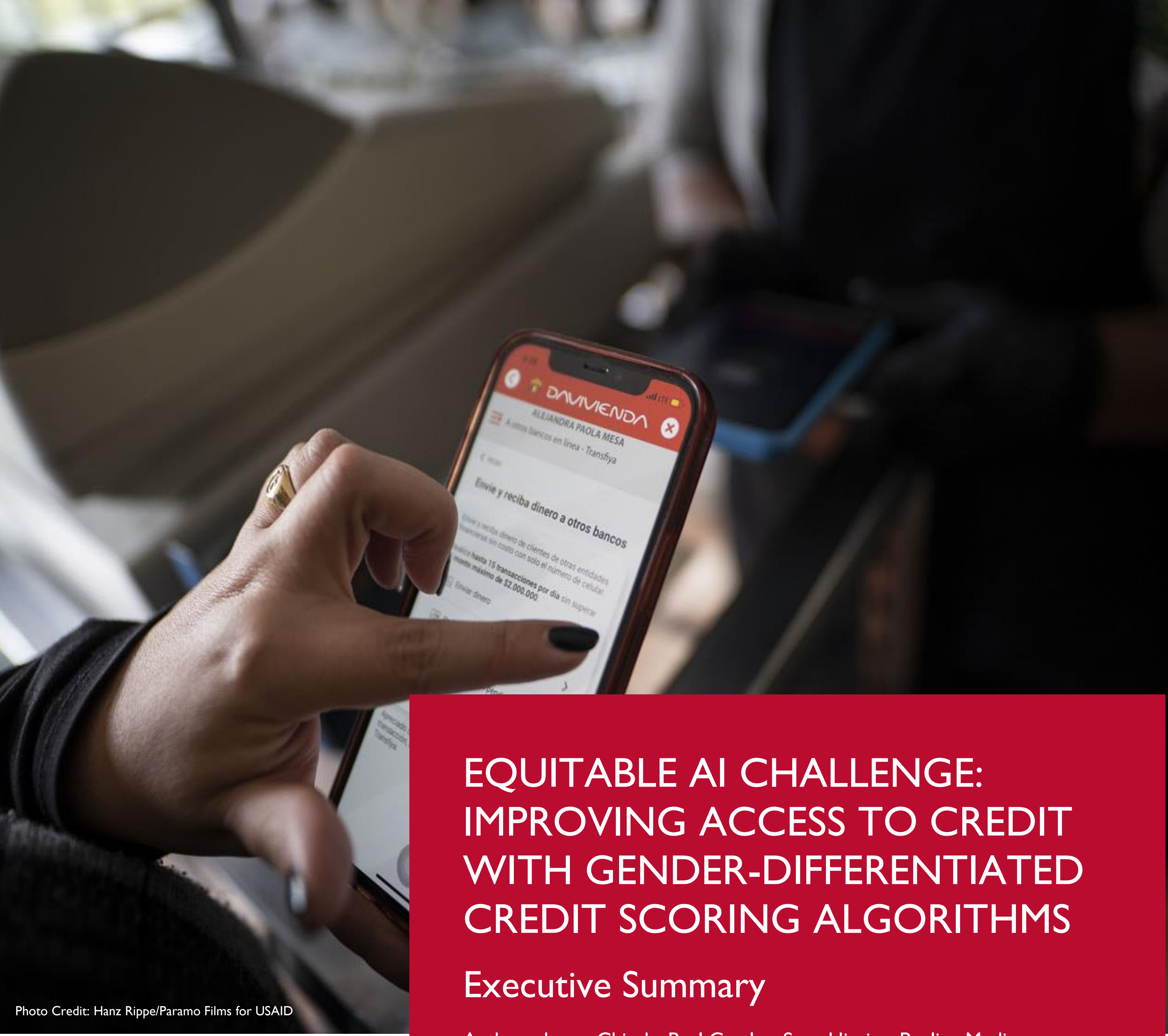

Photo Credit: Hanz Rippe/Paramo Films for USAID

EQUITABLE AI CHALLENGE: IMPROVING ACCESS TO CREDIT WITH GENDER-DIFFERENTIATED CREDIT SCORING ALGORITHMS

Executive Summary

Authors: Laura Chioda, Paul Gertler, Sean Higgins, Paolina Medina

April 2024

USAID FROM THE AMERICAN PEOPLE | DAI Shaping a more livable world. | Berkeley UNIVERSITY OF CALIFORNIA | UNIVERSITY OF HOUSTON

DISCLAIMER This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of DAI and do not necessarily reflect the views of USAID or the United States Government. This publication/report/guide was produced under DAI's Digital Frontiers Project (Cooperative Agreement AID-OAA-A-17-00033) at the request of USAID. This report was produced under USAID's Equitable AI Challenge Grant NO: 2022026.

EXECUTIVE SUMMARY

Although women default on loans less often than men (Karlan and Zinman, 2009; D'Espalier et al., 2011), gender gaps in access to formal credit have persisted despite the automation of evaluations for credit applications (Demirguç-Künt et al., 2015). We argue that a major (and underappreciated) driver of this gap is the way in which credit scoring models are calibrated and deployed.

Our study addresses two constraints that prevent low-income women from accessing credit. First, traditional data sources used in credit scoring models—such as credit histories, asset ownership, and formal earnings—are biased against women, who have historically been left out of credit markets. Second, the practice of pooling the data of men and women when training the credit scoring algorithm does not account for differences in behavior between men and women.

Traditional credit scoring models pool data from men and women and either omit gender entirely due to discrimination concerns (Mester, 1997), or include gender without fully capturing the ways in which gender interacts with other variables (e.g., Johnston and Morduch, 2008). For example, in the US, the Equal Credit Opportunity Act enacted in 1974 made it illegal for any creditor to discriminate based on sex or to consider gender when evaluating creditworthiness. Furthermore, credit scoring models are traditionally trained with data from credit bureau records and, as a result, individuals without a credit history cannot even be scored, leading to a cycle of exclusion.

A recent and fast-growing body of research studies the accuracy and validity of credit scoring models that leverage alternative sources of data such as digital footprints (i.e., the data that an individual creates through their actions online or through specific apps, including the type of device used and distinct patterns of online behavior).[1]

Overall, this literature points to the promise of alternative data to improve the prediction power of credit scoring models (Björkegren et al., 2015 (mobile footprint), Agarwal et al., 2020 (mobile footprint), Lee et al., 2022 (grocery data), DiMaggio et al., 2022 (Upstart fintech lending platform data), Berg et al., 2020 (website footprint)). The increased prominence of fintech lenders in the unsecured loans market and the widespread use of artificial intelligence and other technologies to assess credit risk has highlighted the importance of understanding the value of alternative data. The emerging research suggests that taking into consideration alternative data, in addition to the usual information provided in credit reports, translates into broader access to credit for borrowers with low credit scores (Di Maggio et al 2022).

In response to these challenges, the work, conducted under the Equitable AI Challenge (DAI-USAID grant)[2], aimed to address two primary questions:

---

[1] When a person is online, they leave a trail of information on all the things they have created, viewed, or interacted with. For instance, whenever they post or comment on something, purchase a product or service from an eCommerce site, or share pictures and videos, those actions and interactions become part of their digital footprint (Berg et al., 2020).

[2] We acknowledge complementary funding from the CEGA and the Bill & Melinda Gates Foundation Digital Credit Observatory (DCO) and the Lab for Inclusive Fintech (LIFT), both at UC Berkeley.

A. Can digital footprints, generated in the context of a (goods and services) delivery platform, serve as alternative data sources for individuals with thin or no credit histories, particularly underserved and unbanked females?

B. Can gender-differentiated credit scoring models expand access to credit for women, especially when leveraging nontraditional data sources?

We leverage a new type of digital footprint—obtained from the fintech arm of a large delivery platform—to train a credit scoring AI model. Our data consist of detailed transaction records of goods, services, and cash deliveries (e.g., groceries, pharmacy and medical services, cash, special deliveries, and other retail shops) that took place through a delivery app.

In particular, we study whether gender-differentiated credit scoring models trained with these new data can increase access to formal credit for underserved populations. Specifically, we test a new approach to credit scoring modeling that allows for men and women to have different determinants of loan eligibility. We then compare credit allocations generated by gender-differentiated models with those from a model that pools data on men and women (Mester, 1997, Johnston and Morduch, 2008).

Our preliminary findings confirm that digital footprints can indeed serve as valuable alternative data sources to assess the creditworthiness of populations with no or limited (traditional) credit histories. In particular, the credit scoring models developed for this project predict the likelihood of default more precisely compared to other models trained with alternative data in the economic literature for similar underserved populations.[3]

Furthermore, we find meaningful divergence between credit allocations based on models that pool men and women vs. those that train gender-specific models. Gender-segmented models can lead to an expansion in access to credit for women, without compromising lender's portfolio performance.

In other words, these findings suggest that the fairness and equity of algorithmic decisions—increasingly important concerns for regulation of the machine learning models—can be addressed by adopting gender-segmented models without meaningful losses in predictive accuracy nor a deterioration in portfolio default rates.

Notably, not all digital footprints are equal. We find that the performance of credit scoring models trained with alternative data improves significantly when users have longer transaction histories. In particular, the predictive power of models for users with longer transaction histories was comparable to that of models relying on credit bureau data for individuals with rich credit histories.

A limitation of our data, and of many fintech lending models, is that we only observe repayment and default outcomes for applicants who were approved for credit. That is, there is a substantial portion of applicants

---

[3] The standard measure to assess the quality of prediction in credit scoring models is the probability of correctly identifying the good case if faced with one random good and one random bad case (Hanley and McNeil 1982, Berg et al., 2020), also known as the "area under the curve" (AUC). The AUC ranges from 50% (pure random prediction) to 100% (perfect prediction) and is a simple and widely used metric for judging the discriminatory power of credit scores.

whose default behavior is not observed and who may significantly differ from our current sample. To address this selection bias, a lender would need to either lend to all applicants initially to obtain data on repayment for both populations and include both in the application scoring model, or infer performance of the rejected applicants (i.e., reject inference).[4]

In summary, this work highlights the potential of nontraditional data sources and gender-differentiated credit scoring models, and informs the policy debate on credit origination practices, both in countries that allow the use of protected variables in credit decisions and in those that restrict it. In particular, our findings add to the evidence that regulations requiring machine learning credit scoring models to be gender-blind could exacerbate the inequities that originally motivated them.

These results lay the foundation for a broader research agenda that aims to study welfare impacts of expanded access to credit leveraging gendered-differentiated credit models, as well as assess strategies to expand digital footprints for underserved populations.

[4] Machine learning credit scoring models need to be retrained periodically, typically every 12-18 months, to maintain their accuracy as the applicant pool may change over time. Failing to retrain the scoring algorithms can diminish their predictive power, which is especially important when using the model to expand credit access to underserved populations whose characteristics may differ from the original training data.

REFERENCES

- Agarwal, S., et al. (2020). Financial inclusion and alternate credit scoring for the millennials: role of big data and machine learning in fintech. Business School, National University of Singapore Working Paper. *SSRN 3507827*.
- Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7), 2845–2897.
- Björkegren, D., & Grissen, D. (2015). Behavior revealed in mobile phone usage predicts loan repayment. *SSRN 2611775*.
- D'Espallier, B., Guérin, I., & Mersland, R. (2011). Women and repayment in microfinance: A global analysis. *World Development*, 39(5), 758–772.
- Demirguç-Künt, A., Klapper, L., Singer, D., & Van Oudheusden, P. (2015). The global findex database 2014: Measuring financial inclusion around the world. *World Bank Policy Research Working Paper 7255*.
- Di Maggio, M., Ratnadiwakara, D., & Carmichael, D. (2022). Invisible primes: Fintech lending with alternative data. *NBER Working Paper No. w29840*.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36.
- Johnston, Don, and Jonathan Morduch. “The Unbanked: Evidence from Indonesia.” *The World Bank Economic Review*, vol. 22, no. 3, 2008, 517–37.
- Karlan, D., & Zinman, J. (2009). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77, 1993–2008.
- Lee, J. Y., Yang, J., & Anderson, E. (2022). Banking the unbanked: Using grocery data for credit decisions. Wharton Customer Analytics Research Paper.
- Mester, L. (1997). What’s the point of credit scoring? *Business Review*, 3, 3–16.