FinRegLab

900 19th St NW, Studio 6th Fl.
Washington, D.C. 20005
fingreglab.org

BY EMAIL

July 1, 2021

Board of Governors of the Federal Reserve System
Consumer Financial Protection Bureau
Federal Deposit Insurance Corporation
National Credit Union Administration
Office of the Comptroller of the Currency

RE: Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning, Docket No. OCC-2020-0049, FRB OP-1743, FDIC RIN 3064-ZA24, CFPB-2021-0004, NCUA-2021-0023

FinRegLab is pleased to submit these comments in response to the agencies' Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning released in March 2021 (the RFI). We are actively researching the use of artificial intelligence (AI) and machine learning (ML) technologies in financial services, with a particular focus on their potential implications for fairness and inclusion in credit underwriting. We welcome the agencies' interest in the topics outlined in the RFI, and agree that they are of urgent importance to facilitating the responsible adoption and use of these technologies to the mutual benefit of consumers and financial services providers.

FinRegLab will shortly publish a market context report reflecting our initial research into the state of ML adoption in credit underwriting, and is partnering with researchers at the Stanford University Graduate School of Business to conduct groundbreaking empirical research on issues concerning the transparency, reliability, and fairness of machine learning underwriting models. This comment provides a brief overview of our current understanding of the market, and will be supplemented with our research reports upon publication.

Background

Established in 2018, FinRegLab is an independent, nonpartisan innovation center that tests and monitors the use of new technologies and data to drive the financial services sector toward a responsible and inclusive marketplace. Through our research and policy discourse, we facilitate collaboration across the financial ecosystem to inform public policy and market practices.

Our first empirical evaluation focused on the use of cash-flow data from bank accounts and other sources, concluding that the data have substantial potential to increase inclusion in consumer and small business credit markets.[1] The agencies subsequently recognized the potential benefits of such data in their Interagency Statement on the Use of Alternative Data in Credit Underwriting, and in individual initiatives such as the Office of the Comptroller of the Currency's Project REACh.[2]

As described in further depth below, our new project will empirically evaluate the capability and performance of tools to explain and manage machine learning underwriting models with respect to reliability, fairness, and transparency, among other concerns. We expect that the research will also provide more general insights about the potential inclusion effects of using machine learning techniques in credit underwriting. Our work thus directly intersects with many of the topics raised in the agencies' RFI, and can help to inform policy analyses about the need to adjust market practices and regulatory frameworks to promote efficiency, fairness, and inclusion when using machine learning models.

Discussion

Financial services providers have begun to use machine learning models in a variety of business and operational contexts because they offer potential increases in the accuracy of predictions relative to incumbent models that have been used for decades. Depending on the context, such predictiveness improvements may allow providers to reduce losses due to default and fraud, cut processing times and costs, tailor their products, and/or expand their customer bases. The potential for such changes to benefit consumers and business customers as well—particularly in those segments of the market that currently struggle to access safe and affordable financial services—is motivating FinRegLab's research in this area.

We have launched a broad inquiry into the use of AI in financial services, including publishing a series of FAQs that highlight a number of different use cases and key considerations.[3] For purposes of these comments, however, we will focus primarily on use of machine learning in the credit context since that is the focus of our most intensive empirical and market research.

---

[1] FinRegLab, The Use of Cash-Flow Data in Underwriting Credit: Empirical Research Findings (2019); see also FinRegLab, The Use of Cash-Flow Data in Underwriting Credit: Small Business Spotlight (2019); The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis (2020).

[2] Board of Governors of the Federal Reserve System, Consumer Financial Protection Bureau, Federal Deposit Insurance Corporation, National Credit Union Administration & Office of the Comptroller of the Currency, Interagency Statement on the Use of Alternative Data in Credit Underwriting (2019); Office of the Comptroller of the Currency, OCC Announces Project REACh to Promote Greater Access to Capital and Credit for Underserved Populations (July 10, 2020).

[3] FinRegLab, Frequently Asked Questions, AI in Financial Services: Key Concepts (2020); FinRegLab, Frequently Asked Questions, AI in Financial Services: Explainability in Credit Underwriting (2020); FinRegLab, Frequently Asked Questions, Federated Machine Learning in Anti-Crime Financial Processes (2020); FinRegLab, Frequently Asked Questions, AI in Financial Services: The Data Science of Explainability (2021).

The Promise and Risks of Machine Learning Underwriting Models for Financial Inclusion

Applications of machine learning in the credit context are particularly important because credit plays such a critical role in borrowers' long-term financial health and economic participation. Credit can not only help bridge short-term gaps, but fund long-term investments in housing, transportation, education, and small business formation. The credit system thus both reflects and influences the ability of families, small businesses, and communities to participate in the broader economy.

The initial shift in consumer credit markets from subjective decision-making toward greater use of data and automated underwriting began more than 50 years ago. Research suggests that these changes have tended to lower underwriting costs and default losses, improve consistency of treatment of similarly situated applicants, and increase competition for borrowers.[4] Yet for all of these benefits, traditional scoring methods and underwriting models have limitations because they are dependent on data which is often not available for people in marginalized communities. We know, for instance, that about 20 percent of U.S. adults lack sufficient credit history to be scored under the most widely used models.[5] Prior to the pandemic, another 30 percent may have struggled to access affordable credit because their scores were considered to be "non-prime."[6] And small business owners often struggle to access commercial credit in part because of information gaps that discourage banks from serving the small end of the market.[7]

In each of these cases, communities of color and low-income populations are substantially more likely to be affected by these information barriers than other applicants. For example, nearly 30 percent of African-Americans and Hispanics cannot be scored using the most widely adopted

[4] Board of Governors of the Federal Reserve System, Report to Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit S-2 to S-4, O-2 to O-4, 32-49 (2007); Allen N. Berger & W. Scott Frame, Small Business Credit Scoring and Credit Availability, 47 J. of Small Bus. Mgmt. 5 (2007); Susan Wharton Gates *et al*., Automated Underwriting in Mortgage Lending: Good News for the Underserved?, 13 Housing Policy Debate 369 (2002); FinRegLab, The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis at 11 n.16.

[5] Consumer Financial Protection Bureau, Data Point, Credit Invisibles 4-6, 17 (2015); FinRegLab, The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis § 2.2.

[6] In lower score bands, the majority of applicants may be likely to repay but lenders cannot determine which particular applicants are lower risk without additional information. Lenders may choose not to lend to those cohort or may impose higher prices because default risks for the group as a whole are relatively high. For instance, depending on interest rates, consumers with scores near the typical minimums for approval may pay $7500 more over the life of a $20,000 auto loan and $86,000 more over the life of a $250,000 mortgage than peers with high scores. FinRegLab, The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis §§ 2.1, 2.2; Lyle Daly, Here's How Much Money Bad Credit Will Really Cost You, The Ascent (Apr. 8, 2019). Since the pandemic, average credit scores have risen due to changes in reporting practices, the effect of stimulus payments, and other factors, but there is also some evidence that lenders are relying on them less due to uncertainty. FinRegLab, Research Brief, Covid-19 Credit Reporting and Scoring Update 2 & nn. 7, 10 (2020); Elisabeth Buchwald, A Pandemic Paradox: Americans' Credit Scores Continue to Rise as Economy Struggles — Here's Why, MarketWatch (updated Feb. 20, 2021).

[7] FinRegLab,The Use of Cash-Flow Data in Underwriting Credit: Small Business Spotlight §§ 2.1, 2.2.

credit scoring models, compared to about 16 percent of whites and Asians. Racial disparities regarding access to credit are far greater than for more basic transaction accounts, for instance.[8]

Thus, the combination of machine learning and alternative financial data could be transformational where information gaps and other obstacles currently increase the cost or risk of serving particular consumer and small business populations using traditional models and data. For example, more nuanced models have the potential to assess default risks among consumers who lack extensive credit history and relatively new small businesses. Machine learning may also help analyze changes in economic conditions and detect more quickly nuanced signs of improvement in the financial capabilities of millions of people and small businesses, which may be especially useful for post-Covid recovery.

But the greater predictive power of machine learning models can increase risks as well as potential benefits, due to the models' greater complexity and to their potential to exacerbate historical disparities and other flaws in underlying data. Because the models are more complex, they are often more difficult for lenders and regulators to monitor appropriately over time. These concerns are especially important for two reasons. First, AI and machine learning models can be brittle because they have been overfitted to the data used in initial development and testing. This means that their performance can deteriorate when the conditions in which they are used differ from training and testing. Second, AI and machine learning models might amplify patterns of historical discrimination and financial exclusion due to reliance on flawed data or mistakes made in development and deployment. The greater complexity also makes it more challenging to explain to individual applicants why they were rejected or charged higher prices, and how they might improve their risk profiles over time.

Publicly available research supports the general predictiveness benefits of machine learning models, but provides only limited insights on these more complicated questions about reliability, fairness, and inclusion effects. For example, multiple academic studies have found substantial predictiveness gains from machine learning models relative to conventional credit card algorithms,[9] and a survey of five recent studies on the use of AI and machine learning in commercial lending reported gains of 2-3 percent on average, with one study reporting gains over 15 percent and another reporting 3-4 percent gains independent of alternative data.[10] But

[8] For example, a 2017 Federal Deposit Insurance Corporation survey found that about 10% of black and Hispanic households lacked bank and/or prepaid accounts, while more than 30% of both groups reported not having mainstream credit accounts of the type that are likely to be reported to credit bureaus. FinRegLab, The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis § 2.2; FDIC, 2017 National Survey of Unbanked and Underbanked Households (2018).

[9] Amir E. Khandani et al., Consumer Credit Risk Models via Machine-Learning Algorithms (2010); Florentin Butaru, *et al.*, Risk and Risk Management in the Credit Card Industry, J. of Banking & Finance (2016); Anastasios Petropoulos *et al.,* A Robust Machine Learning Approach for Credit Risk Analysis of Large Loan Level Datasets Using Deep Learning and Extreme Gradient Boosting, Bank for International Settlements (2018).

[10] Dinesh Bacham & Janet Zhao, Building AI in Credit Risk: A Commercial Lending Perspective, Moody's Analytics Risk Perspectives fig. 6 (July 2017).

limited information on the effects of machine learning underwriting models for populations that have historically struggled to access affordable credit creates a more complicated picture:

- VantageScore reports that its use of machine learning to develop scorecard models for consumers who are not scorable under some third-party models because their credit histories have not had an update in the prior six months resulted in a performance improvement of 16.6 percent for bank card originations and 12.5 percent improvement for auto loan originations.[11]
- An academic study of machine learning models using conventional data in the mortgage context concluded that such models would likely lead to modest improvements in application approvals among black and Hispanic applicants, but would increase pricing differentials between different demographic groups due to many minority applicants being evaluated as higher risk than under conventional approaches.[12]
- Another academic study shows that credit scores for minority groups generally reflect significantly more signal noise than other potential borrowers due to thin credit files, which may undercut inclusion effects of using machine learning models with traditional credit data.[13]

Although more research is needed, this suggests that the known inclusion benefits of using alternative financial data, such as cash-flow data,[14] may be enhanced when that data is used to develop machine learning underwriting models.[15]

The State of Adoption of Machine Learning Credit Underwriting Models

Machine learning can be used in a variety of ways in the credit context that could have implications for fairness and inclusion, including marketing, customer onboarding, fraud and illicit activities detection, originations, servicing, and collections. FinRegLab's immediate focus is on the use of machine learning models in underwriting, since such activities are at the heart of the lending process and are subject to the most federal regulatory scrutiny.

Thus, as a precursor to our empirical research on this use of AI and machine learning, we are conducting outreach to financial services stakeholders to understand the state of adoption of machine learning underwriting models and areas of greatest concern and uncertainty for stakeholders. This outreach suggests that some firms are only using machine learning models to help them develop traditional logistic regression models and scorecards by using them to identify variables and relationships that are particularly predictive of credit risk. Other firms, however,

[11] VantageScore, Our Models (undated), https://vantagescore.com/lenders/our-models#vantage-score-4.
[12] Andreas Fuster et al., Predictably Unequal? The Effects of Machine Learning on Credit Markets (Oct. 2020).
[13] Laura Blattner & Scott Nelson, How Costly Is Noise? Data and Disparities in the U.S. Mortgage Market (Jan. 2021).
[14] FinRegLab, The Use of Cash-Flow Data in Underwriting Credit: Empirical Research Findings.
[15] Blattner & Nelson, How Costly is Noise? (suggesting that the combination of machine learning underwriting models and alternative data, such as cash flow data, is required for greater financial inclusion).

are beginning to explore using machine learning algorithms directly in their underwriting models to evaluate individual applications. With regard to this latter group:

- Banks and nonbank lenders are interested in using machine learning underwriting models to make credit decisions due to their potential to improve the accuracy of credit risk assessment and reduce losses, to speed up the process of updating and refitting models, and to keep pace with market competitors. Many also cite the ability of machine learning models to leverage large, diverse data sets as a motivation. Nonbank usage is likely more established due to a number of factors, including reliance on digital business models, newer lending platforms, and differences in the nature and maturity of risk management processes.
- Credit cards and unsecured personal loans are the asset classes in which use of machine learning models to make credit decisions is most advanced. This reflects the historical position of credit cards as being at the analytical forefront of consumer finance and the dominance of digital lending in unsecured personal loans. Auto lending and small business lending are also areas where machine learning underwriting models are in use.
- Individual decisions about whether to use machine learning models to make credit decisions, what forms of machine learning to use, and how to enable appropriate oversight of such models varies based on firm culture and strategy and competitive dynamics in specific asset classes.
- Forms of machine learning used to make credit decisions range from gradient boosted trees and neural networks to ensembles combining multiple machine learning models.[16] Firms frequently introduce constraints to improve model transparency, even if those constraints impose performance tradeoffs. Common constraints include:
    - **Monotonicity constraints:** these constraints make it easier to understand the relationship between input data and predictions by ensuring one-directional relationships between the two; and
    - **Sparsity:** approaches such as regularization limit the number of features that a model uses to make a prediction.
- Decisions about whether to develop machine learning underwriting models in-house or to rely on third-party service providers are most likely to depend on the overall size of the lender and the importance of specific consumer asset classes to the institution. Many firms are likely to lack the resources – foremost among them personnel with appropriate data science and credit expertise – to develop and operate such models on their own.[17]

[16] Office of the Comptroller of the Currency, Credit Card Lending Handbook Version 2.0 at 17 (April 2021).

[17] Across both large and small institutions, approximately 20% of institutions have no in-house staff for credit modeling and rely on third parties to conduct such activities. Even large institutions with credit modeling teams do not devote significant resources, as just 16% of large institutions had four or more full time modelers. Cornerstone Advisors, Credit Monitoring and the Need for Speed: The Case for Advanced Technologies 4, fig. 4 (Q2 2020).

To meet this need, a number of potential third-party providers have entered this market, including score providers, technology firms, and consulting firms.

Prior to the pandemic, surveys of industry executives conducted by other organizations found that the respondents viewed AI/machine learning as a "major differentiator" in their businesses, and about half of the participating institutions lacked AI and machine learning capabilities in some or all of their platforms.[18] The pandemic has likely accelerated interest in and use of machine learning for making credit decisions, as it has accelerated adoption of all forms of AI across financial services and in other industries.[19] Nevertheless, our research suggests that broad conceptual questions about the trustworthiness of AI models (including the reliability in delivering enhanced performance), uncertainty about the net performance benefits of operating AI systems, concerns about compliance with existing federal regulatory requirements, and policy initiatives in other jurisdictions are shaping many firms' decisions in whether and how to use machine learning in connection with predicting default risk.

Core Questions Regarding the Use of Machine Learning Underwriting Models

Our ability to realize the potential of AI and machine learning to significantly enhance the efficiency, fairness, and inclusiveness of credit decisions depends on resolving uncertainty about when we can trust a specific model and when we cannot. Some machine learning underwriting models are by their nature more transparent than others, although a range of techniques has emerged that are designed to explain more opaque models. In the context of credit decisions, this means we need to understand how a model used in credit underwriting makes decisions and reaches particular outcomes such as denying certain credit applicants or charging those applicants more for their loan. Specifically, we need to be able to assess the model's stability and robustness, to provide an explanation for why a particular credit decision was made and a particular price was offered, and to understand if the model is treating people fairly. In this context, the significant jump in financial services leaders who report AI is being adopted "too fast for comfort" warrants close attention.[20]

Questions about the trustworthiness of AI and machine learning models pose a core challenge for all stakeholders in consumer and small business lending: how to enhance our ability to understand and rely on these technologies without unnecessarily diluting their ability to improve predictive power. This challenge is all the more complex in the financial sector, where extensive policy frameworks force consideration of questions about machine learning's trustworthiness more holistically and at an earlier stage than may occur elsewhere. Indeed, implementing AI in the financial system often requires meeting exacting requirements focused on securing the financial system from illicit activity, promoting responsible risk-taking, and providing broad, non-discriminatory access to credit.

---

[18] Leslie Parrish, Risky Business: The State of Play for Risk Executives in the Analytics Ecosystem, Aite 14, fig. 9 (2019).

[19] KPMG, Report: Thriving in an AI World 6-7 (April 2021).

[20] This study reported a 20% increase in financial services executives saying adoption of AI is moving too fast for comfort between 2020 and 2021. *Id.* at 8, chart 3.

There are a number of key issues that need to be addressed by policymakers and firms to foster responsible adoption and use of machine learning for credit underwriting:

Reliability

Some firms continue to explore whether machine learning models that meet governance and compliance requirements deliver sufficient performance gains in the short- and the long-term to make the operational costs of implementing and operating those models worthwhile. Others report significant improvement in lending outcomes. One key area of concern is our ability to understand and manage what happens to a model's predictions when data conditions in deployment differ from the data on which the model was trained. Current machine learning technologies are not generally well equipped for responding to such changes and may not even do well in recognizing them. Emerging tools to monitor this drift in the quality of a model's predictions have not been independently evaluated in the context of financial services.

Transparency

Lenders, as well as academics and other stakeholders, are debating whether certain "black box" technologies are ever appropriate for credit underwriting or lenders should only use interpretable or self-explanatory models.[21] Surveyed industry executives reported that explaining model results, and specifically adverse actions, was the most significant challenge for using AI and machine learning in decisioning applications for credit.[22]

The need to respond to a range of specific regulatory requirements, such as demonstrating general model performance and explaining specific loan decisions, requires different kinds of information and has led to the development of an array of explainability techniques. However, these techniques – which often involve the use of secondary machine learning models or analyses to explain the underwriting model – introduces a second layer of questions and concerns about the trustworthiness of information provided about the model's decision-making. We must ask whether we can trust the information produced by the secondary analyses and whether that information serves important oversight and governance needs.[23] But there is no consensus on what standards to apply to determine whether explanations of a model's decisions are trustworthy or are sufficient to establish compliance with particular legal or regulatory requirements.

Further, these explainability techniques are relatively new and are rapidly evolving. That means there is little independent evidence on their capability and performance in applications like credit underwriting where the stakes for consumers, communities, and firms are high and where exacting legal and regulatory requirements apply. As a result, there is considerable interest in understanding and evaluating the extent to which lenders have appropriate tools to explain

[21] Cynthia Rudin, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, Nature Machine Intelligence (Sep. 2019).

[22] Leslie Parrish, Alternative Data and Advanced Analytics: Table Stakes for Unsecured Personal Loans, Aite 16, fig. 12 (Nov. 2019).

[23] Agus Sudjianto, What We Need Is Interpretable and Not Explainable Machine Learning, presentation at Cogilytica Machine Learning Lifecycle Conference (Jan. 2021), available at https://events.cognilytica.com/wp-content/uploads/2020/12/What-we-need-is-interpretable-and-not-explainable-machine-learning-Agus-Sudjianto-ML-Lifecycle-Slides-.pdf.

machine learning models – from the simplest to the most opaque and complex – with sufficient confidence to satisfy themselves, their regulators, and potential critics.

Fairness

Ensuring that machine learning underwriting models do more than simply replicate historical lending patterns depends heavily on two important factors: (1) the articulation of expectations about what constitutes fair and responsible conduct for those who develop, operate, and monitor such technologies and (2) the choices that individual firms make in the model development and data selection processes, how they test and monitor model performance, and what customers their business and product strategies aim to serve. The transition to machine learning underwriting models holds promise for delivering models that are more accurate and fairer. But the efficacy of emerging approaches to deliver this win-win outcome or even to identify and manage bias responsibly throughout the model lifecycle need to be better understood. Those approaches include data diversification, reweighting or preprocessing data to reduce bias; improving techniques to manage bias in model training, such as adversarial debiasing or learning fair representation; and enhancing reject option classification and other post-processing techniques to reduce discrimination.

More broadly, stakeholders are currently debating what fairness requires in the context of lending, and many are proposing alternative methods for defining, measuring, and evaluating fairness. For example, the advent of AI and machine learning use more broadly has driven data scientists to proliferate metrics for measuring fairness,[24] and some academics and advocates have begun to ask if including protected class information in underwriting models might actually improve credit access for those in protected classes. At the same time, there is also a recognition that fairness is not just a mathematical problem and that it is one area where policy processes are needed to promulgate broadly applicable approaches to what fairness should mean. In considering this question, as well as others on model transparency, it is likely that the standards that emerge will apply even where machine learning models are not used and can thus drive the broader financial system to enhanced fairness and inclusiveness.

Privacy

The potential for machine learning underwriting models to analyze large, diverse data sets raises significant questions about privacy and consumer data controls and protections. Those questions include what data can be used for underwriting, how data are acquired, what constitutes informed consumer consent to a lender's acquisition and use of data, what constraints exist on firms' ability to use and retain that information, and whether and in what circumstances certain applicants should have to provide more access to information to facilitate credit risk assessment.

Specific Compliance Concerns

As the RFI itself recognizes, many of these broad conceptual questions about the trustworthiness of AI systems are also implicated in a number of specific existing federal regulatory requirements that apply to credit underwriting activities regardless of the type of underwriting model that is

[24] Sahil Verma & Julia Rubin, Fairness Definitions Explained, 2018 IEEE/ACM International Workshop on Software Fairness (FairWare), IEEE 1-7 (May 2018).

used. In particular, areas of uncertainty about the application of lending-specific laws and regulations include:

- **Fair lending:** Concern about the state of our ability to identify and control appropriately proxies for protected class information is paramount for firms using machine learning for credit underwriting and those considering such use. This concern is heightened in the context of more complex models and larger, more diverse data sets and includes consideration of whether traditional fair lending analysis and the prohibition on protected class data are suitable for use in the context of machine learning underwriting models.[25]
- **Adverse action reporting:** Required disclosures for applicants whose applications are denied or granted on less favorable terms based on information in a credit report make it necessary for lenders to be able to identify up to four primary bases of their decision. In models with higher numbers of variables and features, as well as models with more complex structures, producing accurate and reliable information for these notices can be a challenge.[26]
- **Model risk management:** The principles-based framework that governs banks' use of models requires model users to demonstrate the conceptual soundness and fitness for purpose of their models and to implement appropriate oversight and risk management measures. Some firms have reported navigating review of machine learning underwriting models in their model risk management programs, applying risk-based controls and governance appropriate to the type of model being used, the product being underwritten, and the customer base being served.

FinRegLab's Research on the Explainability and Fairness of Machine Learning Underwriting Models

FinRegLab decided to launch its new research on machine learning in credit underwriting specifically because of its heightened potential for both benefits and risks relative to traditional models and data. We also believed that credit underwriting could serve as a particularly useful case study with regard to questions about AI adoption in financial services and other contexts more generally because credit decisions are subject to a relatively robust set of regulatory requirements concerning model reliability, fairness, and transparency. These existing requirements provide a useful starting point for evaluating algorithmic decision-making and available tools for managing these models, although they themselves may need to be adjusted in response to evolution in modeling techniques and data.

[25] Talia Gillis, The Input Fallacy, Minn. L. Rev. (forthcoming 2022), April 2021 draft available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3571266.

[26] Parrish, Alternative Data and Advanced Analytics at 16, fig. 12. Surveyed industry executives cited explaining model decisions and adverse actions in particular as the most important challenge for using AI and machine learning.

Our research is designed to provide the first empirical data measuring available model diagnostic tools against these requirements. This research will help inform lenders' decisions about whether and in what circumstances they can trust machine learning underwriting models and policymakers' decisions about how protections and oversight processes need to be adapted to foster fair and responsible use of machine learning underwriting models.

More specifically, in partnership with researchers from the Stanford Graduate School of Business, we are empirically evaluating the performance and capabilities of emerging tools designed to help lenders develop, monitor, and manage machine learning underwriting models. FinRegLab is also conducting research on the market and policy implications of the use of machine learning underwriting models. This will be the first public research shaped by input from key stakeholders – including bank and fintech executives, data scientists, and consumer advocates – to address the questions about explainability and fairness that can promote responsible, fair, and inclusive adoption and use of machine learning in consumer credit.

Over the past year, FinRegLab has engaged these stakeholders to help inform our empirical assessment of a set of proprietary and open-source model diagnostic tools. We are evaluating how well these tools help lenders using machine learning models:

- Demonstrate the conceptual soundness, performance, and reliability of the models at the portfolio or lender level to satisfy prudential regulators and investors;

- Identify, measure, and enable mitigation of fair lending risks, particularly whether models have a disparate impact on protected classes; and

- Provide applicants with individualized "adverse action" notices explaining why they were denied credit or offered less favorable terms where required by law and regulation.

These represent a set of diverse requirements that apply to consumer lending regardless of the type of model being used to make credit decisions. Each one focuses attention on important threshold questions of model transparency related to the shift from incumbent automated underwriting models to machine learning models.

Drawing on nationally representative traditional credit data, we will evaluate the performance and capabilities of a set of proprietary and open-source model diagnostic tools using benchmark underwriting models developed by the research team. We expect to investigate proprietary tools provided by Fiddler Labs; H2O.ai, Relational AI, SolasAI/BLDS, LLC; Stratyfy, and Zest AI. We will assess these tools across a variety of dimensions:

- **Type of machine learning model:** benchmark underwriting models will range from logistic regression and boosted trees to neural networks and ensemble models to identify whether the type of underwriting model being explained affects the accuracy and utility of information produced by the model diagnostic tools;

- **Model complexity:** each form of machine learning being evaluated will have simple and complex forms to help us identify the tradeoffs, if any, between performance and transparency and between performance and fairness;

- **Changes in economic conditions:** test data sets will simulate different economic environments, such as data from 2009-2010, to help assess whether the model diagnostic tools can help lenders identify changes in data conditions and model performance once in operation; and

- **Shifts in applicant distribution:** test data sets will encompass different kinds of borrowers with respect to geographic location and socio-economic status to help us evaluate how well these tools detect fair lending and other risks.

Our set of benchmark models have generally been designed to approximate the models that lenders might use to estimate the risk of default associated with an application. In this evaluation, we will assess how a set of alternative definitions of algorithmic fairness that have emerged in academic literature work in the context of the underwriting models and model diagnostic tools used in our research.

In addition to empirical findings, we expect to put forth a framework that will help all stakeholders – model developers, risk and compliance personnel, and regulators – assess the accuracy and utility of accessible information about a machine learning underwriting model's decision-making. This framework will provide a substantive, thoughtful contribution to the current oversight approach about model transparency – defining the questions we should all be asking about the information that currently available model diagnostic tools produce. Those questions will help us assess whether those tools produce information that is necessary for assessing compliance with legal and regulatory requirements and policy goals. Our aim is that this framework will stimulate debate and evolution of a framework for promoting responsible, fair, and inclusive use of machine learning underwriting models.

Concluding Thoughts

We believe that our empirical, market, and policy research will help to inform a broad dialogue among financial services firms, policymakers, advocates, and others to enable adoption of machine learning in a responsible, fair, and inclusive way. As machine learning uses continue to spread in credit and other financial services contexts, determining how to refine and strengthen market practices and federal regulatory frameworks requires sustained attention and resources going forward.

We also expect that this project and other FinRegLab research initiatives will shed additional light on a range of issues concerning data governance protocols for the use of non-traditional financial information in credit underwriting. As discussed above, those issues are closely intertwined with machine learning underwriting models, and may be a bigger driver of financial inclusion when the two are used in tandem. But data bias and governance issues also arise in contexts that do

not involve machine learning in the first instance, and therefore require continuing direct attention in their own right to further refine best practices and regulatory expectations.

Beyond answering these basic questions about the tools and processes for managing machine learning models and non-traditional data, business and market considerations will also play a critical role in determining if these innovations in fact improve the inclusiveness and fairness of the credit system or of financial services more generally. Individual firms' decisions about strategy and market segmentation, operational barriers to technology and data adoption particularly among smaller firms, and acceptance by investors and secondary market actors of loans originated using non-traditional methods and data will all shape the extent to which model innovations are used to increase access to historically underserved populations or merely to target existing market segments with greater precision.

As our work progresses in the coming weeks and months, we would be pleased to share insights from our efforts. Our market review, empirical study, and policy analysis should help to inform the extent to which current laws and regulations are able to be satisfied in light of the emergence of more complex underwriting models, how well tools to develop and monitor those models perform in identifying effective ways to pursue greater inclusion and fairness, and considerations for policy and market developments that can support the safe, inclusive, and nondiscriminatory adoption of machine learning.

Thank you again for the opportunity to comment on these important topics.

Melissa Koide

Melissa Koide
CEO and Director

P–R Stark

P-R Stark
Director of Machine Learning Research