COMMENTS OF THE
ELECTRONIC FRONTIER FOUNDATION
REGARDING PROPOSED RULE FOR HUD'S IMPLEMENTATION OF
THE FAIR HOUSING ACT'S DISPARATE IMPACT STANDARD

Docket No: FR-6111-P-02

84 Fed. Reg. 42854

Submitted on October 18, 2019 to the
Department of Housing and Urban Development

The Electronic Frontier Foundation (EFF) submits the following comments in response to U.S. Department of Housing and Urban Development's (HUD) proposal to amend the agency's interpretation of the Fair Housing Act's standard for disparate impact claims.[1]

EFF is a non-profit organization that has worked for almost 30 years to protect civil liberties, privacy, consumer interests, and innovation in new technologies. EFF actively encourages and challenges the executive and judiciary to support privacy and safeguard individual rights as emerging technologies become more prevalent in society. With more than 30,000 contributing members, EFF is a leading voice in the global and national effort to ensure that fundamental liberties are respected in the digital environment.

I. Introduction

EFF files these comments to object to HUD's proposed changes to its standard to address disparate impact claims under Title VIII of the Civil Rights Act of 1968 (Fair Housing Act).

Among other mandates, HUD enforces the Fair Housing Act, which prohibits discrimination on the basis of seven protected classes: race, color, national origin, religion, sex, disability, or familial status.[2] The Fair Housing Act was a landmark civil rights law passed in 1968 to counteract decades of government and private policies that

---

[1] 24 C.F.R. §§ 100 *et seq*.
[2] 42 U.S.C. §§ 3604-06, 3617.

promoted segregation in housing—including Jim Crow laws,[3] redlining,[4] and racial covenants.[5]

On February 15, 2013, HUD published a final rule that codified the agency's long-held interpretation[6] that the Fair Housing Act is violated not only by intentional discrimination, but also by "facially neutral practices that have an unjustified discriminatory effect on the basis of a protected characteristic, regardless of intent," or disparate impact.[7] The rule relied on case law under the Fair Housing Act and Title VII of the Civil Rights Act of 1964 to formalize a three-part burden-shifting framework for analyzing claims of disparate impact.[8] Under this burden-shifting framework, a plaintiff can establish a prima facie disparate impact claim by pointing to a challenged practice that resulted in discriminatory effect.[9] The burden shifts to the defendant to prove a legitimate, nondiscriminatory interest, and then shifts back to the plaintiff to present a less discriminatory alternative to the challenged practice.[10]

In 2015, in *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, the Supreme Court held that disparate impact claims were cognizable under the Fair Housing Act.[11] Because the underlying case presented a "novel theory" of disparate impact liability based on allocation of tax credits,[12] the Court "articulated some 'cautionary standards'" to ensure claims were focused on removing

---

[3] *See* Joseph P. Williams, *Segregation's Legacy*, U.S. News & World Report (Apr. 20, 2018), https://www.usnews.com/news/the-report/articles/2018-04-20/us-is-still-segregated-even-after-fair-housing-act.
[4] *See* Terry Gross, *A 'Forgotten History' Of How The U.S. Government Segregated America*, NPR (May 3, 2017), https://www.npr.org/2017/05/03/526655831/a-forgotten-history-of-how-the-u-s-government-segregated-america.
[5] *See* Motoko Rich, *Restrictive Covenants Stubbornly Stay on the Books*, N.Y. Times (Apr. 21, 2005), https://www.nytimes.com/2005/04/21/garden/restrictive-covenants-stubbornly-stay-on-the-books.html.
[6] 78 Fed. Reg. 11460, 11461-62 (demonstrating that HUD has taken the position that disparate impact claims are cognizable under the Fair Housing Act for nearly four decades in formal adjudications, policy statements, regulations implementing the Federal Housing Enterprises Financial Safety and Soundness Act, internal guidance to its staff, and external guidance published in the Federal Register and on its website). All eleven federal circuit courts of appeal that addressed disparate impact under the Fair Housing Act also agreed with HUD's interpretation. *Id.* at 11462 n.28 (citations omitted).
[7] *Id.* at 11461.
[8] *Id.* at 11462-63.
[9] 24 C.F.R. § 100.500(c).
[10] *Id.*
[11] 135 S. Ct. 2507 (2015).
[12] *Id.* at 2522-24.

artificial barriers rather than displacing valid government policies.[13] However, despite this cautionary language, scholars have noted the similarities between the framework articulated by the Court and that established under HUD's 2013 regulations.[14]

HUD's proposed rule claims to address discrepancies between its 2013 regulations and the *Inclusive Communities* decision, but instead uses this opportunity to dismantle the agency's interpretation of disparate impact theory. The proposed rule establishes a new burden-shifting framework for disparate impact claims that is nearly impossible for a plaintiff to meet. To establish a prima facie case, a plaintiff must plead five elements: (1) that the challenged practice is arbitrary, artificial, and unnecessary to achieve a valid interest; (2) that there exists a robust causal link between the challenged practice and a discriminatory effect on a protected class; (3) that the challenged practice has an adverse effect on members of a protected class; (4) that the disparity is significant; and (5) that the injury is directly caused by the challenged practice.[15]

Even if the plaintiff can meet this likely insurmountable prima facie standard, the proposed rule establishes three complete defenses for defendants that rely on algorithmic tools to make housing decisions, which are discussed in Part II. If the plaintiff carries their prima facie burden, the defendant must show that "the policy or practice advances a valid interest."[16] The burden shifts back to the plaintiff to prove that a less discriminatory alternative would serve the interest without imposing materially greater costs on the defendant.[17]

EFF opposes HUD's proposed rule, which marks a stark departure from nearly four decades of agency and court precedent that upholds disparate impact liability under the Fair Housing Act. Particularly, EFF has serious concerns regarding the proposed algorithmic defenses, which would shield defendants that rely on models to make housing determinations from disparate impact liability. These complete defenses are unnecessary, lack any scientific basis, and would allow discriminatory effect to persist without recourse.

---

[13] Robert G. Schwemm, *Fair Housing Litigation After Inclusive Communities: What's New and What's Not*, 115 Colum. L. Rev. Sidebar 106, 111 (2015) (citations omitted).
[14] *See id.* at 121.
[15] 84 Fed. Reg. 42854, 42862.
[16] *Id.* at 42863.
[17] *Id.*

II. HUD's Proposed Rule Creates an Unwarranted Safe Harbor for Defendants That Rely on Algorithmic Models to Make Housing Decisions.

The use of algorithmic models is rapidly expanding across a range of government spheres, including criminal justice,[18] child welfare,[19] education,[20] and immigration.[21] At their best, these models are useful tools that can optimize many areas of life. However, algorithmic models are not without their flaws. For machine-learning risk assessments, while a researcher feeds in certain criteria (e.g. inputs to use, outcomes to predict, learning methods to use), the risk assessment often functions as a "black box," making it difficult to determine why the model predicts certain outcomes.[22] For example, a study found that after Kentucky adopted a risk assessment model to determine pre-trial release, white defendants were more likely to be released than Black defendants—a distinction not present prior to the introduction of the tool.[23]

HUD's proposed rule would offer three affirmative defenses to defendants relying on an algorithmic model in making housing decisions if: (1) the individual inputs are not close proxies for protected class and the model is predictive of risk or other objective; (2) the model is created or maintained by a third party; or (3) a neutral third party has examined the model and determined the individual inputs are not discriminatory and the model is predictive of risk or other objective.[24] These complete defenses would effectively insulate those using algorithmic models from disparate impact lawsuits, even if the algorithmic model produced blatantly discriminatory outcomes. We discuss each of these defenses in turn.

---

[18] *See* Hayley Tsukayama & Jamie Williams, *If A Pre-Trial Risk Assessment Tool Does Not Satisfy These Criteria, It Needs to Stay Out of the Courtroom*, EFF Deeplinks Blog (Nov. 6, 2018), https://www.eff.org/deeplinks/2018/11/if-pre-trial-risk-assessment-tool-does-not-satisfy-these-criteria-it-needs-stay.
[19] *See generally* Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (2018).
[20] *See* Jay Cassano, *NYC students take aim at segregation by hacking an algorithm, Fast Company* (Apr. 16, 2019), https://www.fastcompany.com/90331368/nyc-students-take-aim-at-segregation-by-hacking-an-algorithm.
[21] *See* Chris Welch, *NYCLU sues ICE over changes to immigrant risk assessment algorithm*, The Verge (Dec. 12, 2018), https://www.theverge.com/2018/12/12/18138243/nyclu-lawsuit-ice-immigration-risk-assessment-tool.
[22] Megan Stevenson, *Assessing Risk Assessment in Action*, 103 Minn. L. Rev. 303, 315-16 (Nov. 2018).
[23] *See id.* at 362-68.
[24] 84 Fed. Reg. at 42862.

A. *HUD's Input Defense Erroneously Infers that an Algorithmic Model Cannot Be Discriminatory if It Does Not Use Protected Characteristics as Inputs.*

HUD's first complete defense allows a defendant to defeat a disparate impact claim where the defendant

> Provides the material factors that make up the inputs used in the challenged model and shows that these factors do not rely in any material part on factors that are substitutes or close proxies for protected classes under the Fair Housing Act and that the model is predictive of credit risk or other similar valid objective.[25]

However, the point of sophisticated machine-learning models is that they can learn how combinations of different inputs might predict something that any individual variable might not predict on its own. These combinations can be close proxies for protected characteristics, even if the individual input variables are not.

As a toy example, say a researcher was tasked with training a machine learning model to distinguish between penguins and other birds. As inputs, the model may use information about the bird, but it is forbidden from using any inputs that are close proxies for actually being a penguin. The researcher might choose inputs like whether a particular bird is flightless, where it lives, and what it eats. Being flightless is not a close proxy for being a penguin, because many other birds are flightless. Living in Antarctica is not a close proxy for being a penguin, because many other birds live in Antarctica. However, the combination of being flightless *and* living in Antarctica *is* a close proxy for being a penguin because penguins are the only flightless birds that live in Antarctica. In other words, while the individual inputs may not be close proxies for being a penguin, their combination is. Thus, even when an algorithmic model does not use protected characteristics or close proxies as inputs, the algorithm can still learn to combine inputs to infer someone's protected characteristics, and then potentially make discriminatory predictions based on those inferred protected characteristics.[26]

Apart from combinations of inputs, other factors, such as how an AI has been trained, can also lead to a model having a discriminatory effect. For example, some face classification technologies boast a positive predictive value for a general data set of over 93 percent.[27] When identifying women or people of color, the technology is less accurate—resulting in

---

[25] *Id.*
[26] Samuel Yeom *et al.*, *Hunting for Discriminatory Proxies in Linear Regression Models*, 32nd Conference on Neural Information Processing Systems (2018), at 9, https://www.cs.cmu.edu/~mfredrik/papers/Yeom18-nips.pdf.
[27] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, Proceedings on Machine Learning Research 81:1-15 (2018), at 9, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

more than a 20 percent error rate for dark-skinned women.[28] Unrepresentative training data can account in part for discrepancies like this, as a study found that "[o]ne widely used facial-recognition data set was estimated to be more than 75 percent male and more than 80 percent white."[29] Whether a model is discriminatory as a whole depends on far more than just the express inputs.

HUD claims its input defense allows a defendant to avoid liability when the model is "not the actual cause of the disparate impact alleged."[30] But showing that the individual inputs used in the model are not close proxies for protected characteristics ***does not mean that the model is incapable of discriminatory outcomes***.

Finally, HUD fails to clarify the meaning of the phrase "the model is predictive of credit risk or other similar valid objective."[31] What is the threshold for determining whether a model is predictive? What constitutes a valid objective? The aforementioned face recognition example illustrates that a model with predictive value nonetheless can lead to discriminatory effects. While a face recognition model with a 93 percent positive predictive value may be predictive in identifying the majority of the population, the predictive value of the technology can still be far less for minority groups, such as dark-skinned women. HUD's proposed rule offers a complete defense for defendants relying on models, but fails to articulate specific guidelines about the appropriate uses of such models.

B. *HUD's Third-Party Creator Defense Disincentivizes Housing Providers from Ensuring that Models Are Free from Discriminatory Effect and Raises Questions Regarding Liability.*

The proposed rule's second complete defense allows a defendant to defeat a claim of disparate impact where it

> Shows that the challenged model is produced, maintained, or distributed by a recognized third party that determines industry standards, the inputs and methods within the model are not determined by the defendant, and the defendant is using the model as intended by the third party.[32]

[28] *Id.*
[29] Steve Lohr, *Facial Recognition Is Accurate, if You're a White Guy*, N.Y. Times (Feb. 9, 2018), https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.
[30] 84 Fed. Reg. at 42859.
[31] *Id.* at 42862.
[32] *Id.*

This situation will apply frequently, as many housing providers already use algorithmic models created by someone else, and reliance on these tools is only increasing.[33]

Shielding direct defendants—like housing providers, mortgage lenders, and insurance companies—from liability will remove any incentive to ensure the models these entities rely upon are free from discriminatory impacts, or to pressure model creators to actively work to avoid such outcomes. Research has shown that some of the algorithmic models used in the housing context produce discriminatory outcomes on the basis of a protected characteristic, like race.[34] For example, one recent study of algorithmic discrimination in mortgage rates found that Black and Latinx borrowers paid approximately 5.3 basis points more in interest with online mortgage applications when purchasing homes than non-minority borrowers with similar credit scores and loan-to-value ratios.[35] Given this pervasive discrimination, HUD should be creating more incentives to address and root out systemic discrimination embedded in mortgage and risk assessment models, rather than proposing a rule that would shield from liability any housing entity that relies on these tools.

HUD implies that model makers should be held liable for discriminatory effect rather than entities that rely on such models. The proposed rule states that it offers this second defense because it "recognizes that there are situations in which standard practice is so clearly established that the proper party responsible for the challenged conduct is not the defendant, but the party who establishes the industry standard."[36] HUD further reasons that even "if the plaintiff prevails, the plaintiff would only remove the model from use by one party, whereas suing the party that is actually responsible for the creation and design of the model would remove the disparate impact from the industry as a whole."[37]

It is unclear whether aggrieved parties can actually get relief under the Fair Housing Act by suing the creator of the algorithm instead, as HUD suggests in its proposal. Earlier this

---

[33] *See* Lauren deLisa Coleman, *Inside The Alarming Way The Underbelly Of Algorithms Is Strangling The American Dream*, Forbes (Aug. 27, 2019), https://www.forbes.com/sites/laurencoleman/2019/08/27/inside-the-alarming-way-the-underbelly-of-algorithms-is-strangling-the-american-dream (noting that "landlords use tenant screening companies to apply algorithms to do credit and eviction checks, but now there are growing reports that some even use applicants' social media accounts to track their leisure activities and create additional profiles of what they see to be an ideal tenant").

[34] Annie Nova, *Online lending hasn't removed discrimination, study finds*, CNBC (Nov. 27, 2018), https://www.cnbc.com/2018/11/27/online-lenders-are-charging-minority-borrowers-more-study-concludes.html.

[35] Robert Bartlett *et al.*, *Consumer-Lending Discrimination in the FinTech Era* (May 2019), at 4, https://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf.

[36] 84 Fed. Reg. at 42859.

[37] *Id.*

year, in *Connecticut Fair Housing Center v. Corelogic Rental Property Solutions, LLC*, a federal district court held that a third-party creator could be held liable for a criminal history screening tool that was implemented by a landlord and led to discriminatory effect.[38] However, case law around third-party model creators is sparse.[39] If HUD's proposed rule is implemented, courts first must decide whether third-party model creators can be held liable under the Fair Housing Act for disparate impact discrimination before they can consider the merits of a case.

Because of this uncertainty, we ask that HUD respond to the following questions:

- Does HUD take the position that a plaintiff may sue a third-party model creator under the Fair Housing Act for developing a model that results in an unjustified discriminatory effect?
- If so, how does the new burden-shifting framework affect this position? Specifically, how would a plaintiff demonstrate that the discriminatory effect is "directly caused" by the algorithmic model?

C. *HUD's Neutral Third-Party Defense Falls into the Same Logical Fallacy as Its Input Defense.*

HUD's third proposed defense shields a defendant from disparate impact liability where the defendant

> Shows that the model has been subjected to critical review and has been validated by an objective and unbiased neutral third party that has analyzed the challenged model and found that the model was empirically derived and is a demonstrably and statistically sound algorithm that accurately predicts risk or other valid objectives, and 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.[40]

This defense shields a defendant if a neutral third party analyzed the model in question and determined—similar to the first defense—that the model's inputs are not close proxies for protected characteristics and is predictive of risk or another valid objective. This has the very same problem as the first defense: proving that the express inputs used in a model are not close proxies for a protected characteristic—even when analyzed by a

---

[38] 369 F. Supp. 3d 362 (D. Conn. 2019).
[39] *See, e.g.*, Schwemm, Fair Housing Litigation, *supra* note 13, at 125 (writing in 2015 that "[a]mong the housing policies that future FHA-based impact claims seem likely to challenge are: (1) landlords' screening devices based on an applicant's prior criminal record; . . . and (4) use by mortgage providers and others of credit scores or other financial qualifying techniques that disproportionately exclude racial minorities").
[40] 84 Fed. Reg. at 42862.

"qualified expert"—does not mean that the model itself is incapable of having a discriminatory impact.

Moreover, as mentioned in the first defense, HUD fails to define the meaning of the phrase "the model . . . accurately predicts risk or other valid objective."[41] What type of risk is a model intended to predict? What is the threshold for determining whether a model is predictive? What constitutes a valid objective? HUD offers sweeping defenses that can prevent plaintiffs from reaching discovery altogether, but fails to articulate what constitutes an acceptable use of algorithmic models.

HUD claims that the proposed rule's complete defenses are designed "merely to recognize that additional guidance is necessary in response to the complexity of disparate impact cases challenging these models" and that they are "not intended to provide a special exemption for parties who use algorithmic models."[42] ***But that is precisely what the proposed rule does.*** Under HUD's existing framework, once a plaintiff alleges a discriminatory effect based on an algorithmic model, the burden-shifting framework allows the parties to engage in discovery and determine how a model works, how it was trained, and other information necessary to determining disparate impact. This case-by-case determination, with careful attention paid to the particular facts at hand, is how disparate impact cases should work. In fact, in 2016, HUD rebutted insurance industry comments advocating for a safe harbor for certain practices, stating that "case-by-case adjudication is preferable to creating the requested exemptions or safe harbors for insurance practices" and that it would be "practically impossible for HUD to define the scope of insurance practices covered by an exemption or safe harbor with enough precision to avoid case-by-case disputes over its application."[43] The same inclination toward case-by-case determinations should be applied to algorithmic models, given the diversity in housing decision makers—including landlords, property management companies, banks, insurance providers, and advertisers—and the range of purposes for using such tools—including renting, selling, lending, insuring, and advertising. Instead, these three complete defenses will effectively foreclose disparate impact claims under the Fair Housing Act against any party that uses algorithmic models.

III. HUD's Proposed Rule Would Dismantle Disparate Impact Liability Under the Fair Housing Act.

Apart from the complete defenses, HUD also makes it much harder to bring a disparate impact claim because of the proposed rule's heightened prima facie standard and its disincentivization of data collection.

---

[41] *Id.*
[42] *Id.* at 42859.
[43] *Id.* at 42855 (citations omitted).

Heightened Prima Facie Standard. Even apart from the complete defenses offered to defendants who rely on algorithmic models, HUD's five-part prima facie case would require a herculean effort from a plaintiff to plead a viable disparate impact claim. This is especially concerning given that the widespread use of algorithmic models, particularly those derived using AI and machine learning, is still relatively new to the housing industry. The contours of when it is and is not acceptable to use algorithms—as well as what the standards should be for training datasets, algorithmic transparency, and fairness metrics—have yet to be clarified by HUD or Congress. As such, it falls to the courts to help determine these contours. However, given HUD's heightened pleading standard, it is unlikely that these questions will even reach the stage where a court can examine these questions. Until HUD or Congress provide greater clarity on how the housing industry can use algorithms in a manner that will not violate anti-discrimination laws—other than proposing unwarranted complete defenses—HUD should not make it more difficult for concerned plaintiffs to bring disparate impact cases.

No Adverse Inference on Collection of Protected Class Data. HUD's proposed rule adds in a new provision, section 100.5(d), which states: "Nothing in this part requires or encourages the collection of [protected class] data . . . . The absence of any such collection efforts shall not result in any adverse inference against a party."[44] This new provision potentially conflicts with existing law, which in some cases has required similar data collection in order to monitor and address discriminatory effects. For example, to counteract decades of redlining, Congress passed the Home Mortgage Disclosure Act, which requires financial lenders to report race, ethnicity, gender, and income of mortgage applicants and borrowers.[45] In addition, the Equal Credit Opportunity Act and its associated regulations create incentives for financial institutions to collect certain protected class information from non-mortgage applicants and conduct voluntary self-testing to determine whether its credit practices are resulting in discriminatory outcomes.[46] While EFF has serious concerns about the privacy implications of parties collecting data on individuals' protected characteristics, there may be situations where de-identified, aggregated collection of this data will be necessary to ensure that algorithmic models are not having unintended impacts on people based on their protected characteristics. At minimum, HUD should incentivize housing decision makers who wish to ensure that the models they use do not result in discriminatory outcomes by providing guidelines and incentives for privacy-preserving collection of such data. Instead, HUD does the opposite by discouraging data collection altogether, which may make it much harder to assess when a model results in a disparate impact based on a protected characteristic.

---

[44] *Id.* at 42862.
[45] 12 U.S.C.§§ 2801 *et seq.*; 12 C.F.R. §§ 1003 *et seq*.
[46] 12 C.F.R. §§ 1002 *et seq*.

IV. Conclusion

HUD's proposed rule departs from nearly forty years of consistent agency and court precedent by setting up a framework that would effectively dismantle disparate impact liability. The affirmative algorithmic defenses HUD proposes demonstrate the agency's lack of understanding of machine learning and forebode discriminatory effect without recourse. EFF respectfully urges HUD to rescind its proposed rule and continue to apply its existing burden-shifting framework to algorithmic models.

Respectfully submitted,

Saira Hussain
Staff Attorney
Electronic Frontier Foundation

Jeremy Gillula
Tech Projects Director
Electronic Frontier Foundation