in insurance underwriting and pricing requires insurers provide notice to consumers disclosing details about all information upon which the insurer based the adverse decision, including the source of the specific information relied on. The NYDFS circular further notes that failure to adequately disclose specific reasons for adverse actions may be deemed an unfair or deceptive practice.[28]

Lastly, consumers should have the right to appeal an AI-driven decision for human review. The Colorado AI Act clearly establishes the right for consumers to appeal adverse consequential decisions arising from the deployment of high-risk AI systems for human review.[29] Similarly, the EU General Data Protection Regulation (GDPR) includes similar rights for consumers who have been subject to automated decisions to obtain human intervention and contest the decision.[30] To provide consumers with a genuine, meaningful right to appeal that effectively serves as a check on inaccuracies and discrimination,

[26] Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector. Monetary Authority of Singapore, 2018. https://www.mas.gov.sg/-/media/mas/news-and-publications/monographs-and-information-papers/feat-principles-updated-7-feb-19.pdf

[27] FEAT Transparency Principles Assessment Methodology. Veritas, 2022. https://www.mas.gov.sg/-/media/MAS-Media-Library/news/media-releases/2022/Veritas-Document-3C---FEAT-Transparency-Principles-Assessment-Methodology.pdf

[28] Section IV(E), Art. 38-39, Insurance Circular Letter No. 7 on Use of Artificial Intelligence Systems and External Consumer Data and Information Sources in Insurance Underwriting and Pricing. NYDFS, July 2024. https://www.dfs.ny.gov/industry-guidance/circular-letters/cl2024-07

[29] Art. 6-1-1703(4), Colorado AI Act (SB-205)

[30] Article 22, GDPR

it will be necessary to ensure that the procedures and timeframes for exercising this right to appeal are clearly established and relatively easily accessible for consumers.[31]

GenAI-enabled customer service

The last area of potential but increasing concern is generative AI (GenAI), i.e. a specific subset of AI/ML technologies with the ability to generate new content, including text, images, and sound. GenAI has rapidly captured the attention of the entire ecosystem since the introduction of ChatGPT, particularly large language models (LLMs), a type of GenAI. LLMs are neural network-based models trained on massive amounts of unstructured data. These models learn to predict the probability of the next word in a desired output response, resulting in the capability for LLMs to produce understandable and meaningful text based on a prompt. Because foundation models such as ChatGPT are trained on such massive datasets across the entire internet, they can generate content on a wide range of topics and hence be used for a variety of use cases (though they also raise risks of embedded bias).

Financial institutions are currently in the process of exploring a wide variety of GenAI use cases, including both customer-facing uses as well as for enhancing internal business operations, while also determining how to utilize GenAI safely. The most evident public-facing use of GenAI is powering customer service chatbots and virtual assistants. In recent years, a small but growing number of major banks as well as fintechs have launched customer service applications such as digital chatbots and virtual assistants that combine a range of tools and technologies, including natural language processing (NLP), text-generating technologies using LLMs, voice analytics, predictive analytics and recommendation algorithms, and access to account data. Recent estimates have indicated that 11% of banks are offering 1st generation AI-assisted virtual assistants, while 28% of bank chatbots offer advanced capabilities like NLP.[32] Klarna, a Swedish-based buy now, pay later (BNPL) company, indicated that its OpenAI-powered virtual assistant was already handling two-thirds of all customer service chats (equaling 2.3 million conversations) one month after its global launch.[33]

While traditional chatbots were rule-based and could answer simple questions, virtual assistants engage more dynamically with consumers and provide a much wider array of functionality to support customers in their everyday financial lives. Virtual assistants can understand more complex user queries; handle specific requests such as transferring money, paying bills, or pulling up transaction histories or account balances; send automated alerts and predictive notifications regarding cash shortfalls or duplicate transactions; and track spending and provide personalized insights. As virtual assistants become even more sophisticated, it is anticipated that they will be able to provide personalized financial advice and help customers make higher-value financial decisions.

As mentioned previously, AI often represents a double-edged sword for consumer welfare, and GenAI is no exception. Digital chatbots could enable consumers to take simple actions and resolve simple queries much faster than with current customer service options. Virtual assistants could in theory provide real benefits to consumers if designed and deployed in a safe and responsible manner. They could serve as a

[31] For further discussion on the right to contest AI-driven decisions might best be structured, see Kaminski, Margot E. and Jennifer M. Urban. The Right to Contest AI. Columbia Law Review Vol. 121, No. 7, November 2021. https://columbialawreview.org/wp-content/uploads/2021/11/Kaminski-Urban-The_Right_to_Contest_AI.pdf

[32] https://www.americanbanker.com/opinion/virtual-assistants-are-the-future-but-some-banks-are-falling-behind

[33] https://www.fastcompany.com/91039401/klarna-ai-virtual-assistant-does-the-work-of-700-humans-after-layoffs

virtual financial advisor for consumers who may never be able to access a personal financial advisor otherwise, helping vulnerable consumers to better manage their financial lives and financial well-being.

However, if deployed irresponsibly, digital chatbots and virtual assistants could pose real risks to consumers. We highlight below a few key risks that could arise.[34]

On a practical level, consumers may run into issues with not being able to resolve urgent matters, due to a combination of limited functionality of chatbots as well as barriers to reaching live assistance. In worst case scenarios, digital chatbots and virtual assistants may be used intentionally to stymie consumers' ability to resolve issues or avoid accountability. In a recent nationally representative survey conducted by CR, we found that consumers interacting with digital chatbots were three times more likely to say they did not get the help they were looking for compared to consumers who interacted with a live representative (58% vs 21%). In addition, nearly half of consumers (47%) indicated they had difficulty reaching a live representative.[35]

There is also the risk that consumers may be given inaccurate or unreliable information, or be given inappropriate advice or offered inappropriate products. LLMs produce outputs based on probability, not predictive analytics or logical reasoning. They are known for producing hallucinations, factually inaccurate statements that may be presented in a very plausible sounding manner, particularly when faced with complex questions or situations. Inaccurate information could potentially lead consumers to make poor financial decisions that harm their well-being.

An additional concern is that virtual sales assistants could be used to actively manipulate consumer behavior, particularly if trained solely (or primarily) to optimize profit without consideration for consumer welfare, leading consumers to purchase inappropriate products or services. For example, multiple fintechs have announced collaborations with OpenAI to install plug-ins that will enable users to ask for advice and product recommendations (i.e. a virtual assistant for shopping). More broadly, AI/ML models are being used to develop detailed customer profiles and send targeted digital ads and offers tailored for particular customers. Digital targeted marketing raises potential risks of exploitation of behavioral biases, nudging consumers into purchasing inappropriate, inferior, or higher priced products or services they otherwise would not have. Conversely, such models also raise risks of digital redlining, where certain demographic groups are excluded from marketing of products.[36]

GenAI use cases in the financial sector are still evolving, and many financial institutions appear to be proceeding cautiously given the known risks associated with GenAI and LLMs, including those noted above. However, there is clear pressure for financial institutions to figure out how to leverage GenAI to

---

[34] Discussion on additional risks potentially arising from digital chatbots can be found in *Chatbots in consumer finance*. CFPB, June 2023. https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/

[35] Consumer Reports nationally representative American Experiences Survey of 2,035 U.S. adults (February 2024). https://article.images.consumerreports.org/image/upload/v1710449643/prod/content/dam/surveys/Consumer_Reports_AES_February_2024.pdf

[36] For example, see Evans, Carol and Westra Miller. From Catalogs to Clicks: The Fair Lending Implications of Targeted, Internet Marketing. Federal Reserve Board of Governors, 2019. https://www.consumercomplianceoutlook.org/2019/third-issue/from-catalogs-to-clicks-the-fair-lending-implications-of-targeted-internet-marketing/

increase efficiency and competitiveness. Appropriate solutions to address the risks that GenAI poses are still nascent.

For digital chatbots and virtual assistants, emerging good practices include providing clear disclaimers when consumers are interacting with a digital chatbot or virtual assistant and offering an easily accessible means to escalate to human assistance. Financial institutions should also provide clear information regarding the capabilities and limitations of a digital chatbot or virtual assistant.[37]

In addition, financial institutions that utilize consumer-facing AI systems such as digital chatbots and virtual assistants that respond to customer inquiries and execute transactions should be held to high consumer protection standards, particularly regarding accuracy.[38] Employing GenAI tools to interact directly with consumers should not absolve financial institutions of liability arising from the outputs of such tools, and consumers' reliance on such outputs. Application developers should take proactive measures to improve accuracy and decrease hallucinations, while promoting safety and security. This includes prompt engineering, routing, and model fine-tuning to adapt LLM calls for specific tasks, retrieval augmented generation (RAG) to inform responses with relevant context, input guardrails to prevent problematic questions, output guardrails to prevent problematic responses, and rigorous, repeatable evaluation methods to enable test-driven development process.[39]

With respect to virtual sales assistants, and more generally the risks to consumers that arise from digital targeted marketing, more policy efforts are needed to determine and establish appropriate safeguards. These safeguards should include restrictions on data harvesting, easily accessible opt-outs for cross-marketing and targeted advertising, and clearer guidance on when digital targeted marketing tactics cross the line into becoming unfair, deceptive, or abusive acts or practices (UDAAP) violations.

<u>Question 10: How are financial institutions addressing any increase in fair lending and other consumer-related risks, including identifying and addressing possible discrimination, related to the use of AI, particularly emerging AI technologies? What governance approaches throughout the development, validation, implementation, and deployment phases do financial institutions expect to establish to ensure compliance with fair lending and other consumer-related laws for AI models and tools prior to deployment and application?</u>

As discussed under Question 9, the use of AI/ML raises risks of algorithmic discrimination which can arise from multiple sources. **There are a range of approaches that financial institutions can and should employ throughout the model development pipeline (as opposed to after the fact, or only during the**

---

[37] For example, these practices are highlighted in Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. EIOPA, June 2021. https://www.eiopa.europa.eu/system/files/2021-06/eiopa-ai-governance-principles-june-2021.pdf

[38] For example, the Center for American Progress calls for the CFPB to leverage its unfair, deceptive, or abusive acts or practices (UDAAP) authority to ensure that financial institutions' consumer-facing AI systems are accurate in all respects and to require, through rulemaking, periodic review of their systems to ensure accuracy. See Taking Further Agency Action on AI: Financial Regulatory Agencies. Center for American Progress, June 2024. https://www.americanprogress.org/article/taking-further-agency-action-on-ai/financial-regulatory-agencies-chapter/

[39] For further discussion on responsible approaches to GenAI development based on CR's own experimentation, see CR's Innovation Blog at https://innovation.consumerreports.org/blog/?theme=responsible-tech&author=&daterange=&search=&_nonce=15503a3010

modeling stage) in order to directly address these sources of algorithmic discrimination and effectively mitigate the risk of algorithmic discrimination.

To begin with, financial institutions should take active measures during the data collection stage to avoid employing biased data collection methods that can lead to biased datasets and ensure that training data is representative and complete for different demographic groups. Financial institutions should also be cognizant of the choices they make during data pre-processing, as emerging research has shown that choices made regarding data imputation strategies and data encoding can lead to more or less bias in resulting AI/ML models.[40] Financial institutions should also test for close proxies for protected characteristics and exclude such variables. In addition, clearer guidance would be beneficial on use of variables that are highly likely to embed disparate impact for protected groups, for example standardized test scores. Post-deployment, financial institutions should continue to actively monitor model performance to identify and address emerging issues of discrimination.

During the modeling and training stage, a particularly powerful solution to address algorithmic discrimination would be ensuring that financial institutions conduct a robust search for and implement less discriminatory alternative (LDA) models.[41] Advanced tools and techniques are emerging that enable fine-tuning and debiasing AI/ML models during the development stage to mitigate disparities. Techniques such as adversarial debiasing, joint optimization,[42] or optimized searches for different combinations of variables now enable developers to explore a wide range of alternative models in a much more rapid, efficient manner than was previously feasible. It is now possible to identify alternative models that maintain similar performance levels while minimizing disparity, a win/win for both financial institutions as well as consumers.

Disparate impact involves facially neutral policies that have a disproportionately adverse effect on a protected class, regardless of intent. Disparate impact doctrine has been clearly established in case law, regulatory guidance, and enforcement actions. Disparate impact doctrine typically involves a three-step test: (1) a plaintiff must establish a prima facie case showing that a policy or practice has a disparate impact on a disadvantaged group; (2) the defendant then has the burden of demonstrating a legitimate business justification for the practice; and (3) even where defendants provide a business justification, they can still face liability if there is an alternative approach that serves the same business needs with

[40] For further discussion on these issues, see Black, Emily, Rakshit Naidu, Rayid Ghani, Kit T. Rodolfa, Daniel E. Ho, and Hoda Heidari. Toward Operationalizing Pipeline-aware ML Fairness: A Research Agenda for Developing Practical Guidelines and Tools. Sept 2023. https://arxiv.org/pdf/2309.17337

[41] The content in this section draws heavily from a joint letter from Consumer Reports and the Consumer Federation of American sent to the CFPB in June 2024 regarding LDAs. For further details, see https://advocacy.consumerreports.org/wp-content/uploads/2024/06/240626-CR-CFA-Statement-on-Less-Discriminatory-Algorithms-FINAL.pdf

[42] Adversarial debiasing involves using a second "adversary" model to predict protected class status based on a primary model's predictions, and to make changes to the primary model through an iterative process in order to gradually minimize the adversary model's ability to predict protected class characteristics, eventually resulting in a primary model that is less correlated with protected characteristics. Joint optimization involves developing a model that simultaneously optimizes two objectives, such as predictive accuracy and minimizing disparities. For further details, see Explainability & Fairness in Machine Lending for Credit Underwriting: Policy Analysis. FinRegLab, December 2023. https://finreglab.org/wp-content/uploads/2023/12/FinRegLab_2023-12-07_Research-Report_Explainability-and-Fairness-in-Machine-Learning-for-Credit-Undewriting_Policy-Analysis.pdf

less disparate impact. Multiple regulators, including the CFPB, have affirmed the obligation to search for LDAs as a core element of disparate impact doctrine.[43]

Demonstrating the existence of an LDA can be challenging for plaintiffs in practice and is even more difficult in the context of complex, proprietary "black box" ML models. However, due to the unique capacity of ML models to improve through rapid iteration, it is now much less burdensome or time-intensive to search for and implement LDAs. For example, model multiplicity refers to a phenomenon identified in recent computer science and statistics research that shows that there are multiple possible models that are equally effective at a given task. *"As a result, when an algorithmic system displays a disparate impact, model multiplicity suggests that other models exist that perform equally well but have less discriminatory effects. In other words, in almost all cases, a less discriminatory algorithm exists."*[44] Model multiplicity implies that there no longer needs to be a significant tradeoff between performance and fairness.

Therefore, particularly in the context of ML models, it is no longer a question of whether or not an LDA that meets legitimate business needs can be found – this threshold question has already been answered in the affirmative. For example, the Equal Employment Opportunity Commission (EEOC) has noted that "one advantage of algorithmic decision-making tools is that the process of developing the tool may itself produce a variety of comparably effective alternative algorithms. Failure to adopt a less discriminatory alternative that was considered during the development process may give rise to liability."[45] Instead, the question now turns to how financial institutions should go about finding and implementing LDAs, particularly in the case of ML models.

Requiring companies to proactively mitigate disparate impact by searching for and implementing LDAs is a natural and logical evolution of existing disparate impact doctrine and has significant potential to achieve anti-discriminatory policy objectives more effectively. Particularly when it comes to ML models, the development process already involves weighing a series of choices and making continual refinements to optimize the performance of the resulting model, selected from a universe of potential models. Incorporating disparate impact as an additional lens to apply could be easily integrated into the

[43] For example, the U.S. Department of Housing and Urban Development (HUD)'s discriminatory effects rule codified long-standing case law on discriminatory effects doctrine under the Fair Housing Act (FHA), including noting that plaintiffs can still prevail in cases where defendants demonstrate a legitimate purpose for a challenged practice if they can show that this purpose "could be served by another practice that has a less discriminatory effect." See 24 Code of Federal Regulations § 100.500 Discriminatory effect prohibited. Similarly, the CFPB's official interpretation of Regulation B incorporates the concept of LDAs, stating that a creditor's practice may be prohibited if it is "discriminatory in effect because it has a disproportionately negative impact on a prohibited basis, even though the creditor has no intent to discriminate and the practice appears neutral on its face unless the creditor practice meets a legitimate business need that cannot reasonably be achieved as well by means that are less disparate in their impact." See Comment for 1002.6 - Rules Concerning Evaluation of Applications in Supplement I to Part 1002 - Official Interpretations. CFPB. https://www.consumerfinance.gov/rules-policy/regulations/1002/interp-6/

[44] Black, Emily, John Logan Koepke, Pauline T. Kim, Solon Barocas, and Mingwei Hsu. "Less Discriminatory Algorithms (October 2, 2023)." Georgetown Law Journal, Vol. 113, No. 1, 2024, Washington University in St. Louis Legal Studies Research Paper Forthcoming. https://ssrn.com/abstract=4590481

[45] "Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964." EEOC, May 18, 2023. https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial

typical model development process, leading to the selection of a model that will advance legitimate business interests but with less disparate impact.

Conducting a robust search for LDA models can be a powerful approach to address disparate impact that's actually more effective than current approaches, and can directly address discrimination on an ex ante basis before consumers are negatively impacted. It's common sense and fitting that more sophisticated tools such as AI/ML models should call for more sophisticated techniques to address disparate impact.

A clear obligation to mitigate disparate impact and to search for and implement LDAs is also in line with broader emerging approaches towards rights-impacting AI (which includes decisions relating to insurance and credit, among others). A recent memo from the Office of Management and Budget (OMB) on the governance of agency use of AI states that agencies must mitigate algorithmic discrimination when it is present before using AI, including mitigating disparities that lead to or perpetuate harmful bias or decrease equity.[46]

Policymakers should take proactive steps to mainstream the LDA approach for high-risk AI/ML models. CFPB staff have already publicly stated that financial institutions are expected to conduct searches for LDAs as part of fair lending compliance under the Equal Credit Opportunity Act (ECOA).[47] This regulatory expectation, and what it entails in practice, should be clearly established by the CFPB. The obligation to conduct robust searches for LDAs should also be made to consistently apply wherever AI/ML models are employed for high-risk, consequential decisions that impact financial consumers, beyond just credit products. For example, NYDFS's circular on use of AI in insurance calls for insurance providers to search for LDAs when disparate impact is identified in AI/ML models for underwriting and pricing.[48] While ECOA is limited to credit, it may be feasible to leverage UDAAP authority to address discrimination in other contexts, as discrimination by definition logically fits under the definition of "unfair" acts.[49]

Policymakers should ensure that developers and deployers of AI/ML models conduct robust searches for LDAs and provide necessary guidance on what a robust LDA search should entail. In CR's and the Consumer Federation of America's joint letter to the CFPB from June 2024, we called for the CFPB to provide more guidance and examples on appropriate tools and techniques for conducting robust LDA searches in credit underwriting and pricing. Providing provide greater certainty and clarity to industry would particularly benefit financial institutions that may be hesitant to fully leverage AI/ML because it is

[46] Memorandum on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. OMB, March 28, 2024. https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf
[47] https://ncrc.org/cfpb-puts-lenders-fintechs-on-notice-their-models-must-search-for-less-discriminatory-alternatives-or-face-fair-lending-non-compliance-risk/
[48] Insurance Circular Letter No. 7 on Use of Artificial Intelligence Systems and External Consumer Data and Information Sources in Insurance Underwriting and Pricing. NYDFS, July 2024. https://www.dfs.ny.gov/industry-guidance/circular-letters/cl2024-07
[49] For further details on the application of UDAAP authority to discrimination, see Hayes, Stephen and Kali Schellenberg. Discrimination is "Unfair": Interpreting UDA(AP) to Prohibit Discrimination. Student Borrower Protection Center, April 2021. https://protectborrowers.org/wp-content/uploads/2021/04/Discrimination_is_Unfair.pdf

not clear how to do it responsibly and safely, including smaller community-focused institutions that are less sophisticated technologically. Clear guidance and examples would be particularly helpful regarding:

- The appropriate frequency and depth of LDA searches
- Appropriate metrics for measuring fairness in different contexts, etc. (e.g. adverse impact ratio (AIR), standardized mean differential (SMD))
- Appropriate debiasing techniques for developing LDAs (e.g. adversarial debiasing, joint optimizations, etc.)
- Key considerations when determining the viability of an LDA

Question 14

- As states adopt the NAIC's Model Bulletin on the Use of Artificial Intelligence Systems by Insurers and other states develop their own regulations or guidance, what changes have insurers implemented and what changes might they implement to comply or be consistent with these laws and regulatory guidance?
- How do insurers using AI make certain that their underwriting, rating, and pricing practices and outcomes are consistent with applicable laws addressing unfair discrimination?
- How are insurers currently covering AI-related risks in existing policies? Are the coverage, rates, or availability of insurance for financial institutions changing due to AI risks? Are insurers including exclusions for AI-related risks or adjusting policy wording for AI risks?

AI/ML is being leveraged extensively in the insurance sector for a range of use cases, from targeted marketing to underwriting and pricing to claims management and fraud monitoring. The increasing use of AI/ML in insurance heightens the risk of unfair or discriminatory outcomes. There have already been documented instances of harms to consumer arising from insurance providers' use of AI/ML across multiple use cases.

As noted previously, a fraud monitoring algorithm may systematically flag consumers on the basis of race or proxies for race, as illustrated in the lawsuit against State Farm claiming that its fraud detection software has a disparate impact on Black customers.[50] A pricing algorithm may systematically charge similarly situated consumers differently based on race or other sensitive characteristics, or proxies thereof. For example, telematics programs that obtain consumer-generated driving data for insurance pricing may result in unintended bias and disparate impact.[51] Pricing algorithms may also be used to charge prices based on a consumer's willingness to pay rather than actual risk. A joint investigation by CR and The Markup found that an advanced algorithm Allstate was proposing to deploy for pricing of auto insurance premiums "seemed to determine how much a customer was willing to pay – or overpay – without defecting."[52]

There is an urgent need to ensure that stronger and more consistent safeguards are established in the insurance sector to ensure transparency, accountability, and fairness in the use of AI/ML. CR has long advocated for insurance that is priced fairly based on the risk posed by the insured and underwriting that does not utilize variables with limited causal links to claims risk (such as credit scores). However,

[50] https://www.nytimes.com/2022/12/14/business/state-farm-racial-bias-lawsuit.html

[51] See Heller, Douglas and Michael DeLong. Watch Where You're Going: What's Needed to Make Auto Insurance Telematics Work for Consumers. Consumer Federation of America, May 2021. https://consumerfed.org/wp-content/uploads/2021/05/Insurance-Auto-Telematics-White-Paper-5-26-21.pdf

[52] https://www.consumerreports.org/money/car-insurance/why-you-may-be-paying-too-much-for-your-car-insurance-a5080204954/

while the use of AI/ML in the insurance raises a range of risks for consumers, regulatory safeguards are patchwork with significant gaps. Many state insurance laws do not recognize disparate impact, and only a few states have begun to substantively address the new and enhanced types of risks to consumers that arise from use of AI/ML in insurance.

Greater regulatory clarity would also be beneficial on the appropriateness of optimized pricing models, in the insurance context and beyond. AI/ML has super-charged the ability of companies to employ optimized pricing practices. Pricing practices that are solely designed to optimize profitability for companies while disregarding negative impacts on consumer well-being are problematic. Policymakers should carefully consider and clearly articulate when such practices may be deemed unfair or predatory. For example, EIOPA specifically states that price optimization practices such as those aiming to maximize consumer's "willingness to pay" should be avoided when they unfairly harm consumers, in particular vulnerable consumers or protected classes and in lines of business that are essential for financial inclusion.[53]

Question 18

- What actions are necessary to promote responsible innovation and competition with respect to the use of AI in financial services? What actions do you recommend Treasury take, and what actions do you recommend others take? What, if any, further actions are needed to protect impacted entities, including consumers, from potential risks and harms?
- Please provide specific feedback on legislative, regulatory, or supervisory enhancements related to the use of AI that would promote a financial system that delivers inclusive and equitable access to financial services that meet the needs of consumers and businesses, while maintaining stability and integrity, protecting critical financial sector infrastructure, and combating illicit finance and national security threats. What enhancements, if any, do you recommend be made to existing governance structures, oversight requirements, or risk management practices as they relate to the use of AI, and in particular, emerging AI technologies?

In our responses provided to the above questions, we have highlighted the following regulatory enhancements needed to ensure the safe and responsible use of AI/ML:

- Providing greater clarity and consistency on requirements and expectations regarding internal transparency into AI/ML models, including use cases where inherent interpretability may be required
- Requiring that consumers be clearly informed when an AI tool is being used to help make a consequential decision about them
- Requiring that consumers be provided with clear, specific explanations when they receive a consequential adverse decision
- Providing consumers with the right to appeal an AI-driven decision for human review
- Establishing good practices for digital chatbots and virtual assistants, including clear disclaimers to consumers when interacting with such tools, offering an easily accessible means to escalate to human assistance, and providing clear information regarding the capabilities and limitations of such tools
- Ensuring financial institutions utilizing consumer-facing AI systems meet high consumer protection standards, including regarding accuracy

[53] Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. EIOPA, June 2021. https://www.eiopa.europa.eu/system/files/2021-06/eiopa-ai-governance-principles-june-2021.pdf

- Providing clearer guidance on acceptable and prohibited practices in digital targeted marketing
- Providing greater clarity on the full range of measures financial institutions can and should be employing throughout each stage of the model development pipeline to directly address potential sources of algorithmic discrimination
- Clearly establishing the expectation that institutions should conduct robust searches for LDAs (for credit products as well as in other high-risk contexts), and providing guidance on what a robust LDA search entails
- Establishing stronger and more consistent safeguards to ensure transparency, accountability, and fairness regarding the use of AI/ML in the insurance sector
- Providing greater clarity on the appropriateness of optimized pricing models

Effectively addressing potential harms arising from AI/ML will require fully leveraging existing powers and authorities, while updating laws and regulations where needed to address key gaps. In addition to these specific regulatory enhancements, a number of legislative enhancements would help contribute to establishing strong safeguards regarding the use of AI/ML. There is a continued and urgent need to ensure strong data privacy laws as the current data economy underpins all AI/ML applications. In particular, data minimization requirements should be widely established and consumers should be empowered regarding control of their own data. It would also be beneficial to update ECOA to address any gaps and ensure its effectiveness in addressing new technological challenges, as well as to clarify the extent to which financial institutions can collect demographic data for narrow uses such as fair lending testing. In addition, actions should be taken to ensure that strong and consistent anti-discrimination safeguards apply across more situations where consumers can face significant harm.

The above enhancements will need to be complemented with sufficient resources to implementing agencies for supervision and enforcement. In order to effectively oversee fast-evolving technologies and ensure active supervisory coverage and monitoring across both bank and non-bank sectors, agencies will need the resources to further build up technical expertise and deploy sophisticated tools, for example to facilitate continuous monitoring or to undertake examinations tailored to AI/ML models and risks. As AI/ML use is cross-cutting across the financial sector, close coordination across financial sector agencies will also be necessary, in which Treasury can play a key role.

In conclusion, we would reiterate that AI/ML is a double-edged sword which can provide a number of benefits to consumers, but only if strong safeguards are put in place and effectively enforced. We appreciate the opportunity to submit these comments. For further information, please contact Jennifer Chien at jennifer.chien@consumer.org.

Sincerely,

Jennifer Chien
Senior Policy Counsel, Digital Marketplace

Consumer Reports
607 14th St NW #725
Washington, DC 20004
www.ConsumerReports.org