March 15, 2024

The Honorable Adrienne A. Harris
Superintendent, New York State Department of Financial Services
One Commerce Plaza
Albany, NY 12257

RE: Proposed Circular on the use of AI and ECDIS in Insurance Underwriting and Pricing

Dear Superintendent Harris,

Consumer Reports[1] appreciates the opportunity to comment on the New York State Department of Financial Services' (NYDFS) proposed circular on the use of artificial intelligence systems (AIS) and external consumer data and information sources (ECDIS) in insurance underwriting and pricing. While artificial intelligence and machine learning (AI/ML) has the potential to enhance the provision of insurance for both insurers and customers, it comes with a range of risks, including the risk of discrimination and unfair treatment of consumers. We highly commend the NYDFS for continuing to take proactive action to address these new and enhanced types of risks and would highlight the areas below where further steps would be beneficial.

Algorithmic discrimination in insurance

AI/ML is increasingly being utilized by financial institutions for both back-end and front-end operations across the customer lifecycle, from targeted marketing to pricing and underwriting to claims management, fraud monitoring, and customer service. While AI/ML can bring efficiencies for institutions and financial inclusion benefits for consumers, these technologies also introduce a range of risks for consumers. 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.[2] Algorithmic discrimination occurs when an automated decision system repeatedly creates unfair or inaccurate outcomes for a particular group. While the risk of discrimination exists with traditional models, these risks are exacerbated by ML techniques for automated decision-making that rely on the processing of vast amounts of data using often opaque models.

[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.

[2] For example, see Barocas, Solon and Andrew D. Selbst. "Big Data's Disparate Impact." 104 California Law Review 671, 2016; 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.

Biased results can arise from a number of sources, including underlying data and model design. Unrepresentative, incorrect, or incomplete training data as well as biased data collection methods can lead to poor outcomes in algorithmic decision-making for certain groups. Data may also reflect historical biases, particularly the types of data sources used for underwriting which are tainted by past discriminatory practices. Biases can also be embedded into models through the design process, such as via improper use of protected characteristics directly or through proxies. With complex ML models utilizing hundreds or thousands of input features, chosen features may serve as proxies for protected characteristics. Choices made during the model development process can also affect its predictiveness regarding particular populations. The issue of potential bias and discrimination is further compounded by the lack of transparency for complex ML models.

For example, a fraud monitoring algorithm may systematically flag consumers on the basis of race or proxies for race, as illustrated in the recent lawsuit against State Farm claiming that its fraud detection software has a disparate impact on Black customers.[3] 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 impacts.[4] 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."[5]

CR has long advocated for insurance that is priced fairly based on the risk posed by the insured. The increasing use of AIS and ECDIS in insurance heightens the risk of unfair or discriminatory outcomes due to disparate treatment or disparate impact, particularly where complex "black box" ML models are used. There is a clear need to take proactive steps to ensure transparency, accountability, and fairness in the use of ECDIS and AIS in insurance, across the customer lifecycle. We believe that the proposed circular represents a significant step in this direction, though it could be further strengthened in a few key areas.

Proactive duty to search for and implement less discriminatory algorithms

In Section 14, the proposed circular states that insurers should not use ECDIS or AIS in underwriting or pricing unless the insurer can establish through a comprehensive assessment that its underwriting or pricing guidelines are not unfairly or unlawfully discriminatory. The circular further notes that a comprehensive assessment should entail at a minimum assessing whether the use of ECDIS or AIS produces disproportionate adverse effects in underwriting or pricing on similarly situated insureds or insureds of a protected class; if so, then assessing whether there is a legitimate, lawful, and fair explanation or rationale for the differential effect; and then conducting a search and analysis for a less discriminatory alternative variable or methodology that reasonably meets the insurer's legitimate business needs.

---

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

[4] 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.

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

We recognize that this three-step process generally aligns with existing and long-standing disparate impact doctrine, where similar three-step tests can be found with respect to discrimination in employment, housing, and credit. We believe that the proposed circular's clear and explicit recognition of the need to address disparate impact represents a significant step forward in addressing discrimination in insurance.

However, we would highlight that with ML models in particular, there is an opportunity to update the approach towards disparate impact to be more effective and efficient. In the context of ML models, the third step in the test regarding whether a less discriminatory alternative exists is no longer a real question. A less discriminatory algorithm nearly always exists due to model multiplicity. Model multiplicity refers to a phenomenon identified in recent computer science and statistics research 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."[6] It is no longer a question of whether or not a less discriminatory alternative can be found, but rather a matter of calling for companies to regularly search for and implement less discriminatory algorithms (LDAs).

Therefore, we would suggest the NYDFS consider revising Section 14 to require companies to proactively search for and implement LDAs (in essence, jumping immediately to step 3). This would be a natural and logical evolution of existing disparate impact doctrine, better aligns with modern technology, and has greater potential to achieve the intended policy objective of reducing discrimination. Developing ML models already involves weighing a series of choices and making continual refinements to optimize performance of the resulting model, selected from a universe of potential models. Incorporating disparate impact as an additional lens to apply could be integrated into the typical model development process, leading to the selection of an equally well-performing model with less disparate impact. 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."[7]

Preventing biased models from being developed on an ex ante basis would be a more direct approach to tackling discrimination than relying on ex post assessments, which the current phrasing of Section 14 seems to entail. To be effective, the search for LDAs should be undertaken systematically and throughout the entire ML model development pipeline, including considering implications to discrimination during early stages, as limiting the search for LDAs only to later stages or relying solely on post-processing techniques can be less efficient and lead to sub-optimal outcomes.[8] Rather than a one-

[6] Black, Emily, John Logan Koepke, Pauline T. Kim, Solon Barocas, and Mingwei Hsu. "Less Discriminatory Algorithms." Washington University in St. Louis Legal Studies Research Paper Forthcoming, October 2023.
[7] 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.
[8] For further details, 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." ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, Oct 2023.

off comprehensive assessment, we would suggest the circular reframe the duty to search for and implement LDAs as an integrated, ongoing requirement in the model development process.[9]

Guidance on searching for and selecting LDAs

In addition to reframing language to establish a clear duty to proactively search for and implement LDAs, it would be helpful for the NYDFS to provide further guidance (where possible) on how to conduct such a search and how to select between LDAs. We appreciate that this may be challenging as research on these topics is still limited and good practices are only beginning to emerge. But helping to shape expectations in these areas will be practically necessary to ensure consistent and robust implementation across the market. Additional guidance could be included in the circular itself or in follow-on materials.

In particular, it would be helpful to have further guidance, explanations, or examples on appropriate techniques to employ to search for LDAs that maximize reductions in disparate impact; appropriate fairness metrics to utilize,[10] which may differ by context; general expectations on the extent to which fairness should be optimized and disparate impact minimized; and clarifying the appropriate, limited use of sensitive demographic data in testing for fairness. The circular addresses some of these points in Section 17, but further guidance will likely be necessary.

Transparency requirements

With respect to transparency, the draft circular primarily focuses on external transparency for consumers and less so on challenges regarding internal transparency, i.e. transparency into the inner workings of a ML model for internal stakeholders. The circular briefly touches upon use of explanatory techniques in Section 17(vi), but otherwise does not address broader requirements regarding internal transparency, explainability, or interpretability.[11]

We would suggest expanding on the issue of internal transparency either in the circular or in future guidance, as regulatory expectations and standards in this area remain unclear. In particular, it would be useful to clarify to what extent and in what situations *post hoc* explainability techniques are considered suitably robust and sufficient for transparency purposes, and/or whether some level of interpretability is required to adequately assess and adapt models to address disparate impact. For example, the European Insurance and Occupational Pensions Authority (EIOPA) has stated that insurance firms should endeavor to use as much as possible explainable (i.e. interpretable) AI models, particularly where the AI use case has a significant impact on consumers.[12]

---

[9] Alternatively, it could be clarified that the three-step process in Section 14 will typically involve an iterative testing process. Another option would be to frame the ongoing duty to consider LDAs during the model development process as a complementary requirement to the need to conduct a comprehensive disparate impact assessment prior to deployment.

[10] For example, see examples of fairness and non-discrimination metrics in Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. European Insurance and Occupational Pensions Authority, 2021.

[11] The terms interpretability and explainability are not consistently defined and are at times used interchangeably. For purposes of this letter, we are using the term interpretability to refer to a modeling approach where transparency is baked into the model. In contrast, explainability is being used to refer to more "black box" models where *post hoc* explainability techniques are then used to provide some (limited) visibility into the model.

[12] Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. European Insurance and Occupational Pensions Authority, 2021.

A certain degree of interpretability will also be necessary in order to be able to point to sufficiently specific reasons for adverse actions, as is required in Section IV.E. For example, EIOPA has noted that use of AI specifically in pricing and underwriting of insurance will require a high level of transparency and explainability to achieve fairness for consumers, citing in particular the challenges posed when non-traditional factors such as telematics are processed by "black box" systems where features can only be roughly explained by existing, state-of-the-art *post hoc* explainability techniques.[13]

Adverse action notices

We fully support the provisions in the draft circular (as well as in past NYDFS circulars) to enhance transparency to consumers regarding the reasons for adverse actions as well as the data relied upon for such decisions, including external sources. Consumers have the right to know why their insurance policies have been refused, canceled, or limited. However, we think more could be done to ensure that such information is conveyed to consumers in a way that is user-friendly and actionable. Insurance products are already daunting for consumers to understand, and the use of ECDIS and AIS for pricing or underwriting only adds to the potential complexity and confusion that consumers may face. Providing such information without carefully considering how it is presented and framed to consumers will limit its practical use and helpfulness.

We would suggest that NYDFS provide further guidance or encouragement on how adverse action notifications should best be conveyed to consumers. At a minimum, information in adverse action notices should be provided in a user-friendly manner using plain language and explanations where necessary. 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[14] 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.[15] Financial institutions should ensure they have the answers to the following questions in their explanations:[16]

- 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.

---

[13] Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. European Insurance and Occupational Pensions Authority, 2021.
[14] 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.
[15] FEAT Transparency Principles Assessment Methodology. Veritas, 2022.
[16] FEAT Transparency Principles Assessment Methodology. Veritas, 2022.

In addition to more actionable adverse action notices, we would highlight the need for consumers to have the right to request human review of AIS-driven decisions, i.e. not to be subject only to automated decision-making. Such a right is clearly established in the EU's General Data Protection Regulation (GDPR)[17] as well as in MAS's FEAT Guidelines.[18] Establishing this type of right may require broader regulatory revision that is beyond this immediate circular, but we suggest this issue be considered for future action.

AI use cases beyond pricing and underwriting

Lastly, we would highlight the importance of expanding the types of requirements and safeguards included in the proposed circular to apply to AI use cases beyond pricing and underwriting. Algorithms and ML models are increasingly being used across the insurance customer lifecycle, from marketing to claims settlement to fraud monitoring. These additional stages are currently explicitly carved out of the proposed circular, yet algorithmic discrimination and disparate impact can arise in all of these instances. In fact, as previously noted, there are already examples of fraud monitoring algorithms used by insurance companies discriminating on the basis of race. Therefore, we would strongly urge the NYDFS to take further action to address algorithmic discrimination that may arise across all stages of the insurance lifecycle.

In conclusion, we would like to reiterate our strong support for the NYDFS' concrete actions to address algorithmic discrimination in insurance and 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, Financial Fairness

Consumer Reports
1101 17th St NW #500
Washington, DC 20036
www.ConsumerReports.org
(202) 462-6262

---

[17] Article 22, GDPR

[18] Principle 10 states that "Data subjects are provided with channels to enquire about, submit appeals for and request reviews of AIDA-driven decisions that affect them."