



## ASPEN TECH POLICY HUB

 THE ASPEN INSTITUTE

Scott Hallworth  
Chief Data, Model & Analytics Officer  
Fannie Mae  
1100 15th St. NW  
Washington, D.C. 20005

Frank Nazzaro  
Chief Information Officer  
Freddie Mac  
8200 Jones Branch Dr.  
McLean, VA 22102-3110

July 8, 2020

Dear Mr. Hallworth and Mr. Nazzaro,

We are statistical and legal researchers from universities around the country studying algorithmic lending, the practice of using advanced mathematical models, including machine learning, to decide and price consumer loans. Our work aims to detect and study discrimination in algorithmic lending and to develop alternative lending models that are more accurate and less discriminatory than existing methods.

We are deeply grateful for your agencies' commitment to transparency and for your historical support for the academic community, especially in providing public access to loan performance data.

We write to express our support for the proposal that your agencies grant researchers access to a combined single-family mortgage loan-level dataset that links borrower and loan characteristics, loan approval and pricing, and loan

performance. Such a dataset would lower the cost of building fair machine-learning (ML) lending models and increase these models' accuracy.

Access to this merged dataset would have a strong positive impact for the following reasons:

1. **Algorithmic lending has potential to discriminate less than face-to-face lending.** A [2019 study from researchers at the University of California, Berkeley](#) found that financial technology (fintech) algorithms discriminate 40 percent less than face-to-face lenders. We believe that, with more intentional work, algorithm-based lending could significantly improve fairness in lending.
2. **Crafting algorithmic models to predict loan default rates requires a clean training data set that links borrower characteristics to loan performance, not just to loan approval.** Using solely historic loan approval information, like that provided by the Home Mortgage Disclosure Act (HMDA), risks re-encoding discrimination in future loan approval decisions. Integrating the actual performance of loans would better ensure that a model is predicting the likelihood of default, as opposed to predicting whether a particular borrower would be granted a mortgage under current lending practices.
3. **With a more accurate merged dataset, researchers will be far more effective at detecting and studying discrimination** in loan pricing, as well as developing experimental machine-learning algorithms that more accurately predict default and are less discriminatory than their alternatives.

We respectfully request that you grant this request to make a merged loan-approval and loan-performance dataset available to researchers who are using machine learning to make lending more accurate — and less discriminatory.

Thank you very much for your consideration.

Signed,

Samara Trilling  
Aspen Institute Tech Policy Hub Fellow

Talia Gillis  
Associate Professor, Columbia Law School

Robert P. Bartlett, III

I. Michael Heyman Professor of Law & Faculty Co-Director of the Berkeley Center for Law and Business, UC Berkeley School of Law

Nancy Wallace

Professor & Lisle and Roslyn Payne Chair in Real Estate Capital Markets & Co-Chair, Fisher Center for Real Estate and Urban Economics, UC Berkeley Haas School of Business

Richard Stanton

Professor & Kingsford Capital Management Chair in Business Economics, UC Berkeley Haas School of Business

Adair Morse

Associate Professor & Solomon P. Lee Chair in Business Ethics, UC Berkeley Haas School of Business

Ryan Bressler, Principal Scientist