



October 18, 2019

Office of General Counsel
Department of Housing and Urban Development
451 7th St. SW
Room 10276
Washington, DC 20410

Re: ***HUD's Consideration of the Fair Housing Act's Disparate Impact Standard***
Docket No. FR-611-P-02; RIN 2529-AA98

Dear Assistant Secretary Farías:

BSA | The Software Alliance (BSA) is the leading advocate for the global software industry before governments and in the international marketplace.¹ Our members are at the forefront of software-enabled innovation that is fueling economic growth in every industry sector. As global leaders in the development of data-driven enterprise software solutions, BSA's members have a keen interest in working with policymakers to establish a legal environment that helps engender the public's trust and confidence in the technologies that are driving today's digital economy. We therefore welcome this opportunity to provide comments to the Department of Housing and Urban Development's (HUD) proposed rule concerning the interpretation of the Fair Housing Act's disparate impact standard.²

Given BSA's focus on the intersection of technology and policy, these comments focus narrowly on aspects of the Proposed Rule bearing on the use of Artificial Intelligence (AI) and the creation of potential safe harbors in circumstances where a "plaintiff identifies an offending policy or practice that relies on an algorithmic model."³ We are concerned that

¹ BSA's members include: Adobe, Akamai, Apple, Autodesk, Bentley Systems, Box, Cadence, CNC/Mastercam, DataStax, DocuSign, IBM, Informatica, Intel, MathWorks, Microsoft, Okta, Oracle, PTC, Salesforce, ServiceNow, Siemens PLM Software, Sitecore, Slack, Splunk, Symantec, Trend Micro, Trimble Solutions Corporation, Twilio, and Workday.

² 84 Fed. Reg. 42854 (August 19, 2019) [hereinafter "Proposed Rule"].

³ Proposed Rule at 42859.

the proposed safe harbors – as currently drafted – could undermine trust in digital technologies that increasingly are involved in high-stakes decisions that impact people’s lives.

BSA members are firmly committed to ensuring that their technologies enhance fairness and mitigate the potential for discrimination. As digital technologies are deployed in ways that implicate the public’s ability to obtain access to housing and finance, it is critical that HUD has the resources and authorities it needs to robustly enforce the Fair Housing Act’s prohibitions on discrimination. As a matter of principle, the public must be confident that the Fair Housing Act (FHA) will continue to afford the same level of protection irrespective of whether a lending or housing decision was made by a person, or a person assisted by a machine. One objective of this proceeding should therefore be to ensure that the use of technology will not hinder the enforcement of legitimate FHA claims. Simply put, existing laws should apply to the use of new technologies, and decisions that would otherwise incur liability under the FHA’s disparate impact standard should not benefit from a safe harbor merely because they involve the use of an AI system.

The use of advanced technologies in connection with housing and lending decisions presents both opportunities and risks. On the one hand, the adoption of AI by financial institutions has the potential to reduce discrimination and promote fairness by facilitating a data-driven approach to decision-making that is less vulnerable to human biases.⁴ For instance, the use of AI can improve access to credit and housing to historically marginalized communities by enabling lenders to evaluate a greater array of data than is ordinarily accounted for in traditional credit reports. At the same time, researchers caution that flaws in the design, development and/or deployment of AI systems have the potential to perpetuate existing social biases.⁵ Such biases can arise in a variety of ways, including circumstances in which an AI system is “trained” using data that reflects historical biases or when AI systems are deployed in populations that do not reflect the demographics of the data upon which they were trained.

Developing mechanisms for identifying and mitigating the risks of AI bias has emerged as an area of intense focus for experts in industry, academia, and government. In just the past few years, a vast body of research has identified a range of organizational best practices, governance safeguards, and technical tools that can help manage risks of bias throughout

⁴ See, e.g., Jennifer Sukis, *The origins of bias and how AI may be the answer to ending its reign*, Medium (Jan. 13, 2019), <https://medium.com/design-ibm/the-origins-of-bias-and-how-ai-might-be-our-answer-to-ending-it-acc3610d6354>.

⁵ See, e.g., Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*, Brookings (May 22, 2019), <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.

the AI lifecycle. Because static evaluations of AI models cannot account for all potential issues that may arise when AI systems are deployed in the field, experts agree that mitigating risks of AI bias requires a lifecycle approach, including ongoing monitoring by end-users to ensure that the system is operating as intended.

In light of the continuing evolution of this field of research, we urge HUD to take a cautious approach as it considers potential safe harbors for disparate impact claims arising from a defendant's use of algorithmic models. We appreciate HUD's clarification that the safe harbors are "not intended to provide a special exemption for parties who use algorithmic models" and instead are aimed at providing defendants with guidance about how they "can show their models achieve 'legitimate objectives.'"⁶ However, for the reasons outlined below, we are concerned that the proposed safe harbors may ultimately create greater uncertainty for entities that use and/or develop algorithmic systems and potentially exacerbate the risks associated with AI bias. We outline below the basis of our concerns.

I. The Proposed Rule's Inconsistent Use of Terminology Creates Uncertainty

HUD's explanation of the Proposed Rule describes the "first defense" (hereinafter Safe Harbor #1) and "third defense" (hereinafter Safe Harbor #3) as functionally "similar." HUD indicates that Safe Harbor #1 enables a defendant to prevail if it shows that the "model is not the actual cause of the disparate impact" through a "piece-by-piece" examination to determine whether "a factor used in the model is correlated with a protected class."⁷ HUD likewise characterizes Safe Harbor #3 as enabling a defendant to prevail if it proves (through the use of a qualified expert) that the "model is not the actual cause of the disparate impact."

Notwithstanding HUD's characterization of these defenses as functionally "similar," the proposed text for the defenses seems to employ terminology differently:

- Safe Harbor #1 can be invoked if a defendant shows that the "material factors that make up the inputs used in the challenged model...do not rely in any material part on factors that are substitutes or close proxies for protected classes under the Fair Housing Act."⁸
- Safe Harbor #3 can be invoked if a neutral third party validates that "none of the factors used in the algorithm rely in any material part on factors that are substitutes or close proxies for protected classes under the Fair Housing Act."⁹

⁶ Proposed Rule at 42859.

⁷ Id.

⁸ Id. at 42862 (§ 100.500 (c)(2)(i)).

⁹ Id. (§ 100.500 (c)(2)(iii)).

To avoid confusion, HUD should clarify whether the use of different terminology in Safe Harbor #1 and Safe Harbor #3 is intentional. To the extent the inquiries under Safe Harbor #1 and Safe Harbor #3 are intended to focus on different aspects of a challenged model, HUD should provide additional guidance in the final rule.

II. The Proposed Rule’s Focus on Individual Inputs is Both Over- and Under-Inclusive

Safe Harbor #1 and Safe Harbor #3 appear to create a bright line rule that would excuse disparate impacts that arise from a defendant’s use of an AI system that does not rely on individual inputs that are “substitutes or close proxies for protected classes under the Fair Housing Act.” Although the Proposed Rule lacks specific guidance about how HUD will assess whether an input to an algorithmic model is a “substitute” or “close proxy” to a protected class, HUD notes that the defenses would be unavailable if a plaintiff is able to demonstrate “that a factor used in the model is correlated with a protected class.”¹⁰ Conditioning eligibility for the safe harbor on an analysis that focuses on individual inputs would result in a range of unintended outcomes.

On the one hand, such a safe harbor would be unduly narrow. As a practical matter, it could have the effect of preventing lending institutions from relying on data inputs, such as income, that bear a close nexus to creditworthiness, but which may also be correlated to protected classes. Such a safe harbor could also preclude AI systems from containing features that have the effect of mitigating potential biases. Precluding the use of variables that are correlated to protected classes could deter lenders from using AI systems that leverage such variables for the explicit purpose of *preventing* disparate impacts.¹¹ Foreclosing the use of AI models that use protected classes (or proxies thereof) for the express purpose of de-biasing the model would of course be counterintuitive to the purpose of the Proposed Rule.

On the other hand, Safe Harbors #1 and #3 would also be overly broad, potentially privileging systems that produce discriminatory results based on inputs that bear no reasonably intuitive relationship to credit risk. The focus on individual inputs misapprehends the risk that a model may rely on a *combination* of facially neutral inputs that amount to a proxy for a protected class. By focusing only on the individual inputs to a model, the safe

¹⁰ Id. at 42859.

¹¹ See, e.g., Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*, Brookings (May 22, 2019), <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>

harbors could theoretically be invoked in circumstances where a model relies on facially neutral variables that produce extremely discriminatory outcomes. The risk is particularly pronounced if the facially neutral variables do not bear a reasonably explainable relationship to the target variable (e.g., credit risk) that the model is intended to measure.¹²

III. The Proposed Rule’s Reference to Industry Standards is Unclear

Safe Harbor #2 can be invoked by a defendant who uses an algorithmic model that is “produced, maintained, or distributed by a recognized third party that determines industry standards.”¹³ We seek additional clarity about the types of “industry standards” and “recognized” third parties to which this provision refers. Under a narrow reading, Safe Harbor #2 may only apply in circumstances where an international standard-setting body, such as the International Organization for Standards, both develops a standard and then distributes an associated algorithmic model that implements the standard. Under a broader reading, HUD may be referring to widely deployed technologies produced by an individual company. Alternatively, HUD may be referring to automated underwriting systems built on algorithmic models that are produced by government-sponsored entities (e.g., Fannie Mae and Freddie Mac).

IV. The Proposed Rule Creates Perverse Incentives that Exacerbate Risks of Bias

Safe Harbor #2 could also have the perverse effect of discouraging institutions from closely monitoring their own use of AI systems for unintended impacts. In the explanation of Safe Harbor #2, HUD suggests that liability for bias caused by algorithmic models that are “standard in the industry” should be borne by “the party that is actually responsible for the

¹² Such concerns prompted the Federal Reserve Board to issue a 2017 advisory bulletin cautioning against the use of facially neutral data inputs that do not bear a reasonably intuitive connection to creditworthiness. See Carol A. Evans, *Keeping Fintech Fair: Thinking About Fair Lending and UDAP Risks*, Consumer Compliance Outlook (Fed. Res. Sys., Phila, Pa.), 2017, <https://www.consumercomplianceoutlook.org/2017/second-issue/keeping-fintech-fair-thinking-about-fair-lending-and-udap-risks/> (“Careful analysis is particularly warranted when data may not only be correlated with race or national origin but may also closely reflect the effects of historical discrimination, such as redlining and segregation. For example, it’s been reported that some lenders consider whether a consumer’s online social network includes people with poor credit histories, which can raise concerns about discrimination against those living in disadvantaged areas. Instead of expanding access to responsible credit, the use of data correlated with race or national origin could serve to entrench or even worsen existing inequities in financial access. **Finally, it is important to consider that some data may not appear correlated with race or national origin when used alone but may be highly correlated with prohibited characteristics when evaluated in conjunction with other fields.**”) (Emphasis added.)

¹³ Id. at 42862 (§ 100.500 (c)(2)(ii)).

creation and design of the model.”¹⁴ Setting aside the uncertainty (noted above) about the type of industry standards this refers to, such a bright line rule overlooks the complexity of the AI ecosystem and threatens to establish a one-size-fits-all policy that may deter end-users from monitoring their own usage of an algorithmic model to ensure that is not creating a disparate impact. As noted above, the risk of bias must be continuously monitored because the performance of a model can be impacted if it is deployed into an environment in which the demographics differ from the data upon which it was trained. In many circumstances, only the entity that has deployed the model will be in a position to monitor its operation. However, Safe Harbor #2 could create a disincentive to perform such monitoring if doing so could increase their exposure to liability from which they would otherwise be shielded.

Conclusion

The growing ubiquity of AI has the potential to improve the delivery of services that will impact almost every facet of our daily lives. As AI is integrated into business processes that have consequential impacts on people – such as their ability to obtain access to credit or housing – it is imperative to ensure that existing legal protections apply even as technologies evolve. The public must be confident that these protections apply regardless of whether a decision is made by a person or by a machine. The safe harbors in the Proposed Rule would undermine that confidence, create uncertainty, and ultimately exacerbate the risks associated with AI bias.

The Proposed Rule’s safe harbors constitute the first intervention by a US government agency to define how civil rights protections will apply to the use of Artificial Intelligence. The complex issues that are implicated by the safe harbors strike at the heart of one of the most active areas of AI research and will arise again as other government agencies evaluate how the use of AI will impact their missions. Accordingly, we urge HUD to be very cautious and to consider whether these issues might benefit from a coordinated interagency consultation process.

The Executive Order on Maintaining American Leadership in AI tasked the Office of Science Technology and Policy and the Office of Management with the development of guidance for the heads of all agencies that is intended to “reduce barriers to the use of AI technologies in order promote their innovative application while protecting civil liberties.” Given that this Proposed Rule bears squarely on uses of AI that implicate core civil liberties protections, we urge HUD to consult closely with OSTP and OMB before issuing a final rule. We would likewise urge HUD to consult with the National Institute of Standards and Technology about the potential for convening a multistakeholder process for the purpose of developing an AI

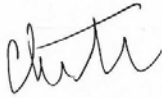
¹⁴ Id. at 42859.

lifecycle risk management framework. Such a process would enable experts from government, industry, and academia to collaborate on the development of a framework for identifying and mitigating the risks of bias that can emerge during the various phases of the AI lifecycle. The development of an AI risk management framework would be valuable not only for government agencies – such as HUD – that are developing AI policy, but also the companies that are developing and deploying AI technologies.

* * * * *

Thank you again for the opportunity to share our views on these important issues.

Sincerely,



Christian Troncoso
Director, Policy